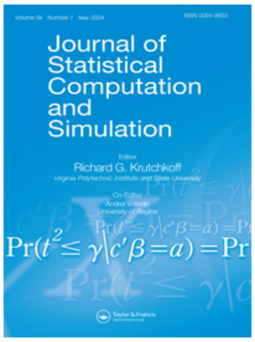

Journal of Statistical Computation and Simulation

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/gscs20

Higher-order asymptotic refinements in a multivariate regression model with general parameterization

Tatiane F. N. Melo, Tiago M. Vargas, Artur J. Lemonte & Alexandre G. Patriota

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION
https://doi.org/10.1080/00949655.2024.2361824

Higher-order asymptotic refinements in a multivariate regression model with general parameterization

Tatiane F. N. Melo[a], Tiago M. Vargas[a], Artur J. Lemonte[b] and Alexandre G. Patriota[c]

[a]Institute of Mathematics and Statistics, Federal University of Goiás, Goiás, Brazil; [b]Department of Statistics, Federal University of Rio Grande do Norte, Natal, Brazil; [c]Department of Statistics, University of São Paulo, São Paulo, Brazil

ABSTRACT

This paper derives a general Bartlett correction formula to improve the inference based on the likelihood ratio test in a multivariate model under a quite general parameterization, where the mean vector and the variance-covariance matrix can share the same vector of parameters. This approach includes a number of models as special cases such as non-linear regression models, errors-in-variables models, mixed-effects models with non-linear fixed effects, and mixtures of the previous models. We also employ the Skovgaard adjustment to the likelihood ratio statistic in this class of multivariate models and derive a general expression of the correction factor based on Skovgaard approach. Monte Carlo simulation experiments are carried out to verify the performance of the improved tests, and the numerical results confirm that the modified tests are more reliable than the usual likelihood ratio test. Applications to real data are also presented for illustrative purposes.

ARTICLE HISTORY
Received 8 December 2023
Accepted 23 May 2024

KEYWORDS
Bartlett correction; likelihood ratio test; parametric inference; Skovgaard adjustment

MATHEMATICS SUBJECT CLASSIFICATIONS
62F03; 62F05; 62F10

1. Introduction

A commonly used testing procedure is the well-known and time-honoured likelihood ratio (LR) test. The asymptotic chi-squared distribution (χ^2) of the LR statistic (denoted here as ω) is frequently used to test hypotheses of interest in parametric (regression) models, since its exact distribution is not easy to obtain in finite-sized samples. The null rejection rates (i.e. the rejection rates under the null hypothesis) based on the asymptotic distribution may be not close to the adopted significance level. In this case, therefore, the asymptotic approximation provided by the chi-squared distribution to the exact null distribution of the test statistic (that is, the distribution obtained under the null hypothesis) produces unreliable results. On the other hand, higher-order asymptotic theory may help attenuate this issue, since more terms in the Taylor expansion are considered in the approximation. Based on this procedure, Bartlett [1] proposed an improved LR statistic. His argument goes as follows: suppose that, under the null hypothesis, $\mathrm{E}(\omega) = k + B/n + \mathcal{O}(n^{-2}) = k\{1 + k^{-1}B/n + \mathcal{O}(n^{-2})\}$, where B is a constant that can be consistently estimated under

CONTACT Artur J. Lemonte arturlemonte@gmail.com Departamento de Estatistica, CCET, Universidade Federal do Rio Grande do Norte, Lagoa Nova, Natal/RN 59078970, Brazil

the null hypothesis and k is the difference of the dimensions of the parameter spaces under the alternative and null hypotheses. Then, the expected value of the transformed statistic $\omega^* = \omega/(1 + k^{-1}B/n)$ would be closer to the one from a χ^2_k distribution than the expected value of the original LR statistic. This became widely known as the Bartlett correction. Lawley [2] showed the remarkable fact that the Bartlett correction, under mild regularity conditions, guarantees that all the moments of the adjusted statistic ω^* are equal to those of the χ^2_k distribution up to order $\mathcal{O}(n^{-1})$. This implies that the Bartlett correction factor $1 + k^{-1}B/n$ also corrects the distribution of the statistic ω to order $\mathcal{O}(n^{-1})$. Moreover, the null distribution of the modified statistic ω^* remains χ^2_k with approximation error of order $\mathcal{O}(n^{-2})$ when B is replaced by a $\sqrt{n}$-consistent estimate (see, for example, [3]).

Bartlett correction has become a widely used method for improving the large-sample chi-squared approximation to the null distribution of the LR statistic when the sample size is finite. Several papers have been published giving closed-form expressions for computing Bartlett correction factors in special models. To mention a few, but not limited to, the reader is referred to Zucker et al. [4], Cordeiro [5], Barroso and Cordeiro [6], Cordeiro et al. [7], Fujita et al. [8], Araújo et al. [9], Guedes et al. [10] and Larsson [11], among many others. DiCiccio and Stern [12] provides a good discussion on Bartlett correction in both frequentist and Bayesian settings. For a detailed survey of Bartlett correction, the reader is referred to Cribari–Neto and Cordeiro [13] and Cordeiro and Cribari–Neto [14] and the references cited therein. It is worth mentioning that the vast majority of the works regarding Bartlett's corrections to the LR statistic consider that the mean vector and variance-covariance matrix do not share any parameters. In such cases, the Bartlett correction factor does not need some derivatives of the log-likelihood function with respect to the model parameters (and their corresponding expected values), simplifying its computation in some sense; see the above-mentioned references. There are, however, models where the mean vector and variance-covariance matrix share parameters as, e.g. the class of structural errors-in-variables models (see Section 2). Models with this complex parametric structure require also the mixed derivatives of the log-likelihood function with respect to the model parameters (and their expected values) to attain a general Bartlett correction factor to the LR statistic. This complexity may explain why the derivation of a Bartlett correction to the LR statistic is noticeable only in simple models where the mean vector and variance-covariance matrix share parameters. For instance, Wong [15] derived the Bartlett correction factor by means of orthogonal parameterization for a simple linear regression model when both variables are subject to measurement errors under a specific identifiability condition (see also Arellano-Valle and Bolfarine [16] for Bartlett corrections under different identifiability conditions), whereas Wong [17] computed the Bartlett correction factor under a measurement error model which compares several measurement instruments with a standard one. It should be noted that these works were carried out under simple errors-in-variables models. As we shall see in what follows, we have been able to derive a general Bartlett correction factor to the LR statistic in a multivariate model where the mean vector and the variance-covariance matrix can share parameters.

In this paper, we shall derive a general closed-form expression for the Bartlett correction factor to the LR statistic in the class of multivariate models with general parameterization to improve the inference based on the LR test in this class of models. This class of multivariate models was introduced by Patriota and Lemonte [18]. In the multivariate normal model with general parameterization, the mean vector and the variance-covariance matrix can share parameters. This approach unifies several important multivariate normal models, even errors-in-variables models in a more complex framework than the above mentioned such as, for instance, models under different kinds of heteroscedastic regimes. We also derive a closed-form expression for the Skovgaard [19]'s correction to the LR statistic under normality. Skovgaard's corrections were previously studied in Melo et al. [20] under a general class of elliptical distributions, but only an algorithm to compute the corrected version was offered. Here, instead, we provide a closed-form analytical expression to the Skovgaard correction under normality. It is worth stressing that the closed-form expression of the Skovgaard correction factor is provided using matrix notation, which has numerical advantages since it requires only simple operations on matrices and vectors.

One may argue that the use of improved tests becomes less appealing in the era of big data, since the null distribution of the conventional and improved test statistics will closely follow the asymptotic distribution for large sample sizes. This is true, but, naturally, there exist actual data examples where the sample size is small, and so the chi-squared approximation to the null distribution of the usual LR statistic will produce unreliable results. In these cases, where small-sized samples are observed, it is known that the asymptotic tests with no correction are less reliable than the corrected ones. Empirical examples with small samples are reported in some works. For example, Lemonte et al. [21] considered an empirical application to a biaxial fatigue dataset on the lifetime of $n = 15$ metal pieces in the process of metal extrusion; Vargas et al. [22] presented an empirical example related to an experiment to study the size of squid eaten by sharks and tuna, where the study involved measurements taken on $n = 22$ squids; and Medeiros and Ferrai [23] considered a real data example where it is investigated the effect of emulsion components on $n = 20$ orange beverage emulsion properties; among others. In short, even in the era of big data, small samples are observed in different areas and, consequently, the use of improved tests in such a case is always justifiable.

We provide general expressions to the correction factors in matrix notation. Although the algebraic forms of the Bartlett correction factor as well as the Skovgaard adjustment are cumbersome, they can be incorporated into some software with numerical linear algebra facilities such as object-oriented matrix programming language `Ox`; see Doornik [24].

This paper is organized as follows. Section 2 presents the model and some moments of derivatives of the log-likelihood function. In Section 3, we present a general matrix formula for computing the Bartlett correction factor. For the sake of completeness, we also present an expression in matrix notation to compute the Skovgaard adjustment to the LR statistic in this section. In Section 4, we consider some useful examples of the proposed formulation. Numerical evidence of the effectiveness of the finite sample corrections is presented and discussed in Section 5. The numerical results show that the modified tests are more reliable in finite samples than the usual LR test. Section 6 contains applications to real data. Finally, Section 7 concludes the paper.

2. The model

According to Patriota and Lemonte [18], the multivariate normal model under a general parameterization can be represented as

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where $\boldsymbol{Y}_i$ is a random q_i-vector of dependent variables, $\boldsymbol{\mu}_i := \boldsymbol{\mu}_i(\boldsymbol{\theta})$ is a mean function (whose shape is assumed to be known) which is three times continuously differentiable with respect to each element of $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^\top$. Also, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ is a p-vector of unknown parameters of interest. We assume that the independent random errors $\boldsymbol{\epsilon}_i$ ($i = 1, 2, \ldots, n$) follow a multivariate normal distribution, $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{q_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\Sigma}_i := \boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ is a positive definite $q_i \times q_i$ variance-covariance matrix, and the entries of $\boldsymbol{\Sigma}_i$ are assumed to be four times continuously differentiable in each element of $\boldsymbol{\theta}$. The functions $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ have to be defined in such a way that all elements of $\boldsymbol{\theta}$ are identifiable in the general model (1). It is evident that covariates can also be included in both $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$.

It is worth mentioning that the model in (1) covers many important regression models, probably it includes all multivariate normal models with independent vector responses. For example, it may be applied in an errors-in-variables model in which we observe two variables, namely Y_i and X_i, and the relationship between them is given by $Y_i = \alpha + \beta x_i + e_i$ and $X_i = x_i + u_i$, where $x_i \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $e_i \sim \mathcal{N}(0, \sigma^2)$ and $u_i \sim \mathcal{N}(0, \sigma_u^2)$, with σ_u^2 known. Additionally, consider that x_i, e_i and u_i are mutually uncorrelated. Then, denoting $\boldsymbol{Y}_i = (Y_i, X_i)^\top$ and $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_x^2, \sigma^2)^\top$, we can write this model as the general formulation (1), where $\boldsymbol{Y}_i \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and

$$\boldsymbol{\mu} := \boldsymbol{\mu}(\boldsymbol{\theta}) = \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix}, \quad \boldsymbol{\Sigma} := \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \beta^2\sigma_x^2 + \sigma^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix}.$$

The above model corresponds to the simple linear regression where the covariate x_i is not observed directly, instead of it, it is observed a surrogate variable, X_i. Notice that the mean vector and the variance-covariance matrix have a parameter in common, and so it corresponds to a simple example where the usual approach (assuming that the mean vector and the variance-covariance matrix do not share any parameter) is not applicable. The effort to derive a Bartlett correction factor in this simple model is very hard, as mentioned earlier. However, the general expression that we will derive for the Bartlett correction factor as well as for the Skovgaard adjustment term can be applied to this model with minimal effort.

To simplify notation, define $\boldsymbol{\mu}_i := \boldsymbol{\mu}_i(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_i := \boldsymbol{\Sigma}_i(\boldsymbol{\theta})$, $\boldsymbol{u}_i = \boldsymbol{Y}_i - \boldsymbol{\mu}_i$, $\boldsymbol{Y} = \mathrm{vec}(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n)$, $\boldsymbol{\mu} := \boldsymbol{\mu}(\boldsymbol{\theta}) = \mathrm{vec}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \ldots, \boldsymbol{\mu}_n)$, $\boldsymbol{\Sigma} := \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathrm{diag}\{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots, \boldsymbol{\Sigma}_n\}$ and $\boldsymbol{u} = \boldsymbol{Y} - \boldsymbol{\mu}$, where $\mathrm{vec}(\cdot)$ is the vec operator, which transforms a matrix into a vector by stacking the columns of the matrix one underneath the other. The log-likelihood function for $\boldsymbol{\theta}$, apart from an unimportant constant, is given by

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{i=1}^{n} \log\{\det(\boldsymbol{\Sigma}_i)\} - \frac{1}{2}\sum_{i=1}^{n} \mathrm{tr}\{\boldsymbol{\Sigma}_i^{-1}\boldsymbol{u}_i\boldsymbol{u}_i^\top\}$$

where $\mathrm{tr}\{\cdot\}$ means the trace operator. Note that we can also express $\ell(\boldsymbol{\theta}) = -(1/2) \log\{\det(\boldsymbol{\Sigma})\} - (1/2)\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\boldsymbol{u}^\top\}$. Here, we assume valid the usual regular conditions [25, Ch. 9] on the behaviour of $\ell(\boldsymbol{\theta})$ as the sample size n approaches infinity, such as the regularity of the first three derivatives of $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and the uniqueness of the maximum likelihood (ML) estimate $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_p)^\top$ of $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^\top$. Moreover, define the following quantities (we naturally assume that they do exist and are finite):

$$\boldsymbol{a}_{i(r)} = \frac{\partial \boldsymbol{\mu}_i}{\partial \theta_r}, \quad \boldsymbol{a}_{i(rs)} = \frac{\partial^2 \boldsymbol{\mu}_i}{\partial \theta_r \partial \theta_s}, \quad \boldsymbol{C}_{i(r)} = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \theta_r},$$

$$\boldsymbol{C}_{i(rs)} = \frac{\partial^2 \boldsymbol{\Sigma}_i}{\partial\theta_r\partial\theta_s}, \quad \boldsymbol{A}_{i(r)} = \frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial\theta_r} = -\boldsymbol{\Sigma}_i^{-1}\boldsymbol{C}_{i(r)}\boldsymbol{\Sigma}_i^{-1},$$

$$\boldsymbol{a}_r = \frac{\partial \boldsymbol{\mu}}{\partial\theta_r}, \quad \boldsymbol{a}_{rs} = \frac{\partial^2 \boldsymbol{\mu}}{\partial\theta_r\partial\theta_s}, \quad \boldsymbol{a}_{rst} = \frac{\partial^3 \boldsymbol{\mu}}{\partial\theta_r\partial\theta_s\partial\theta_t}, \quad \boldsymbol{C}_r = \frac{\partial \boldsymbol{\Sigma}}{\partial\theta_r},$$

$$\boldsymbol{C}_{rs} = \frac{\partial^2 \boldsymbol{\Sigma}}{\partial\theta_r\partial\theta_s}, \quad \boldsymbol{C}_{rst} = \frac{\partial^3 \boldsymbol{\Sigma}}{\partial\theta_r\partial\theta_s\partial\theta_t}, \quad \boldsymbol{C}_{rstv} = \frac{\partial^4 \boldsymbol{\Sigma}}{\partial\theta_r\partial\theta_s\partial\theta_t\partial\theta_v},$$

$$\boldsymbol{A}_r = \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial\theta_r} = -\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_r\boldsymbol{\Sigma}^{-1},$$

$$\boldsymbol{A}_{rs} = \frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial\theta_r\partial\theta_s} = -\boldsymbol{A}_r\boldsymbol{C}_s\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{rs}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_s\boldsymbol{A}_r,$$

$$\begin{aligned}\boldsymbol{A}_{rst} = \frac{\partial^3 \boldsymbol{\Sigma}^{-1}}{\partial\theta_r\partial\theta_s\partial\theta_t} &= -\boldsymbol{A}_{rs}\boldsymbol{C}_t\boldsymbol{\Sigma}^{-1} - \boldsymbol{A}_s\boldsymbol{C}_{rt}\boldsymbol{\Sigma}^{-1} - \boldsymbol{A}_s\boldsymbol{C}_t\boldsymbol{A}_r \\ &- \boldsymbol{A}_r\boldsymbol{C}_{st}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{rst}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{st}\boldsymbol{A}_r \\ &- \boldsymbol{A}_r\boldsymbol{C}_t\boldsymbol{A}_s - \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{rt}\boldsymbol{A}_s - \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_t\boldsymbol{A}_{rs},\end{aligned}$$

$$\begin{aligned}\boldsymbol{A}_{rstv} = \frac{\partial^4 \boldsymbol{\Sigma}^{-1}}{\partial\theta_r\partial\theta_s\partial\theta_t\partial\theta_v} &= -\boldsymbol{A}_{rst}\boldsymbol{C}_v\boldsymbol{\Sigma}^{-1} - \boldsymbol{A}_{st}\boldsymbol{C}_{rv}\boldsymbol{\Sigma}^{-1} - \boldsymbol{A}_{st}\boldsymbol{C}_v\boldsymbol{A}_r - \boldsymbol{A}_{rt}\boldsymbol{C}_{sv}\boldsymbol{\Sigma}^{-1} \\ &- \boldsymbol{A}_t\boldsymbol{C}_{rsv}\boldsymbol{\Sigma}^{-1} - \boldsymbol{A}_t\boldsymbol{C}_{sv}\boldsymbol{A}_r - \boldsymbol{A}_{rt}\boldsymbol{C}_v\boldsymbol{A}_s - \boldsymbol{A}_t\boldsymbol{C}_{rv}\boldsymbol{A}_s \\ &- \boldsymbol{A}_t\boldsymbol{C}_v\boldsymbol{A}_{rs} - \boldsymbol{A}_{rs}\boldsymbol{C}_{tv}\boldsymbol{\Sigma}^{-1} - \boldsymbol{A}_s\boldsymbol{C}_{rtv}\boldsymbol{\Sigma}^{-1} - \boldsymbol{A}_s\boldsymbol{C}_{tv}\boldsymbol{A}_r \\ &- \boldsymbol{A}_r\boldsymbol{C}_{stv}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{rstv}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{stv}\boldsymbol{A}_r - \boldsymbol{A}_r\boldsymbol{C}_{tv}\boldsymbol{A}_s \\ &- \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{rtv}\boldsymbol{A}_s - \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{tv}\boldsymbol{A}_{rs} - \boldsymbol{A}_{rs}\boldsymbol{C}_v\boldsymbol{A}_t - \boldsymbol{A}_s\boldsymbol{C}_{rv}\boldsymbol{A}_t \\ &- \boldsymbol{A}_s\boldsymbol{C}_v\boldsymbol{A}_{rt} - \boldsymbol{A}_r\boldsymbol{C}_{sv}\boldsymbol{A}_t - \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{rsv}\boldsymbol{A}_t - \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{sv}\boldsymbol{A}_{rt} \\ &- \boldsymbol{A}_r\boldsymbol{C}_v\boldsymbol{A}_{st} - \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{rv}\boldsymbol{A}_{st} - \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_v\boldsymbol{A}_{rst},\end{aligned}$$

where $r, s, t, v = 1, 2, \ldots, p$. In addition, define the following quantities: $\boldsymbol{D}_i = [\boldsymbol{a}_{i(1)}, \boldsymbol{a}_{i(2)}, \ldots, \boldsymbol{a}_{i(p)}]$, $\boldsymbol{V}_i = [\text{vec}(\boldsymbol{C}_{i(1)}), \text{vec}(\boldsymbol{C}_{i(2)}), \ldots, \text{vec}(\boldsymbol{C}_{i(p)})]$,

$$\boldsymbol{F}_i = \begin{bmatrix}\boldsymbol{D}_i \\ \boldsymbol{V}_i\end{bmatrix}, \quad \boldsymbol{H}_i = \begin{bmatrix}\boldsymbol{\Sigma}_i & \boldsymbol{0} \\ \boldsymbol{0} & 2(\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i)\end{bmatrix}^{-1}, \quad \boldsymbol{u}_i^* = \begin{pmatrix}\boldsymbol{u}_i \\ -\text{vec}(\boldsymbol{\Sigma}_i - \boldsymbol{u}_i\boldsymbol{u}_i^\top)\end{pmatrix},$$

$$\boldsymbol{F} = \begin{bmatrix}\boldsymbol{D} \\ \boldsymbol{V}\end{bmatrix}, \quad \boldsymbol{H} = \begin{bmatrix}\boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & 2(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\end{bmatrix}^{-1}, \quad \boldsymbol{u}^* = \begin{pmatrix}\boldsymbol{u} \\ -\text{vec}(\boldsymbol{\Sigma} - \boldsymbol{u}\boldsymbol{u}^\top)\end{pmatrix},$$

where $\boldsymbol{D} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p]$, $\boldsymbol{V} = [\text{vec}(\boldsymbol{C}_1), \text{vec}(\boldsymbol{C}_2), \ldots, \text{vec}(\boldsymbol{C}_p)]$, and $\otimes$ is the Kronecker product. We assume that $\boldsymbol{F}_i$ has rank p. From Patriota and Lemonte [18], the score function has the form $\boldsymbol{U} := \boldsymbol{U}(\boldsymbol{\theta}) = \boldsymbol{F}^\top \boldsymbol{H}\boldsymbol{u}^*$, and the expected Fisher information matrix is given by $\boldsymbol{K} := \boldsymbol{K}(\boldsymbol{\theta}) = \boldsymbol{F}^\top\boldsymbol{H}\boldsymbol{F}$. The observed Fisher information matrix can be expressed as $\boldsymbol{J} := \boldsymbol{J}(\boldsymbol{\theta}) = \sum_{i=1}^n (\boldsymbol{F}_i^\top \boldsymbol{H}_i\boldsymbol{M}_i\boldsymbol{H}_i\boldsymbol{F}_i + \boldsymbol{G}_i)$, where $\boldsymbol{G}_i$ is a $p \times p$ matrix whose the (r, s)th element is given by $-(\boldsymbol{a}_{i(rs)}^\top \text{vec}(\boldsymbol{C}_{i(rs)})^\top)\boldsymbol{H}_i\boldsymbol{u}_i^*$ for $r, s = 1, 2, \ldots, p$, and

$$\boldsymbol{M}_i = \begin{bmatrix}\boldsymbol{\Sigma}_i & 2\boldsymbol{\Sigma}_i \otimes \boldsymbol{u}_i^\top \\ 2\boldsymbol{\Sigma}_i \otimes \boldsymbol{u}_i & \boldsymbol{M}_{i;22}\end{bmatrix},$$

with $\boldsymbol{M}_{i;22} = 2[(\boldsymbol{u}_i\boldsymbol{u}_i^\top) \otimes \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i \otimes (\boldsymbol{u}_i\boldsymbol{u}_i^\top) - \boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i]$. The Fisher scoring method can be used to estimate $\boldsymbol{\theta}$ by iteratively solving the equation $(\boldsymbol{F}^{(m)\top}\boldsymbol{H}^{(m)}\boldsymbol{F}^{(m)})\boldsymbol{\theta}^{(m+1)} = \boldsymbol{F}^{(m)\top}\boldsymbol{H}^{(m)}\boldsymbol{u}^{**(m)}$, where $m = 0, 1, 2, \ldots$ is the iteration counter, and $\boldsymbol{u}^{**(m)} = \boldsymbol{F}^{(m)}\boldsymbol{\theta}^{(m)} + \boldsymbol{u}^{*(m)}$. Under regularity conditions, the ML estimators obtained using this iterative equation are asymptotically normally distributed and we may write $\widehat{\boldsymbol{\theta}} \overset{a}{\sim} \mathcal{N}_p(\boldsymbol{\theta}, \boldsymbol{K}^{-1})$, when n is large, $\overset{a}{\sim}$ denoting approximately distributed. The reader is referred to Patriota and Lemonte [18] for further details.

3. Improved LR tests

Let the p-vector $\boldsymbol{\theta}$ be partitioned as $\boldsymbol{\theta} = (\boldsymbol{\psi}^\top, \boldsymbol{\xi}^\top)^\top$, where the ν-vector $\boldsymbol{\psi}$ represents the parameter of interest, and the $(p-\nu)$-vector $\boldsymbol{\xi}$ is the nuisance parameter vector. Consider the problem of testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\psi} = \boldsymbol{\psi}^{(0)}$ against the composite alternative hypothesis $\mathcal{H}_1 : \boldsymbol{\psi} \neq \boldsymbol{\psi}^{(0)}$, where $\boldsymbol{\psi}^{(0)}$ is a fixed ν-vector. Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\psi}}^\top, \widehat{\boldsymbol{\xi}}^\top)^\top$ be the ML estimate of $\boldsymbol{\theta}$, and $\widetilde{\boldsymbol{\theta}} = (\boldsymbol{\psi}^{(0)\top}, \widetilde{\boldsymbol{\xi}}^\top)^\top$ be the ML estimate of $\boldsymbol{\theta}$ obtained by imposing that the null hypothesis holds. The LR statistic for testing $\mathcal{H}_0 : \boldsymbol{\psi} = \boldsymbol{\psi}^{(0)}$ is given by

$$\omega = 2[\ell(\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\xi}}) - \ell(\boldsymbol{\psi}^{(0)}, \widetilde{\boldsymbol{\xi}})].$$

Under the null hypothesis, the limiting distribution of the test statistic ω is χ^2_ν (i.e. chi-squared distribution with ν degrees of freedom). The null hypothesis is rejected for a given nominal level, α say, if the observed value of the test statistic ω exceeds the upper $100(1-\alpha)\%$ quantile of the central χ^2_ν distribution. It is worth emphasizing that the chi-squared distribution χ^2_ν may be a poor approximation to the exact null distribution of the above statistic when the sample size is not sufficiently large. It is thus important to obtain refinements for inference based on this test from second-order asymptotic theory. Next, we derive a general Bartlett correction factor to the LR statistic, and we also present an adjusted LR statistic based on the Skovgaard [19] procedure.

The Bartlett-corrected LR statistic for testing $\mathcal{H}_0 : \boldsymbol{\psi} = \boldsymbol{\psi}^{(0)}$ is given by

$$\omega_{\mathrm{B}} = \frac{\omega}{1 + \nu^{-1}\Delta},$$

where $\Delta := \Delta(\boldsymbol{\theta}) = \epsilon_{(p)} - \epsilon_{(p-\nu)}$, and $\epsilon_{(p)}$ and $\epsilon_{(p-\nu)}$ can be attained from formula (2). It follows that $\Pr(\omega_{\mathrm{B}} \le z) = \Pr(\chi^2_\nu \le z) + \mathcal{O}(n^{-2})$, whereas $\Pr(\omega \le z) = \Pr(\chi^2_\nu \le z) + \mathcal{O}(n^{-1})$, a clear improvement. After long and tedious algebra, we arrive at the following proposition.

Proposition 3.1: *The Bartlett correction factor in the multivariate normal model with general parameterization is obtained from*

$$\epsilon_{(p)} = \mathrm{tr}\{\boldsymbol{K}^{-1}[\boldsymbol{L} - \boldsymbol{M} - \boldsymbol{N}]\}, \tag{2}$$

where the $p \times p$ matrices $\boldsymbol{L}$, $\boldsymbol{M}$ and $\boldsymbol{N}$ are defined through the entries

$$L_{rs} = \mathrm{tr}\{\boldsymbol{K}^{-1}\boldsymbol{E}^{(rs)}\},$$

$$M_{rs} = -\frac{1}{6}\mathrm{tr}\{\boldsymbol{K}^{-1}\boldsymbol{P}^{(r)}\boldsymbol{K}^{-1}\boldsymbol{P}^{(s)}\} + \mathrm{tr}\{\boldsymbol{K}^{-1}\boldsymbol{P}^{(r)}\boldsymbol{K}^{-1}\boldsymbol{Q}^{(s)}\} - \mathrm{tr}\{\boldsymbol{K}^{-1}\boldsymbol{Q}^{(r)}\boldsymbol{K}^{-1}\boldsymbol{Q}^{(s)}\},$$

$$N_{rs} = -\frac{1}{4}\mathrm{tr}\{\boldsymbol{P}^{(r)}\boldsymbol{K}^{-1}\}\mathrm{tr}\{\boldsymbol{P}^{(s)}\boldsymbol{K}^{-1}\} + \mathrm{tr}\{\boldsymbol{P}^{(r)}\boldsymbol{K}^{-1}\}\mathrm{tr}\{\boldsymbol{Q}^{(s)}\boldsymbol{K}^{-1}\}$$
$$- \mathrm{tr}\{\boldsymbol{Q}^{(r)}\boldsymbol{K}^{-1}\}\mathrm{tr}\{\boldsymbol{Q}^{(s)}\boldsymbol{K}^{-1}\},$$

where $r, s = 1, 2, \ldots, p$. The (t, v)th elements of the matrices $\boldsymbol{E}^{(rs)}$, $\boldsymbol{P}^{(r)}$, and $\boldsymbol{Q}^{(s)}$ become

$$E_{tv}^{(rs)} = -\frac{1}{8}\mathrm{tr}\{\boldsymbol{A}_{tvr}\boldsymbol{C}_s + 4\boldsymbol{A}_{str}\boldsymbol{C}_v + \boldsymbol{A}_t\boldsymbol{C}_{vrs} + \boldsymbol{A}_v\boldsymbol{C}_{trs} + \boldsymbol{A}_r\boldsymbol{C}_{tvs} + \boldsymbol{A}_{tv}\boldsymbol{C}_{rs} + \boldsymbol{A}_{vr}\boldsymbol{C}_{ts}$$
$$+ 5\boldsymbol{A}_{tr}\boldsymbol{C}_{vs} + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{tvrs} + \boldsymbol{A}_{tvrs}\boldsymbol{\Sigma}\} - \frac{1}{4}(\boldsymbol{a}_t^\top\boldsymbol{A}_{vs}\boldsymbol{a}_r + \boldsymbol{a}_t^\top\boldsymbol{A}_{vr}\boldsymbol{a}_s - 3\boldsymbol{a}_t^\top\boldsymbol{A}_{rs}\boldsymbol{a}_v$$
$$- 3\boldsymbol{a}_v^\top\boldsymbol{A}_{ts}\boldsymbol{a}_r + \boldsymbol{a}_s^\top\boldsymbol{A}_{tv}\boldsymbol{a}_r + \boldsymbol{a}_s^\top\boldsymbol{A}_{tr}\boldsymbol{a}_v + \boldsymbol{a}_t^\top\boldsymbol{A}_s\boldsymbol{a}_{vr} - 3\boldsymbol{a}_t^\top\boldsymbol{A}_r\boldsymbol{a}_{vs} + \boldsymbol{a}_t^\top\boldsymbol{A}_v\boldsymbol{a}_{rs}$$
$$+ \boldsymbol{a}_s^\top\boldsymbol{A}_t\boldsymbol{a}_{vr} + \boldsymbol{a}_s^\top\boldsymbol{A}_v\boldsymbol{a}_{tr} + \boldsymbol{a}_s^\top\boldsymbol{A}_r\boldsymbol{a}_{tv} + \boldsymbol{a}_r^\top\boldsymbol{A}_v\boldsymbol{a}_{ts} - 3\boldsymbol{a}_r^\top\boldsymbol{A}_t\boldsymbol{a}_{vs} + \boldsymbol{a}_r^\top\boldsymbol{A}_s\boldsymbol{a}_{tv}$$
$$- 3\boldsymbol{a}_v^\top\boldsymbol{A}_s\boldsymbol{a}_{tr} - 3\boldsymbol{a}_v^\top\boldsymbol{A}_r\boldsymbol{a}_{ts} - 3\boldsymbol{a}_v^\top\boldsymbol{A}_t\boldsymbol{a}_{rs} + \boldsymbol{a}_{tv}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{rs} - 3\boldsymbol{a}_{tr}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{vs}$$
$$+ \boldsymbol{a}_{vr}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{ts} + \boldsymbol{a}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{vrs} - 3\boldsymbol{a}_v^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{trs} + \boldsymbol{a}_r^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tvs} + \boldsymbol{a}_s^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tvr}),$$

$$P_{tv}^{(r)} = -\frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_{tv}\boldsymbol{C}_r + \boldsymbol{A}_v\boldsymbol{C}_{tr} + \boldsymbol{A}_t\boldsymbol{C}_{vr} + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{tvr} + \boldsymbol{A}_{tvr}\boldsymbol{\Sigma}\} - \boldsymbol{a}_t^\top\boldsymbol{A}_r\boldsymbol{a}_v - \boldsymbol{a}_v^\top\boldsymbol{A}_t\boldsymbol{a}_r$$
$$- \boldsymbol{a}_t^\top\boldsymbol{A}_v\boldsymbol{a}_r - \boldsymbol{a}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{vr} - \boldsymbol{a}_v^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tr} - \boldsymbol{a}_r^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tv},$$

$$Q_{tv}^{(s)} = -\frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_{tv}\boldsymbol{C}_s + \boldsymbol{A}_{vs}\boldsymbol{C}_t + \boldsymbol{A}_v\boldsymbol{C}_{ts} + \boldsymbol{A}_t\boldsymbol{C}_{vs} + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{tvs} + \boldsymbol{A}_{tvs}\boldsymbol{\Sigma}\}$$
$$- \boldsymbol{a}_v^\top\boldsymbol{A}_t\boldsymbol{a}_s - \boldsymbol{a}_s^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tv} - \boldsymbol{a}_v^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{ts},$$

where $t, v = 1, 2, \ldots, p$.

Proof: Appendix. ■

We have that $\epsilon_{(p-\nu)}$ is obtained from (2) by taking into account only the nuisance parameters in $\boldsymbol{\xi}$; that is, r and s vary over all $(p-\nu)$ nuisance parameters in $\boldsymbol{\xi}$. Also, all unknown parameters in the quantities that define the Bartlett-corrected LR statistic are replaced by their restricted ML estimates; that is, the order of the approximation remains unchanged when the unknown parameters in $\Delta(\boldsymbol{\theta})$ are replaced by their restricted ML estimates. The improved LR test that uses $\omega_{\mathrm{B}} = \omega/(1+\nu^{-1}\Delta)$ as test statistic follows from the comparison of the observed value of ω_{B} with the critical value obtained as the appropriate χ^2_ν quantile. The Bartlett-corrected LR test statistic is usually expressed as $\omega/[1+\nu^{-1}\Delta]$. There is, however, another equivalent specification that delivers the same order of accuracy and has the advantage of being non-negative, that is, $\omega_{\mathrm{B}}^* = \omega\exp(-\nu^{-1}\Delta)$. The Bartlett-corrected LR statistics ω_{B} and ω_{B}^* are equivalent to order $\mathcal{O}(n^{-1})$.

Another way of improving the LR test statistic is to consider the procedure developed by Skovgaard [19]. The author generalized the results in Skovgaard [26] to improve LR testing inference in a general setting. The adjusted likelihood ratio statistics proposed by

Skovgaard [19] is given by

$$\omega_{\mathrm{S}} = \omega - 2\log(\rho),$$

where

$$\rho = \frac{(\widetilde{\boldsymbol{U}}^\top \bar{\boldsymbol{\Upsilon}}^{-1}\widehat{\boldsymbol{K}}\widehat{\boldsymbol{J}}^{-1}\bar{\boldsymbol{\Upsilon}}\widetilde{\boldsymbol{K}}^{-1}\widetilde{\boldsymbol{U}})^{\nu/2}}{\omega^{\nu/2-1}\widetilde{\boldsymbol{U}}^\top \bar{\boldsymbol{\Upsilon}}^{-1}\bar{\boldsymbol{q}}}\sqrt{\frac{\det(\widetilde{\boldsymbol{K}})\det(\widehat{\boldsymbol{K}})\det(\widetilde{\boldsymbol{J}}_{\boldsymbol{\psi}\boldsymbol{\psi}})}{\det(\bar{\boldsymbol{\Upsilon}})^2\det([\widetilde{\boldsymbol{K}}\bar{\boldsymbol{\Upsilon}}^{-1\top}\widehat{\boldsymbol{J}}\widehat{\boldsymbol{K}}^{-1}\bar{\boldsymbol{\Upsilon}}]_{\boldsymbol{\xi}\boldsymbol{\xi}})}},$$

and $\bar{\boldsymbol{q}}$ and $\bar{\boldsymbol{\Upsilon}}$ come respectively from $\boldsymbol{q} = \mathrm{E}_{\boldsymbol{\omega}_1}[\boldsymbol{U}(\boldsymbol{\omega}_1)(\ell(\boldsymbol{\omega}_1) - \ell(\boldsymbol{\omega}_2))]$ and $\boldsymbol{\Upsilon} = \mathrm{E}_{\boldsymbol{\omega}_1}[\boldsymbol{U}(\boldsymbol{\omega}_1)\boldsymbol{U}(\boldsymbol{\omega}_2)^\top]$ by inserting $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\omega}_1$ and $\widetilde{\boldsymbol{\theta}}$ for $\boldsymbol{\omega}_2$ after the expected values were computed. Also, we write $\boldsymbol{J}_{\boldsymbol{\xi}\boldsymbol{\xi}}(\boldsymbol{\theta})$ to denote the $(p-\nu)\times(p-\nu)$ observed information matrix for the nuisance vector $\boldsymbol{\xi}$. Additionally, $\widehat{\boldsymbol{K}} = \boldsymbol{K}(\widehat{\boldsymbol{\theta}})$, $\widetilde{\boldsymbol{K}} = \boldsymbol{K}(\widetilde{\boldsymbol{\theta}})$, $\widehat{\boldsymbol{J}} = \boldsymbol{J}(\widehat{\boldsymbol{\theta}})$, $\widetilde{\boldsymbol{U}} = \boldsymbol{U}(\widetilde{\boldsymbol{\theta}})$, $\widetilde{\boldsymbol{J}} = \boldsymbol{J}(\widetilde{\boldsymbol{\theta}})$ and $\widetilde{\boldsymbol{J}}_{\boldsymbol{\psi}\boldsymbol{\psi}} = \boldsymbol{J}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\widetilde{\boldsymbol{\theta}})$. After lengthy algebra, we arrive at the following proposition.

Proposition 3.2: *For the multivariate normal model with general parameterization, we have that*

$$\boldsymbol{q} = \sum_{i=1}^{n} \boldsymbol{F}_i(\boldsymbol{\omega}_1)^\top \boldsymbol{H}_i(\boldsymbol{\omega}_1)\boldsymbol{v}_i(\boldsymbol{\omega}_1;\boldsymbol{\omega}_2), \quad \boldsymbol{\Upsilon} = \sum_{i=1}^{n} \boldsymbol{F}_i(\boldsymbol{\omega}_1)^\top \boldsymbol{H}_i(\boldsymbol{\omega}_2)\boldsymbol{F}_i(\boldsymbol{\omega}_2),$$

where

$$\boldsymbol{v}_i(\boldsymbol{\omega}_1;\boldsymbol{\omega}_2) = \begin{pmatrix} \boldsymbol{\Sigma}_i(\boldsymbol{\omega}_1)\boldsymbol{\Sigma}_i(\boldsymbol{\omega}_2)^{-1}(\boldsymbol{\mu}_i(\boldsymbol{\omega}_1) - \boldsymbol{\mu}_i(\boldsymbol{\omega}_2)) \\ \mathrm{vec}(\boldsymbol{\Sigma}_i(\boldsymbol{\omega}_1)\boldsymbol{\Sigma}_i(\boldsymbol{\omega}_2)^{-1}\boldsymbol{\Sigma}_i(\boldsymbol{\omega}_1) - \boldsymbol{\Sigma}_i(\boldsymbol{\omega}_1)) \end{pmatrix}.$$

Proof: Appendix. ■

According to Skovgaard [19], the statistic $\omega_{\mathrm{S}}^* = \omega(1-\omega^{-1}\log(\rho))^2$ is asymptotically equivalent to ω_{S}. As pointed out by Skovgaard, the version ω_{S} is the one that arises naturally in the theoretical development, but the version ω_{S}^*, unlike ω_{S}, has the advantage of being non-negative. The statistics ω_{S} and ω_{S}^* follow approximately the asymptotic reference χ^2_ν distribution under the null hypothesis with high accuracy [19]. For a detailed discussion on the development of ω_{S} and ω_{S}^*, the reader is referred to Skovgaard [19, § 5.7]. Although Melo et al. [20] studied the Skovgaard correction under a general class of elliptical distributions, which includes the normal one, they only provided an algorithm to attain the required quantities. On the other hand, our Proposition 3.2 provides a closed-form expression for the ingredients of the Skovgaard correction under normality. The formulas in Proposition 3.2 have not been published elsewhere.

It is noteworthy that the general expressions which define the improved LR statistics ω_{B}, ω_{B}^*, ω_{S} and ω_{S}^* only involve operations on matrices and vectors and can be implemented in computational programming languages with support for matrix operations, such as `Ox` [24] or `R` [27], among many others. Unfortunately, the correction factors are not easy to interpret in generality and provide no indication as to what structural aspects of the model contribute significantly to their magnitude.

4. Special models

In what follows, we consider some special cases of the main model in (1) to illustrate the usefulness of the general results derived in the previous section.

4.1. Errors-in-variables model

Consider the following multivariate errors-in-variables model:

$$\boldsymbol{z}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\boldsymbol{x}_i + \boldsymbol{q}_i, \quad i = 1, 2, \dots, n, \tag{3}$$

where $\boldsymbol{z}_i$ is a $(v \times 1)$ latent response vector, $\boldsymbol{x}_i$ is a $(m \times 1)$ latent vector of covariates, $\boldsymbol{\beta}_0$ is a $(v \times 1)$ vector of intercepts, $\boldsymbol{\beta}_1$ is a $(v \times m)$ matrix which elements are inclinations and $\boldsymbol{q}_i$ is the equation error having multivariate normal distribution with zero mean and variance-covariance matrix $\boldsymbol{\Sigma}_q$. The variables $\boldsymbol{z}_i$ and $\boldsymbol{x}_i$ are not directly observed, instead surrogate variables $\boldsymbol{Z}_i$ and $\boldsymbol{X}_i$ are measured with the following additive structure $\boldsymbol{Z}_i = \boldsymbol{z}_i + \boldsymbol{\eta}_{z_i}$ and $\boldsymbol{X}_i = \boldsymbol{x}_i + \boldsymbol{\eta}_{x_i}$, respectively. The errors $\boldsymbol{\eta}_{z_i}$ and $\boldsymbol{\eta}_{x_i}$ are assumed to follow a normal distribution given by

$$\begin{pmatrix} \boldsymbol{\eta}_{z_i} \\ \boldsymbol{\eta}_{x_i} \end{pmatrix} \sim \mathcal{N}_{v+m}\left[\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\tau}_{z_i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\tau}_{x_i} \end{pmatrix}\right],$$

where the variance-covariance matrices $\boldsymbol{\tau}_{z_i}$ and $\boldsymbol{\tau}_{x_i}$ are assumed to be known for all $i = 1, 2, \dots, n$. These matrices may be attained through an analytical treatment of the data collection process, replications, or machine precision.

Notice that $\boldsymbol{Z}_i = \boldsymbol{z}_i + \boldsymbol{\eta}_{z_i}$ and $\boldsymbol{X}_i = \boldsymbol{x}_i + \boldsymbol{\eta}_{x_i}$ have equation errors for all lines, that is, $\boldsymbol{z}_i$ and $\boldsymbol{x}_i$ are not perfectly related. These equation errors are justified by the influence of other factors than $\boldsymbol{x}_i$ in the variation of $\boldsymbol{z}_i$. It is very reasonable to consider equation errors in (3) to capture extra variability, since the variances $\boldsymbol{\tau}_{z_i}$ are fixed and whether any other factor affects the variation of $\boldsymbol{z}_i$, the estimation of the line parameters will be clearly affected. Supposing that $\boldsymbol{x}_i \sim \mathcal{N}_m(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ and considering that the model errors ($\boldsymbol{q}_i$, $\boldsymbol{\eta}_{z_i}$ and $\boldsymbol{\eta}_{x_i}$) and $\boldsymbol{x}_i$ are independent, we have that the joint distribution of the observable variables can be expressed as

$$\boldsymbol{Y}_i = \begin{pmatrix} \boldsymbol{Z}_i \\ \boldsymbol{X}_i \end{pmatrix} \sim \mathcal{N}_q\left[\begin{pmatrix} \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\boldsymbol{\mu}_x \\ \boldsymbol{\mu}_x \end{pmatrix}, \begin{pmatrix} \boldsymbol{\beta}_1\boldsymbol{\Sigma}_x\boldsymbol{\beta}_1^\top + \boldsymbol{\Sigma}_q + \boldsymbol{\tau}_{z_i} & \boldsymbol{\beta}_1\boldsymbol{\Sigma}_x \\ \boldsymbol{\Sigma}_x\boldsymbol{\beta}_1^\top & \boldsymbol{\Sigma}_x + \boldsymbol{\tau}_{x_i} \end{pmatrix}\right], \tag{4}$$

where $q = v + m$. Here, the vector $\boldsymbol{\theta} = (\boldsymbol{\beta}_0^\top, \mathrm{vec}(\boldsymbol{\beta}_1)^\top, \boldsymbol{\mu}_x^\top, \mathrm{vech}(\boldsymbol{\Sigma}_x)^\top, \mathrm{vech}(\boldsymbol{\Sigma}_q)^\top)^\top$ has dimension $p = v(m+1) + m + m(m+1)/2 + v(v+1)/2$, and 'vech' transforms a symmetric matrix into a vector by stacking into columns its diagonal and superior diagonal elements. From (4), the mean vector and the variance-covariance matrix have the matrix $\boldsymbol{\beta}_1$ in common. In other words, they share mv parameters.

Next, we provide the main quantities that are used in the Fisher scoring method, Fisher information, and Bartlett correction factor derived in the previous section. First, let us define the following quantities:

$$\boldsymbol{F}_{\boldsymbol{\beta}_0}^{(r)} = \frac{\partial \boldsymbol{\beta}_0}{\partial \theta_r}, \quad \boldsymbol{F}_{\boldsymbol{\beta}_1}^{(r)} = \frac{\partial \boldsymbol{\beta}_1}{\partial \theta_r}, \quad \boldsymbol{F}_{\boldsymbol{\mu}_x}^{(r)} = \frac{\partial \boldsymbol{\mu}_x}{\partial \theta_r},$$

$$\boldsymbol{F}_{\boldsymbol{\Sigma}_x}^{(r)} = \frac{\partial \boldsymbol{\Sigma}_x}{\partial \theta_r}, \quad \boldsymbol{F}_{\boldsymbol{\Sigma}_q}^{(r)} = \frac{\partial \boldsymbol{\Sigma}_q}{\partial \theta_r},$$

where $r = 1, 2, \ldots, p$. The above quantities correspond to vectors or matrices of zeros with one in the position referring to the rth element of the parameter vector $\boldsymbol{\theta}$. Hence,

$$\boldsymbol{a}_r = \begin{cases} \mathbf{1}_n \otimes \begin{pmatrix} \boldsymbol{F}_{\beta_0}^{(r)} \\ \mathbf{0} \end{pmatrix}, & \text{if } r = 1, 2, \ldots, v; \\ \mathbf{1}_n \otimes \begin{pmatrix} \boldsymbol{F}_{\beta_1}^{(r)} \boldsymbol{\mu}_x \\ \mathbf{0} \end{pmatrix}, & \text{if } r = v+1, \ldots, v(m+1); \\ \mathbf{1}_n \otimes \begin{pmatrix} \boldsymbol{\beta}_1 \boldsymbol{F}_{\mu_x}^{(r)} \\ \boldsymbol{F}_{\mu_x}^{(r)} \end{pmatrix}, & \text{if } r = v(m+1)+1, \ldots, v(m+1)+m; \\ \mathbf{1}_n \otimes \mathbf{0}_q, & \text{if } r = (v+1)(m+1), \ldots, p, \end{cases}$$

where $\mathbf{1}_n$ denotes an n-vector of ones, and $\mathbf{0}_q$ denotes an q-vector of zeros. Also,

$$\boldsymbol{a}_{rs} = \mathbf{1}_n \otimes \begin{pmatrix} \boldsymbol{F}_{\beta_1}^{(s)} \boldsymbol{F}_{\mu_x}^{(r)} \\ \mathbf{0} \end{pmatrix}, \quad \boldsymbol{a}_{rst} = \mathbf{0}, \quad r, s, t = 1, 2, \ldots, p.$$

In addition, it follows that

$$\boldsymbol{C}_r = \begin{cases} \boldsymbol{I}_n \otimes \begin{pmatrix} \boldsymbol{F}_{\beta_1}^{(r)} \boldsymbol{\Sigma}_x \boldsymbol{\beta}_1^\top + \boldsymbol{\beta}_1 \boldsymbol{\Sigma}_x \boldsymbol{F}_{\beta_1}^{(r)\top} & \boldsymbol{F}_{\beta_1}^{(r)} \boldsymbol{\Sigma}_x \\ \boldsymbol{F}_{\beta_1}^{(r)} \boldsymbol{\Sigma}_x & \mathbf{0} \end{pmatrix}, & \text{if } r = v+1, \ldots, v(m+1); \\ \boldsymbol{I}_n \otimes \begin{pmatrix} \boldsymbol{\beta}_1 \boldsymbol{F}_{\Sigma_x}^{(r)} \boldsymbol{\beta}_1^\top & \boldsymbol{\beta}_1 \boldsymbol{F}_{\Sigma_x}^{(r)} \\ \boldsymbol{F}_{\Sigma_x}^{(r)} \boldsymbol{\beta}_1^\top & \mathbf{0} \end{pmatrix}, & \text{if } r = (v+1)(m+1), \ldots, p'; \\ \boldsymbol{I}_n \otimes \begin{pmatrix} \boldsymbol{F}_{\Sigma_q}^{(r)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, & \text{if } r = p'+1, \ldots, p; \\ \boldsymbol{I}_n \otimes \mathbf{0}_{q \times q}, & \text{otherwise,} \end{cases}$$

where $p' = v(m+1) + m + m(m+1)/2$, $\mathbf{0}_{q\times q}$ denotes an $q \times q$ matrix of zeros, and $\boldsymbol{I}_n$ denotes an $n \times n$ identity matrix. Additionally, we have

$$\boldsymbol{C}_{rs} = \begin{cases} \boldsymbol{I}_n \otimes \begin{pmatrix} \boldsymbol{F}_{\beta_1}^{(s)} \boldsymbol{\Sigma}_x \boldsymbol{F}_{\beta_1}^{(r)\top} + \boldsymbol{F}_{\beta_1}^{(r)} \boldsymbol{\Sigma}_x \boldsymbol{F}_{\beta_1}^{(s)\top} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, & \text{if } r = v+1, \ldots, v(m+1); \\ \boldsymbol{I}_n \otimes \begin{pmatrix} \boldsymbol{F}_{\beta_1}^{(u)} \boldsymbol{F}_{\Sigma_x}^{(t)} \boldsymbol{\beta}_1^\top + \boldsymbol{\beta}_1 \boldsymbol{F}_{\Sigma_x}^{(t)} \boldsymbol{F}_{\beta_1}^{(s)\top} & \boldsymbol{F}_{\beta_1}^{(u)} \boldsymbol{F}_{\Sigma_x}^{(t)} \\ \boldsymbol{F}_{\Sigma_x}^{(t)} \boldsymbol{F}_{\beta_1}^{(u)\top} & \mathbf{0} \end{pmatrix}, & \text{if } r = (v+1)(m+1), \ldots, p'; \\ \boldsymbol{I}_n \otimes \mathbf{0}_{q \times q}, & \text{otherwise,} \end{cases}$$

where $s = v+1, \ldots, v(m+1)$. Finally, we have that

$$
\boldsymbol{C}_{rst} = \begin{cases}
\boldsymbol{I}_n \otimes \begin{pmatrix} \boldsymbol{F}^{(s)}_{\boldsymbol{\beta}_1}\boldsymbol{F}^{(t)}_{\boldsymbol{\Sigma}_x}\boldsymbol{F}^{(r)\top}_{\boldsymbol{\beta}_1} + \boldsymbol{F}^{(r)}_{\boldsymbol{\beta}_1}\boldsymbol{F}^{(t)}_{\boldsymbol{\Sigma}_x}\boldsymbol{F}^{(s)\top}_{\boldsymbol{\beta}_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, & \text{if } r, s = v+1, \ldots, v(m+1), \\
& t = (v+1)(m+1), \ldots, p'; \\
\boldsymbol{I}_n \otimes \begin{pmatrix} \boldsymbol{F}^{(t)}_{\boldsymbol{\beta}_1}\boldsymbol{F}^{(s)}_{\boldsymbol{\Sigma}_x}\boldsymbol{F}^{(r)\top}_{\boldsymbol{\beta}_1} + \boldsymbol{F}^{(r)}_{\boldsymbol{\beta}_1}\boldsymbol{F}^{(s)}_{\boldsymbol{\Sigma}_x}\boldsymbol{F}^{(t)\top}_{\boldsymbol{\beta}_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, & \text{if } r, t = v+1, \ldots, v(m+1), \\
& s = (v+1)(m+1), \ldots, p'; \\
\boldsymbol{I}_n \otimes \begin{pmatrix} \boldsymbol{F}^{(t)}_{\boldsymbol{\beta}_1}\boldsymbol{F}^{(r)}_{\boldsymbol{\Sigma}_x}\boldsymbol{F}^{(s)\top}_{\boldsymbol{\beta}_1} + \boldsymbol{F}^{(s)}_{\boldsymbol{\beta}_1}\boldsymbol{F}^{(r)}_{\boldsymbol{\Sigma}_x}\boldsymbol{F}^{(t)\top}_{\boldsymbol{\beta}_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, & \text{if } s, t = v+1, \ldots, v(m+1), \\
& r = (v+1)(m+1), \ldots, p'; \\
\boldsymbol{I}_n \otimes \boldsymbol{0}_{q\times q}, & \text{otherwise.}
\end{cases}
$$

Therefore, from the above quantities, the Bartlet correction factor can be obtained from the general result in Proposition 3.1. These quantities can be consistently estimated under the null hypothesis. Regarding the Skovgaard adjustment, it follows that

$$
\bar{\boldsymbol{q}} = \sum_{i=1}^{n} \begin{pmatrix} \widehat{\boldsymbol{D}}_i^{(1)\top}\widetilde{\boldsymbol{\Sigma}}_i^{-1}(\widehat{\boldsymbol{\mu}}_i - \widetilde{\boldsymbol{\mu}}_i) + \frac{1}{2}\widehat{\boldsymbol{V}}_i^{(1)\top}(\widehat{\boldsymbol{\Sigma}}_i \otimes \widehat{\boldsymbol{\Sigma}}_i)^{-1}\mathrm{vec}\big(\widehat{\boldsymbol{\Sigma}}_i\widetilde{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{\Sigma}}_i - \widehat{\boldsymbol{\Sigma}}_i\big) \\ \frac{1}{2}\widehat{\boldsymbol{V}}_i^{(2)\top}(\widehat{\boldsymbol{\Sigma}}_i \otimes \widehat{\boldsymbol{\Sigma}}_i)^{-1}\mathrm{vec}\big(\widehat{\boldsymbol{\Sigma}}_i\widetilde{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{\Sigma}}_i - \widehat{\boldsymbol{\Sigma}}_i\big) \end{pmatrix},
$$

and

$$
\bar{\boldsymbol{\Upsilon}} = \sum_{i=1}^{n} \begin{bmatrix} \widehat{\boldsymbol{D}}_i^{(1)\top}\widetilde{\boldsymbol{\Sigma}}_i^{-1}\widetilde{\boldsymbol{D}}_i^{(1)} + \frac{1}{2}\widehat{\boldsymbol{V}}_i^{(1)\top}(\widetilde{\boldsymbol{\Sigma}}_i \otimes \widetilde{\boldsymbol{\Sigma}}_i)^{-1}\widetilde{\boldsymbol{V}}_i^{(1)} & \frac{1}{2}\widehat{\boldsymbol{V}}_i^{(1)\top}(\widetilde{\boldsymbol{\Sigma}}_i \otimes \widetilde{\boldsymbol{\Sigma}}_i)^{-1}\widetilde{\boldsymbol{V}}_i^{(2)} \\ \frac{1}{2}\widehat{\boldsymbol{V}}_i^{(2)\top}(\widetilde{\boldsymbol{\Sigma}}_i \otimes \widetilde{\boldsymbol{\Sigma}}_i)^{-1}\widetilde{\boldsymbol{V}}_i^{(1)} & \frac{1}{2}\widehat{\boldsymbol{V}}_i^{(2)\top}(\widetilde{\boldsymbol{\Sigma}}_i \otimes \widetilde{\boldsymbol{\Sigma}}_i)^{-1}\widetilde{\boldsymbol{V}}_i^{(2)} \end{bmatrix},
$$

where the quantities distinguished by the addition of '$\widehat{\ }$' and '$\widetilde{\ }$' are evaluated at $\widehat{\boldsymbol{\theta}}$ (unrestricted estimates) and $\widetilde{\boldsymbol{\theta}}$ (restricted estimates), respectively. Also,

$$
\boldsymbol{D}_i^{(1)} = \frac{\partial \boldsymbol{\mu}_i}{\partial(\boldsymbol{\beta}_0^\top, \ \mathrm{vec}(\boldsymbol{\beta})^\top, \ \ \boldsymbol{\mu}_x^\top)}, \quad \boldsymbol{V}_i^{(1)} = \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_i)}{\partial(\boldsymbol{\beta}_0^\top, \ \ \mathrm{vec}(\boldsymbol{\beta})^\top, \ \boldsymbol{\mu}_x^\top)} = \left[\boldsymbol{0}, \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_i)}{\partial \mathrm{vec}(\boldsymbol{\beta})^\top}, \boldsymbol{0}\right],
$$

and

$$
\boldsymbol{V}_i^{(2)} = \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_i)}{\partial(\mathrm{vech}(\boldsymbol{\Sigma}_x)^\top, \ \mathrm{vech}(\boldsymbol{\Sigma}_q)^\top)}.
$$

4.2. Mixed-effects model

Consider the following mixed-effects model:

$$
\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \quad i = 1, 2, \ldots, n,
$$

where $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iq_i})^\top$ is a $(q_i \times 1)$ vector of responses of the ith experimental unit, $\boldsymbol{\beta}$ is a $(v \times 1)$ parameter vector (fixed effects), $\boldsymbol{X}_i$ is a $(q_i \times v)$ fixed effects specification matrix, $\boldsymbol{b}_i$ is a $(m \times 1)$ random effects vector, $\boldsymbol{Z}_i$ is a $(q_i \times m)$ random effects specification matrix, and $\boldsymbol{e}_i = (e_{i1}, e_{i2}, \ldots, e_{iq_i})^\top$ is a $(q_i \times 1)$ vector of random errors. Suppose that $\boldsymbol{e}_i \sim \mathcal{N}_{q_i}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{q_i})$ and $\boldsymbol{b}_i \sim \mathcal{N}_m(\boldsymbol{0}, \boldsymbol{G})$, where $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n, \boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ are independent,

$\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{\varrho})$ is a $(m \times m)$ positive definite matrix, and $\boldsymbol{\varrho}$ is a m^*-vector of unknown parameters with $m^* = m(m+1)/2$. In matrix form, we have that $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{e}$, where $\boldsymbol{Y} = (\boldsymbol{y}_1^\top, \boldsymbol{y}_2^\top, \ldots, \boldsymbol{y}_n^\top)^\top$ is an T-vector with $T = \sum_{i=1}^N q_i$, $\boldsymbol{X} = [\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top, \ldots, \boldsymbol{X}_n^\top]^\top$ is an $T \times v$ matrix, $\boldsymbol{Z} = \mathrm{diag}\{\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n\}$ is an $T \times nm$ matrix, $\boldsymbol{b} = (\boldsymbol{b}_1^\top, \boldsymbol{b}_2^\top, \ldots, \boldsymbol{b}_n^\top)^\top$ is an nm-vector, and $\boldsymbol{e} = (\boldsymbol{e}_1^\top, \boldsymbol{e}_2^\top, \ldots, \boldsymbol{e}_n^\top)^\top$ is an T-vector. Thus, $\boldsymbol{b} \sim \mathcal{N}_{nm}(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{G})$. In short, we have that $\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$, where $\boldsymbol{\mu} := \boldsymbol{\mu}(\boldsymbol{\theta}) = \boldsymbol{X}\boldsymbol{\beta}$, $\boldsymbol{\epsilon} = \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{e}$, and $\boldsymbol{\epsilon} \sim \mathcal{N}_T(\boldsymbol{0}, \boldsymbol{\Sigma})$, being $\boldsymbol{\Sigma} := \boldsymbol{\Sigma}(\boldsymbol{\gamma}) = \boldsymbol{Z}(\boldsymbol{I}_n \otimes \boldsymbol{G})\boldsymbol{Z}^\top + \sigma^2\boldsymbol{I}_T$ and $\boldsymbol{\gamma} = (\boldsymbol{\varrho}^\top, \sigma^2)^\top$ an (m^*+1)-vector of unknown parameters. Here, the vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ has dimension $p = v + m^* + 1$.

The main quantities used to obtain the Bartlett correction factor are as follows. After some algebra, we have that $\boldsymbol{a}_r = \boldsymbol{X}\boldsymbol{F}_{\boldsymbol{\beta}}^{(r)}$ for $r = 1, 2, \ldots, v$, and $\boldsymbol{a}_r = \boldsymbol{0}$ for $r = v+1, \ldots, p$, where $\boldsymbol{F}_{\boldsymbol{\beta}}^{(r)} = \partial\boldsymbol{\beta}/\partial\theta_r$ corresponds to a null vector except for the rth position $(r = 1, \ldots, v)$ which is equal to one. We have also that $\boldsymbol{a}_{rs} = \boldsymbol{a}_{rst} = \boldsymbol{0}$ for all $r, s, t = 1, 2, \ldots, p$. In addition, the matrices $\boldsymbol{C}_r$, $\boldsymbol{C}_{rs}$, and $\boldsymbol{C}_{rst}$ are null except for $r, s, t = v+1, \ldots, p-1$ which are given by $\boldsymbol{C}_r = \boldsymbol{Z}(\boldsymbol{I}_n \otimes \boldsymbol{G}_r)\boldsymbol{Z}^\top$, $\boldsymbol{C}_{rs} = \boldsymbol{Z}(\boldsymbol{I}_n \otimes \boldsymbol{G}_{rs})\boldsymbol{Z}^\top$, and $\boldsymbol{C}_{rst} = \boldsymbol{Z}(\boldsymbol{I}_n \otimes \boldsymbol{G}_{rst})\boldsymbol{Z}^\top$, where $\boldsymbol{G}_r = \partial\boldsymbol{G}/\partial\theta_r$, $\boldsymbol{G}_{rs} = \partial^2\boldsymbol{G}/\partial\theta_r\partial\theta_s$, and $\boldsymbol{G}_{rst} = \partial^3\boldsymbol{G}/\partial\theta_r\partial\theta_s\partial\theta_t$. Finally, it follows that $\boldsymbol{C}_p = \boldsymbol{I}_T$ and $\boldsymbol{C}_{pp} = \boldsymbol{C}_{ppp} = \boldsymbol{0}_{T\times T}$. Hence, the Bartlet correction factor can be obtained from the general result in Proposition 3.1. The above quantities can be consistently estimated under the null hypothesis. The quantities related to the Skovgaard adjustment are given by

$$\bar{\boldsymbol{q}} = \sum_{i=1}^{n} \begin{pmatrix} \widehat{\boldsymbol{D}}_i^{(1)\top}\widetilde{\boldsymbol{\Sigma}}_i^{-1}(\widehat{\boldsymbol{\mu}}_i - \widetilde{\boldsymbol{\mu}}_i) \\ \frac{1}{2}\widehat{\boldsymbol{V}}_i^{(2)\top}(\widehat{\boldsymbol{\Sigma}}_i \otimes \widehat{\boldsymbol{\Sigma}}_i)^{-1}\mathrm{vec}\big(\widehat{\boldsymbol{\Sigma}}_i\widetilde{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{\Sigma}}_i - \widehat{\boldsymbol{\Sigma}}_i\big) \\ \frac{1}{2}\widehat{\boldsymbol{V}}_i^{(3)\top}(\widehat{\boldsymbol{\Sigma}}_i \otimes \widehat{\boldsymbol{\Sigma}}_i)^{-1}\mathrm{vec}\big(\widehat{\boldsymbol{\Sigma}}_i\widetilde{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{\Sigma}}_i - \widehat{\boldsymbol{\Sigma}}_i\big) \end{pmatrix},$$

and

$$\bar{\boldsymbol{\Upsilon}} = \sum_{i=1}^{n} \begin{bmatrix} \widehat{\boldsymbol{D}}_i^{(1)\top}\widetilde{\boldsymbol{\Sigma}}_i^{-1}\widetilde{\boldsymbol{D}}_i^{(1)} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{2}\widehat{\boldsymbol{V}}_i^{(2)\top}(\widetilde{\boldsymbol{\Sigma}}_i \otimes \widetilde{\boldsymbol{\Sigma}}_i)^{-1}\widetilde{\boldsymbol{V}}_i^{(2)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \frac{1}{2}\widehat{\boldsymbol{V}}_i^{(3)\top}(\widetilde{\boldsymbol{\Sigma}}_i \otimes \widetilde{\boldsymbol{\Sigma}}_i)^{-1}\widetilde{\boldsymbol{V}}_i^{(3)} \end{bmatrix},$$

where $\boldsymbol{D}_i^{(1)} = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}^\top = \boldsymbol{F}_{\boldsymbol{\beta}}^{(r)}$, $\boldsymbol{V}_i^{(2)} = (\boldsymbol{Z}_i \otimes \boldsymbol{Z}_i)\partial\mathrm{vec}(\boldsymbol{I}_n \otimes \mathbf{G})/\partial\boldsymbol{\rho}^\top$ and $\boldsymbol{V}_i^{(3)} = \partial\mathrm{vec}(\sigma^2\boldsymbol{I}_T)/\partial\sigma^2 = \boldsymbol{1}_T$. Finally, quantities distinguished by the addition of '$\widehat{\ }$' and '$\widetilde{\ }$' are evaluated at $\widehat{\boldsymbol{\theta}}$ (unrestricted estimates) and $\widetilde{\boldsymbol{\theta}}$ (restricted estimates), respectively.

4.3. Nonlinear model

Consider the following univariate nonlinear model:

$$Y_i = f(\boldsymbol{x}_i, \boldsymbol{\beta}) + \epsilon_i, \quad i = 1, 2, \ldots, n,$$

where Y_i is the response variable, $\boldsymbol{x}_i^\top = (x_{i1}, \ldots, x_{iq})$ are observations on q known regressors, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathrm{R}^p$ is a p-vector of regression parameters $(q \le p < n)$, and f is a nonlinear function of $\boldsymbol{\beta}$. Assume that $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ for $i = 1, 2, \ldots, n$. The univariate nonlinear model above defined is clearly a special case of (1) with $q_i = 1$. Consequently, we have that $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$ is a $(p+1)$-vector, $\mu_i = f(\boldsymbol{x}_i, \boldsymbol{\beta})$ and $\Sigma_i = \sigma^2$.

The quantities related to the Bartlett correction factor are as follows. The ith element of the n-dimensional vector $\boldsymbol{a}_r$ is given by $\partial f(x_i, \boldsymbol{\beta})/\partial\theta_r$ for $r = 1, 2, \ldots, p$, and zero for $r = p + 1$. Note that the ith elements of n-dimensional vectors $\boldsymbol{a}_{rs}$ and $\boldsymbol{a}_{rst}$ are given, respectively, by $\partial^2 f(x_i, \boldsymbol{\beta})/\partial\theta_r\partial\theta_s$ and $\partial^3 f(x_i, \boldsymbol{\beta})/\partial\theta_r\partial\theta_s\partial\theta_t$ for $r, s, t = 1, 2, \ldots, p$, and zero for $r, s, t = p + 1$. Also, $\boldsymbol{C}_r = \boldsymbol{0}$ for $r = 1, 2, \ldots, p$, and $\boldsymbol{C}_r = \boldsymbol{I}_p$ for $r = p + 1$. Finally, $\boldsymbol{C}_{rs} = \boldsymbol{C}_{rst} = \boldsymbol{0}$ for $r, s, t = 1, 2, \ldots, p + 1$. Thus, the Bartlet correction factor can be obtained from the general result in Proposition 3.1. All the above quantities can be consistently estimated under the null hypothesis. Regarding the Skovgaard adjustment, it follows that

$$\bar{\boldsymbol{q}} = \sum_{i=1}^{n} \begin{pmatrix} \frac{1}{\widehat{\sigma}^2} \widehat{\boldsymbol{D}}_i^{(1)\top} (\widehat{\mu}_i - \widetilde{\mu}_i) \\ \frac{1}{2}\left(\frac{1}{\widetilde{\sigma}^2} - \frac{1}{\widehat{\sigma}^2}\right) \end{pmatrix},$$

and

$$\bar{\boldsymbol{\Upsilon}} = \sum_{i=1}^{n} \begin{bmatrix} \frac{1}{\widehat{\sigma}^2} \widehat{\boldsymbol{D}}_i^{(1)\top} \widetilde{\boldsymbol{D}}_i^{(1)} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{2\widehat{\sigma}^4} \end{bmatrix},$$

where $\boldsymbol{D}_i^{(1)} = \partial\mu_i/\partial\boldsymbol{\beta}^\top$. The addition of '$\widehat{\ }$' and '$\widetilde{\ }$' means quantities evaluated at $\widehat{\boldsymbol{\theta}}$ (unrestricted estimates) and $\widetilde{\boldsymbol{\theta}}$ (restricted estimates), respectively.

5. Finite-sample performance

In this section, we shall report Monte Carlo simulation experiments to illustrate the performance of the usual LR test which uses the usual statistic ω, and the improved LR tests that use the corrected statistics ω_{B}, ω_{B}^*, ω_{S}, and ω_{S}^* in small and moderate-sized samples. The number of Monte Carlo replications was 15,000, and the nominal levels of the tests were $\alpha = 10\%, 5\%$ and 1%. The simulations were carried out using the matrix programming language Ox [24], which is freely distributed for academic purposes and available at http://www.doornik.com. We report the null rejection rates of the null hypothesis of all LR tests at the 10%, 5%, and 1% nominal significance levels; that is, the percentage of times that the corresponding observed value of the statistic exceeds the 10%, 5% and 1% upper points of the reference χ^2 distribution. These rates estimate the type I error probability of the tests. The Monte Carlo simulation results are listed in Tables 1–3 whose entries are percentages. As will see in what follows, the usual LR test can be quite size-distorted when the sample size is small, and so the improved LR tests become good alternatives in such a case.

First, we consider a particular case of the errors-in-variables model described in Section 4.1. It is given by

$$Z_i = \beta_0 + \beta_1 x_i + \eta_{z_i} \quad \text{and} \quad X_i = x_i + \eta_{x_i}, \tag{5}$$

where

$$\begin{pmatrix} x_i \\ \eta_{z_i} \\ \eta_{x_i} \end{pmatrix} \sim \mathcal{N}_3 \left[\begin{pmatrix} \mu_x \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_x & 0 & 0 \\ 0 & \Sigma_e & 0 \\ 0 & 0 & \Sigma_u \end{pmatrix} \right],$$

Table 1. Null rejection rates (%) under the errors-in-variables model for the usual LR statistic ω, and its corrected versions, namely, ω_{B}, ω_{B}^*, ω_{S}, and ω_{S}^*.

Statistic	$n=15$	$n=25$	$n=35$	$n=45$	$n=55$
			$\alpha=10\%$		
ω	12.4	11.9	11.5	11.0	10.6
ω_{B}	11.0	11.2	10.9	10.5	10.4
ω_{B}^*	10.9	11.1	10.9	10.5	10.4
ω_{S}	9.3	10.4	10.3	9.9	10.0
ω_{S}^*	9.4	10.4	10.3	9.9	9.9
			$\alpha=5\%$		
ω	6.4	5.8	6.1	5.9	5.5
ω_{B}	5.6	5.3	5.8	5.6	5.4
ω_{B}^*	5.5	5.3	5.8	5.6	5.4
ω_{S}	4.5	4.9	5.3	5.3	5.1
ω_{S}^*	4.4	4.9	5.3	5.3	5.1
			$\alpha=1\%$		
ω	1.3	1.4	1.4	1.4	1.0
ω_{B}	1.1	1.2	1.2	1.3	1.0
ω_{B}^*	1.1	1.2	1.2	1.3	1.0
ω_{S}	0.8	1.1	1.1	1.1	0.9
ω_{S}^*	0.8	1.1	1.0	1.1	0.9

Table 2. Null rejection rates (%) under the mixed-effect model for the usual LR statistic ω, and its corrected versions, namely, ω_{B}, ω_{B}^*, ω_{S}, and ω_{S}^*.

Statistic	$n=15$	$n=25$	$n=35$
		$\alpha=10\%$	
ω	18.5	14.9	13.1
ω_{B}	12.8	11.9	11.0
ω_{B}^*	12.3	11.3	10.9
ω_{S}	11.9	12.8	11.3
ω_{S}^*	10.1	12.1	11.0
		$\alpha=5\%$	
ω	10.3	8.1	7.0
ω_{B}	6.6	5.9	5.7
ω_{B}^*	6.5	5.8	5.7
ω_{S}	5.2	5.1	5.1
ω_{S}^*	4.5	4.7	5.1
		$\alpha=1\%$	
ω	2.7	1.3	1.6
ω_{B}	1.6	0.5	1.1
ω_{B}^*	1.3	0.5	1.1
ω_{S}	0.6	0.3	1.2
ω_{S}^*	0.5	0.3	1.1

and so

$$\boldsymbol{Y}_i = \begin{pmatrix} Z_i \\ X_i \end{pmatrix} \sim \mathcal{N}_2\left[\begin{pmatrix} \beta_0 + \beta_1\mu_x \\ \mu_x \end{pmatrix}, \begin{pmatrix} \beta_1^2\Sigma_x + \Sigma_e & \beta_1\Sigma_x \\ \Sigma_x\beta_1 & \Sigma_x + \Sigma_u \end{pmatrix}\right].$$

To ensure the identifiability of the above model, we assume that $\lambda = \Sigma_e\Sigma_u^{-1}$ is known. In particular, we assume that $\Sigma_e = \Sigma_u$. The parameter vector becomes $\boldsymbol{\theta} = (\beta_0, \beta_1, \mu_x, \Sigma_x, \Sigma_u)^\top$. The true parameter values we consider are $\beta_0 = -3.5$, $\beta_1 = 1.0$, $\mu_x = 20$, $\Sigma_x = 15$ and $\Sigma_u = 2.5$. The null hypothesis is $\mathcal{H}_0 : \beta_1 = 1.0$, which is tested against $\mathcal{H}_1 : \beta_1 \neq 1.0$. The rejection rates of $\mathcal{H}_0$ are listed in Table 1. Notice that the usual LR test is slightly more size-distorted than the corrected LR tests, mainly when the sample

Table 3. Null rejection rates (%) under the nonlinear regression model for the usual LR statistic ω, and its corrected versions, namely, ω_B, ω_B^*, ω_S, and ω_S^*.

Statistic	$n=7$	$n=15$	$n=25$	$n=35$	$n=45$	$n=55$	$n=100$
				$\alpha=10\%$			
ω	34.4	18.0	13.7	12.4	12.2	11.5	10.2
ω_B	22.5	13.5	10.8	10.6	10.6	10.6	10.1
ω_B^*	20.2	13.0	10.6	10.5	10.6	10.6	10.2
ω_S	13.2	11.5	9.9	10.2	10.2	10.3	10.0
ω_S^*	8.3	10.9	9.8	10.1	10.2	10.2	9.9
				$\alpha=5\%$			
ω	24.8	11.0	7.4	6.5	6.2	6.1	5.3
ω_B	14.7	7.4	5.5	5.1	5.4	5.4	5.0
ω_B^*	12.5	7.0	5.3	5.0	5.4	5.4	5.1
ω_S	6.4	5.9	5.0	4.6	5.2	5.1	5.1
ω_S^*	4.0	5.2	4.9	4.5	5.1	5.0	5.0
				$\alpha=1\%$			
ω	12.0	2.9	1.6	1.3	1.6	1.4	1.1
ω_B	5.0	1.4	1.0	0.9	1.1	0.9	1.0
ω_B^*	3.8	1.3	1.0	0.9	1.1	0.9	1.1
ω_S	0.4	1.0	0.8	0.8	1.0	0.9	1.0
ω_S^*	0.5	0.8	0.8	0.8	0.9	0.9	1.0

size is small. In addition, the size of all tests tends to the significance level as the sample size increases, as expected.

Next, we consider the mixed-effects model given by

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 x_{ij} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}, \qquad (6)$$

where $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, 6$. The values of t_{ij} are $t_{i1} = 0$, $t_{i2} = 0.5$, $t_{i3} = 1$, $t_{i4} = 1.5$, $t_{i5} = 2$ and $t_{i6} = 3$, and x_{ij} is a dummy variable. Also, $\boldsymbol{b}_i = (b_{0i} b_{1i})^\top \sim \mathcal{N}_2(\mathbf{0}, \mathbf{G})$, where

$$\mathbf{G} = \begin{bmatrix} \rho_1 & \rho_2 \\ \rho_2 & \rho_3 \end{bmatrix}.$$

Additionally, $\boldsymbol{\epsilon}_i$ and $\boldsymbol{b}_i$ are assumed to be independent, where $\boldsymbol{\epsilon}_i \sim \mathcal{N}_6(0, \sigma^2 \boldsymbol{I}_6)$. The true parameter values are $\beta_0 = 3.5$, $\beta_1 = -0.3$, $\beta_2 = 0$, $\rho_1 = 0.35$, $\rho_2 = -0.04$, $\rho_3 = 0.02$, and $\sigma^2 = 0.3$. Here, the null hypothesis of interest is $\mathcal{H}_0 : \beta_2 = 0$, which is tested against $\mathcal{H}_1 : \beta_2 \neq 0$. Table 2 lists the rejection rates of $\mathcal{H}_0$. From this table, note that the usual LR test is quite liberal (over-rejecting the null hypothesis more frequently than expected based on the selected nominal level). For example, for $\alpha = 5\%$ and $n = 15$, the null rejection rate of the usual LR test was 10.3%, more than twice the nominal significance level. On the other hand, the null rejection rates of the improved LR tests are close to the nominal significance levels. As expected, the null rejection rates of all tests tend to the significance level as the sample size increases.

Now, we consider the nonlinear regression model given by

$$y_i = \theta_1 + \frac{\theta_2 - \theta_1}{1 + \exp\left(-\frac{x_i - \theta_3}{\theta_4}\right)} + \epsilon_i, \quad i = 1, 2, \ldots, n, \qquad (7)$$

where $\epsilon_i \sim \mathcal{N}(0; \sigma^2)$, and $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \sigma^2)^\top$. We set $\theta_1 = 6.0$, $\theta_2 = 0$, $\theta_3 = 0.08$, $\theta_4 = 0.03$, and $\sigma^2 = 0.02$. Here, the covariate values were taken as random draws of the

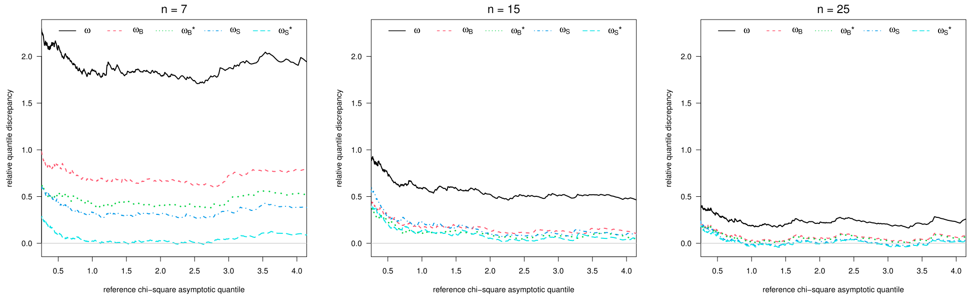

Figure 1. Quantile relative discrepancies.

standard normal distribution. Suppose we want to test the null hypotheses $\mathcal{H}_0 : \theta_2 = 0$ against the alternative hypothesis $\mathcal{H}_1 : \theta_2 \neq 0$. The rejection rates of $\mathcal{H}_0$ are listed in Table 3. From the numerical results, it is evident that the usual LR test is markedly liberal (over-rejecting the null hypothesis much more frequently than expected based on the selected nominal level). For example, for $n = 7$ and $\alpha = 10\%$, the null rejection rate of the LR test was 34.4%, more than threewice the nominal significance level. Again, the improved LR tests present null rejection rates closer to the significance level than the usual LR test. However, it is worth stressing that the improved LR tests based on the corrected statistics ω_S and ω_S^* present the best performances. Notice that the null rejection rates of all the tests approach the corresponding nominal levels as the sample size grows, as expected. Figure 1 presents curves of quantile relative discrepancies versus the correspondent asymptotic quantiles for the statistics ω, ω_B, ω_B^*, ω_S and ω_S^*. Relative quantile discrepancy is defined as the difference between exact (estimated by simulation) and asymptotic quantiles divided by the latter. The closer to zero the relative quantile discrepancy, the better is the approximation of the exact null distribution of the test statistic by the limiting chi-square distribution. Figure 1 reveals that the distribution of the usual LR statistic ω is poorly approximated by the reference χ_1^2 distribution when the sample size is small. On the other hand, the distributions of the corrected statistics ω_B, ω_B^*, ω_S and ω_S^* are closer to the reference χ_1^2 distribution than the original LR statistic. As expected, the relative quantile discrepancy of all test statistics decreases when the sample size increases. Finally, we turn to a brief study of finite-sample power properties of the all tests. As the simulation results in Table 3 show, the tests have different sizes when one uses their asymptotic χ^2 distribution in small- and moderate-sized samples. In evaluating the power of these tests, it is important to ensure that they all have the correct size under the null hypothesis. To overcome this difficulty, we used 250,000 Monte Carlo simulated samples, drawn under the null hypothesis, to estimate the exact critical value of each test for the chosen nominal level. We set $n = 80$ and $\gamma = 5\%$. For the power simulations, we computed the rejection rates under the alternative hypothesis $\beta_2 = \delta$, for δ ranging from -2.0 to 2.0. Figure 2 shows that all tests have similar power behaviour. As expected, the powers of the tests approach 1 as $|\delta|$ grows.

Overall, in small to moderate-sized samples, the best-performing tests are the improved LR tests based on the adjusted LR statistics, since they are less size-distorted than the usual LR test. Hence, these tests may be recommended for testing hypotheses on the parameters in the above class of models, which are special cases of the multivariate regression model

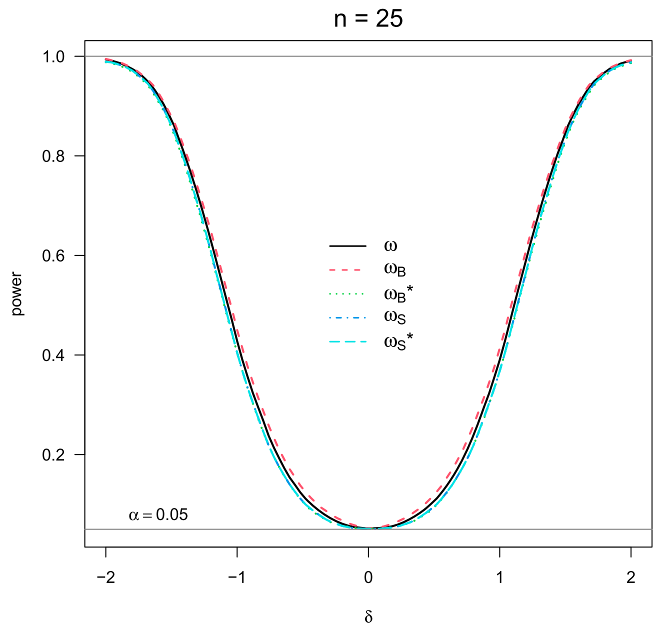

Figure 2. Power of all tests.

with general parameterization studied in this paper. Among the improved tests, the LR test on the basis of the adjusted statistics ω_S and ω_S^* presented the best results. It is evident that much more special cases of the multivariate regression model with general parameterization could be considered, but the main aim here is to show that the correction factors (Bartlett and Skovgaard) derived for this class of multivariate models are always in the right direction and provide substantial improvements on the usual LR test.

6. Real data examples

In the following, we consider real data applications to illustrate the use of the improved LR tests in practice. All computations were performed using the `Ox` program. The ML estimates of the model parameters are obtained by maximizing the log-likelihood function using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton non-linear optimization algorithm. According to Mittelhammer et al. [28, p.199], the BFGS method is generally considered the best non-linear optimization procedure. The `Ox` codes used in the real data applications can be found in the Supplementary Material.

6.1. Premature babies data

These real data were obtained from Kelly [29]. The observations are measures of serum kanamycin levels in blood samples from 20 premature babies. One of the measurements was obtained by the heelstick method (X), and the other using an umbilical catheter (Z). Since both methods are subject to measurement errors, we can consider the errors-in-variables model given in (5) to model these data. Following Kelly (29), we assume $\Sigma_e = \Sigma_u$. The model parameters are β_0, β_1, μ_x, Σ_x and Σ_u, and the corresponding ML estimates are $\widehat{\beta}_0 = -1.1601$, $\widehat{\beta}_1 = 1.0698$, $\widehat{\mu}_x = 20.855$, $\widehat{\Sigma}_x = 20.352$ and $\widehat{\Sigma}_u = 4.3737$. Suppose we want to test the null hypothesis $\mathcal{H}_0 : \beta_1 = 1$ against the alternative hypothesis $\mathcal{H}_1 : \beta_1 \neq 1$. The observed values of the LR statistics (and p-values between parentheses) are given by $\omega = 0.2046$ (p-value: 0.6510), $\omega_{\rm B} = 0.1810$ (p-value: 0.6705), $\omega^*_{\rm B} = 0.1796$ (p-value: 0.6717), $\omega_{\rm S} = 0.1793$ (p-value: 0.6715), and $\omega^*_{\rm S} = 0.1790$ (p-value: 0.6722). It is evident that all tests lead to the non-rejection of the null hypothesis at the 10% nominal level. However, the results could be different if the sample size were smaller. To illustrate this, a subset of the real data set with 15 premature babies was chosen randomly. In this case, the ML estimates become $\widehat{\beta}_0 = -4.0014$, $\widehat{\beta}_1 = 1.2510$, $\widehat{\mu}_x = 20.8270$, $\widehat{\Sigma}_x = 16.7540$ and $\widehat{\Sigma}_u = 2.3838$. Now, to test $\mathcal{H}_0 : \beta_1 = 1$, we obtain the observed values $\omega = 2.8249$ (p-value: 0.0928), $\omega_{\rm B} = 2.4190$ (p-value: 0.1199), $\omega^*_{\rm B} = 2.3885$ (p-value: 0.1222), $\omega_{\rm S} = 2.3751$ (p-value: 0.1233), $\omega^*_{\rm S} = 2.3556$ (p-value: 0.1248). Hence, at the 10% significance level, the adjusted LR tests do not reject the null hypothesis unlike the unmodified LR test, which is oversized as evidenced by the simulation results. Note also that the conclusion obtained using the corrected LR tests is compatible with that reached based on the complete data set.

6.2. *Plasma inorganic phosphate data*

These data were obtained from Davis [30, Table 3.10]. The real data refer to plasma inorganic phosphate measurements obtained from 33 patients (with 13 control (group 1) and 20 obese (group 2)) in 0.0, 0.5, 1.0, 1.5, 2.0, and 3.0 hours after an oral glucose challenge. The goal here is to investigate the effect of the group on the plasma inorganic phosphate. We consider the mixed-effect model given in (6) to model these data, where y_{ij} is the plasma inorganic phosphate measurements of the ith individual in the jth instant of time, t_{ij} is the jth instant of time (in hours) in which the plasma inorganic phosphate measurements of the ith individual, and x_i is a dummy variable that receives the value 1 if the ith individual belongs to group 2 (obese) and 0 if the ith individual belongs to group 1 (control), for $i = 1, 2, \ldots, 33$ and $j = 1, 2, 3, 4, 5$ and 6. The fixed-effect parameters are β_0, β_1, and β_2, and the respective ML estimates (standard errors between parentheses) are $\widehat{\beta}_0 = 3.6315(0.1722)$, $\widehat{\beta}_1 = -0.3209(0.0466)$, and $\widehat{\beta}_2 = 0.5331(0.2027)$. In addition, $\widehat{\rho}_1 = 0.3480$, $\widehat{\rho}_2 = -0.0419$, $\widehat{\rho}_3 = 0.0238$, and $\widehat{\sigma}^2 = 0.2802$. Here, we are interested in testing the null hypothesis $\mathcal{H}_0 : \beta_2 = 0$ against the alternative hypothesis $\mathcal{H}_1 : \beta_2 \neq 0$. The observed values of the LR statistics (and p-values between parentheses) are $\omega = 3.9354$ (p-value: 0.0473), $\omega_{\rm B} = 3.4027$ (p-value: 0.0651), $\omega^*_{\rm B} = 3.3651$ (p-value: 0.0666), $\omega_{\rm S} = 3.4367$ (p-value: $=$ 0.0638), and $\omega^*_{\rm S} = 3.4198$ (p-value: 0.0644). Note that, at the 5% nominal level, all corrected LR tests lead to the non-rejection of the null hypothesis, while the original LR test leads to the rejection of the null hypothesis. In other words, the test based on the original LR statistic ω leads to a different conclusion than the one obtained through its modified versions.

6.3. *Secalonic acid*

These data were studied in Gong et al. [31], whose objective was to explore the toxicity of secalonic acid in grass roots. The response variable corresponds to the root length (in *cm*), and the covariate is given by the dose of secalonic acid applied (in *mM*). The real data were also presented in Ritz and Streibig [32], who considered the nonlinear regression model given in (7) to model these data. The ML estimates (standard errors between parentheses) are $\widehat{\theta}_1 = 6.0536(0.2589), \widehat{\theta}_2 = 0.3539(0.1271), \widehat{\theta}_3 = 0.0752(0.0039), \widehat{\theta}_4 = 0.0293(0.0043)$, and $\widehat{\sigma}^2 = 0.0176(0.0094)$. Suppose we want to test the null hypothesis $\mathcal{H}_0 : \theta_2 = 0$ against $\mathcal{H}_1 : \theta_2 \neq 0$. We have that $\omega = 5.1953$ (p-value: 0.0226), $\omega_{\mathrm{B}} = 3.3023$ (p-value: 0.0692), $\omega_{\mathrm{B}}^* = 2.9285$ (p-value: 0.0870), $\omega_{\mathrm{S}} = 2.3686$ (p-value: 0.1238), and $\omega_{\mathrm{S}}^* = 1.8206$ (p-value: 0.1772). It is noteworthy that one rejects the null hypothesis at the 5% nominal level when the inference is based on the usual LR test, but a different inference is reached when the modified LR tests are used. Recall from the previous section that the unmodified LR test is markedly oversized when the sample is small (here, $n = 7$), which leads us to mistrust the inference delivered by the original LR test. Note also that the p-values of the improved tests that use ω_{S} and ω_{S}^* as test statistics are greater p-values the other ones, which is in accordance with the simulations provided in Section 5 for the non-linear regression model, that is, based on these corrected LR tests, the null hypothesis should not be rejected at any usual significance level.

7. Concluding remarks

We addressed the issue of performing hypothesis tests in the multivariate normal regression model with general parameterization introduced by Patriota and Lemonte [18] when the sample size is small. It includes many of the existing (univariate as well as multivariate) regression models as special cases, namely: nonlinear models, mixed-effect models, and errors-in-variables models, among others. The main theoretical contributions of this paper were the derivation of a Bartlett correction factor in matrix notation to the likelihood ratio statistic, and the derivation of Skovgaard's adjustment factor in matrix notation to the likelihood ratio statistic in this class of models. The Monte Carlo simulation results clearly indicate that the original likelihood ratio test can be considerably oversized (liberal) and should not be recommended to test hypotheses in the multivariate normal regression model with general parameterization when the sample is small or of moderate size. Also, the simulation results have convincingly shown that the inference based on the modified likelihood ratio test statistics can be much more accurate than that based on the unmodified likelihood ratio test statistic. Overall, the numerical results favour the tests obtained from applying the Bartlett correction to the likelihood ratio statistic, as well as from applying the Skovgaard's adjustment to the likelihood ratio statistic. The empirical applications with real data show that the uncorrected likelihood ratio test may lead to misleading conclusions if the sample is not large. We, therefore, recommend the use of the corrected likelihood ratio tests to make inference in the multivariate normal regression model with general parameterization in practice, mainly when the sample size is small.

Acknowledgments

The authors would like to thank the anonymous reviewers for the insightful comments and suggestions which have significantly improved the current work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Tatiane Melo and Tiago Vargas gratefully acknowledge the financial support of the Brazilian agency FAPEG [grant number 201410267001779]. Artur Lemonte gratefully acknowledges the financial support of the Brazilian agency Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [grant number 303554/2022–3].

References

[1] Bartlett MS. Properties of sufficiency and statistical tests. Proc R Soc A. 1937;160:268–282.
[2] Lawley DN. A general method for approximating to the distribution of likelihood ratio criteria. Biometrika. 1956;43:295–303. doi: 10.1093/biomet/43.3-4.295
[3] Barndorff-Nielsen OE, Hall P. On the level-error after Bartlett adjustment of the likelihood ratio statistic. Biometrika. 1988;75:374–378. doi: 10.1093/biomet/75.2.374
[4] Zucker DM, Lieberman O, Manor O. Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. J R Stat Soc B. 2000;62:827–838. doi: 10.1111/1467-9868.00267
[5] Cordeiro GM. Corrected likelihood ratio tests in symmetric nonlinear regression models. J Stat Comput Simul. 2004;74:609–620. doi: 10.1080/00949650310001638591
[6] Barroso LP, Cordeiro GM. Bartlett corrections in heteroskedastic *t* regression models. Stat Probab Lett. 2005;75:86–96. doi: 10.1016/j.spl.2005.05.025
[7] Cordeiro GM, Cysneiros AH, Cysneiros FJ. Bartlett adjustments for overdispersed generalized linear models. Commun Stat Theory Methods. 2006;35:937–952. doi: 10.1080/03610920500202796
[8] Fujita A, Kojima K, Patriota AG, et al. A fast and robust statistical test based on likelihood ratio with Bartlett correction to identify Granger causality between gene sets. Bioinformatics. 2010;26:2349–2351. doi: 10.1093/bioinformatics/btq427
[9] Araújo MC, Cysneiros AHMA, Montenegro LC. Improved heteroskedasticity likelihood ratio tests in symmetric nonlinear regression models. Stat Pap. 2020;61:167–188. doi: 10.1007/s00362-017-0933-5
[10] Guedes AC, Cribari-Neto F, Espinheira PL. Bartlett-corrected tests for varying precision beta regressions with application to environmental biometrics. PLoS ONE. 2021;16:e0253349. doi: 10.1371/journal.pone.0253349
[11] Larsson R. Bartlett correction of an independence test in a multivariate Poisson model. Stat Neerl. 2022;76:391–417. doi: 10.1111/stan.v76.4
[12] DiCiccio TJ, Stern SE. Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. J R Stat Soc B. 1994;56:397–408. doi: 10.1111/j.2517-6161.1994.tb01989.x
[13] Cribari-Neto F, Cordeiro GM. On Bartlett and Bartlett-type corrections. Econ Lett. 1996;15:339–367.
[14] Cordeiro GM, Cribari-Neto F. An introduction to Bartlett correction and bias reduction. New York: Springer; 2014.
[15] Wong MY. Likelihood estimation of a simple linear regression model when both variables have error. Biometrika. 1989;76:141–148. doi: 10.1093/biomet/76.1.141
[16] Arellano-Valle RB, Bolfarine H. A note on the simple structural regression model. Annal Intitute Stat Math. 1996;48:111–125. doi: 10.1007/BF00049293
[17] Wong MY. Bartlett adjustment to the likelihood ratio statistic for testing several slopes. Biometrika. 1991;78:221–224. doi: 10.1093/biomet/78.1.221
[18] Patriota AG, Lemonte AJ. Bias correction in a multivariate normal regression model with general parameterization. Stat Probab Lett. 2009;79:1655–1662. doi: 10.1016/j.spl.2009.04.018
[19] Skovgaard IM. Likelihood asymptotics. Scand J Stat. 2001;28:3–32. doi: 10.1111/sjos.2001.28.issue-1

[20] Melo TFN, Ferrari SLP, Patriota AG. Improved hypothesis testing in a general multivariate elliptical model. J Stat Comput Simul. 2017;87:1416–1428. doi: 10.1080/00949655.2016.1269330
[21] Lemonte AJ, Ferrari SLP, Cribari-Neto F. Improved likelihood inference in Birnbaum–Saunders regressions. Comput Stat Data Anal. 2010;54:1307–1316. doi: 10.1016/j.csda.2009.11.017
[22] Vargas TM, Ferrari SLP, Lemonte AJ. Improved likelihood inference in generalized linear models. Comput Stat Data Anal. 2014;74:110–124. doi: 10.1016/j.csda.2013.12.002
[23] Medeiros FMC, Ferrari SLP. Small-sample testing inference in symmetric and log-symmetric linear regression models. Stat Neerl. 2017;71:200–224. doi: 10.1111/stan.v71.3
[24] Doornik JA. Object-oriented matrix programming using OX. 9th ed. London: Timberlake Consultants Press; 2021.
[25] Cox DR, Hinkley DV. Theoretical statistics. London: Chapman and Hall; 1974.
[26] Skovgaard IM. An explicit large-deviation approximation to one-parameter tests. Bernoulli. 1996;2:145–165. doi: 10.2307/3318548
[27] R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2022.
[28] Mittelhammer RC, Judge GG, Miller DJ. Econometric foundations. New York: Cambridge University Press; 2000.
[29] Kelly G. The influence function in the errors in variables problem. Annal Stat. 1984;12:87–100. doi: 10.1214/aos/1176346394
[30] Davis CS. Statistical methods for the analysis of repeated measurements. New York: Springer-Verlag; 2002.
[31] Gong X, Zeng R, Luo S, et al. Two new secalonic acids from Aspergillus japonicus and their allelopathic effects on higher plants. In: Proceedings of International Symposium on Allelopathy Research and Application, 27–29 April, Shanshui, Guangdong, China. Guangzhou: Zongkai University of Agriculture and Technology; 2004. p. 209–217.
[32] Ritz C, Streibig JC. Nonlinear regression with R. New York: Springer; 2008.
[33] Cordeiro GM. General matrix formulae for computing Bartlett corrections. Stat Probab Lett. 1993;16:11–18. doi: 10.1016/0167-7152(93)90115-Y

Appendix. Proofs

Here, we provide the sketch of the proofs of Propositions 3.1 and 3.2. The log-likelihood function for $\boldsymbol{\theta}$, apart from an unimportant constant, is given by

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{i=1}^{n}\log\{\det(\boldsymbol{\Sigma}_i)\} - \frac{1}{2}\sum_{i=1}^{n}\mathrm{tr}\{\boldsymbol{\Sigma}_i^{-1}\boldsymbol{u}_i\boldsymbol{u}_i^\top\} = -\frac{1}{2}\log\{\det(\boldsymbol{\Sigma})\} - \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\boldsymbol{u}^\top\},$$

where $\boldsymbol{\mu}_i := \boldsymbol{\mu}_i(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_i := \boldsymbol{\Sigma}_i(\boldsymbol{\theta})$, $\boldsymbol{u}_i := \boldsymbol{u}_i(\boldsymbol{\theta}) = \boldsymbol{Y}_i - \boldsymbol{\mu}_i$, $\boldsymbol{Y} = \mathrm{vec}(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n)$, $\boldsymbol{\mu} := \boldsymbol{\mu}(\boldsymbol{\theta}) = \mathrm{vec}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \ldots, \boldsymbol{\mu}_n)$, $\boldsymbol{\Sigma} := \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathrm{diag}\{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots, \boldsymbol{\Sigma}_n\}$ and $\boldsymbol{u} := \boldsymbol{u}(\boldsymbol{\theta}) = \boldsymbol{Y} - \boldsymbol{\mu}$.

A.1 Proof of Proposition 3.1

After lengthy algebra, we have that

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_v} &= -\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_v\} - \frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_v\boldsymbol{u}\boldsymbol{u}^\top - 2\boldsymbol{u}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_v\}, \\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_t \partial \theta_v} &= -\frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_t\boldsymbol{C}_v + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{tv} + \boldsymbol{A}_{tv}\boldsymbol{u}\boldsymbol{u}^\top - 2\boldsymbol{A}_v\boldsymbol{a}_t\boldsymbol{u}^\top + 2\boldsymbol{a}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_v \\
&\quad - 2\boldsymbol{u}^\top\boldsymbol{A}_t\boldsymbol{a}_v - 2\boldsymbol{u}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tv}\},
\end{aligned}$$

$$\frac{\partial^3 \ell(\boldsymbol{\theta})}{\partial\theta_s\partial\theta_t\partial\theta_v} = -\frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_{st}\boldsymbol{C}_v + \boldsymbol{A}_t\boldsymbol{C}_{sv} + \boldsymbol{A}_s\boldsymbol{C}_{tv} + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{stv}\} + \boldsymbol{A}_{stv}\boldsymbol{u}\boldsymbol{u}^\top$$
$$- 2\boldsymbol{A}_{tv}\boldsymbol{a}_s\boldsymbol{u}^\top - 2\boldsymbol{A}_{sv}\boldsymbol{a}_t\boldsymbol{u}^\top - 2\boldsymbol{A}_v\boldsymbol{a}_{st}\boldsymbol{u}^\top + 2\boldsymbol{A}_v\boldsymbol{a}_t\boldsymbol{a}_s^\top$$
$$+ 2\boldsymbol{a}_{st}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_v + 2\boldsymbol{a}_t^\top\boldsymbol{A}_s\boldsymbol{a}_v + 2\boldsymbol{a}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{sv} + 2\boldsymbol{a}_s^\top\boldsymbol{A}_t\boldsymbol{a}_v$$
$$- 2\boldsymbol{u}^\top\boldsymbol{A}_{st}\boldsymbol{a}_v - 2\boldsymbol{u}^\top\boldsymbol{A}_t\boldsymbol{a}_{sv} + 2\boldsymbol{a}_s^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tv} - 2\boldsymbol{u}^\top\boldsymbol{A}_s\boldsymbol{a}_{tv}$$
$$- 2\boldsymbol{u}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{stv},$$
$$\frac{\partial^4 \ell(\boldsymbol{\theta})}{\partial\theta_r\theta_s\partial\theta_t\partial\theta_v} = -\frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_{rst}\boldsymbol{C}_v + \boldsymbol{A}_{st}\boldsymbol{C}_{rv} + \boldsymbol{A}_{rt}\boldsymbol{C}_{sv} + \boldsymbol{A}_t\boldsymbol{C}_{rsv} + \boldsymbol{A}_{rs}\boldsymbol{C}_{tv}$$
$$+ \boldsymbol{A}_s\boldsymbol{C}_{rtv} + \boldsymbol{A}_r\boldsymbol{C}_{stv} + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{rstv} + \boldsymbol{A}_{rstv}\boldsymbol{u}\boldsymbol{u}^\top - 2\boldsymbol{A}_{stv}\boldsymbol{a}_r\boldsymbol{u}^\top$$
$$- 2\boldsymbol{A}_{rtv}\boldsymbol{a}_s\boldsymbol{u}^\top - 2\boldsymbol{A}_{tv}\boldsymbol{a}_{rs}\boldsymbol{u}^\top + 2\boldsymbol{A}_{tv}\boldsymbol{a}_s\boldsymbol{a}_r^\top - 2\boldsymbol{A}_{rsv}\boldsymbol{a}_t\boldsymbol{u}^\top$$
$$- 2\boldsymbol{A}_{sv}\boldsymbol{a}_{rt}\boldsymbol{u}^\top + 2\boldsymbol{A}_{sv}\boldsymbol{a}_t\boldsymbol{a}_r^\top - 2\boldsymbol{A}_{rv}\boldsymbol{a}_{st}\boldsymbol{u}^\top - 2\boldsymbol{A}_v\boldsymbol{a}_{rst}\boldsymbol{u}^\top$$
$$+ 2\boldsymbol{A}_v\boldsymbol{a}_{st}\boldsymbol{a}_r^\top + 2\boldsymbol{A}_{rv}\boldsymbol{a}_t\boldsymbol{a}_s^\top + 2\boldsymbol{A}_v\boldsymbol{a}_{rt}\boldsymbol{a}_s^\top + 2\boldsymbol{A}_v\boldsymbol{a}_t\boldsymbol{a}_{rs}^\top$$
$$+ 2\boldsymbol{a}_{rst}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_v + 2\boldsymbol{a}_{st}^\top\boldsymbol{A}_r\boldsymbol{a}_v + 2\boldsymbol{a}_{st}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{rv} + 2\boldsymbol{a}_{rt}^\top\boldsymbol{A}_s\boldsymbol{a}_v$$
$$+ 2\boldsymbol{a}_t^\top\boldsymbol{A}_{rs}\boldsymbol{a}_v + 2\boldsymbol{a}_t^\top\boldsymbol{A}_s\boldsymbol{a}_{rv} + 2\boldsymbol{a}_{rt}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{sv} + 2\boldsymbol{a}_t^\top\boldsymbol{A}_r\boldsymbol{a}_{sv}$$
$$+ 2\boldsymbol{a}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{rsv} + 2\boldsymbol{a}_{rs}^\top\boldsymbol{A}_t\boldsymbol{a}_v + 2\boldsymbol{a}_s^\top\boldsymbol{A}_{rt}\boldsymbol{a}_v + 2\boldsymbol{a}_s^\top\boldsymbol{A}_t\boldsymbol{a}_{rv}$$
$$+ 2\boldsymbol{a}_r^\top\boldsymbol{A}_{st}\boldsymbol{a}_v - 2\boldsymbol{u}^\top\boldsymbol{A}_{rst}\boldsymbol{a}_v - 2\boldsymbol{u}^\top\boldsymbol{A}_{st}\boldsymbol{a}_{rv} + 2\boldsymbol{a}_r^\top\boldsymbol{A}_t\boldsymbol{a}_{sv}$$
$$- 2\boldsymbol{u}^\top\boldsymbol{A}_{rt}\boldsymbol{a}_{sv} - 2\boldsymbol{u}^\top\boldsymbol{A}_t\boldsymbol{a}_{rsv} + 2\boldsymbol{a}_{rs}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tv} + 2\boldsymbol{a}_s^\top\boldsymbol{A}_r\boldsymbol{a}_{tv}$$
$$+ 2\boldsymbol{a}_s^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{rtv} + 2\boldsymbol{a}_r^\top\boldsymbol{A}_s\boldsymbol{a}_{tv} - 2\boldsymbol{u}^\top\boldsymbol{A}_{rs}\boldsymbol{a}_{tv} - 2\boldsymbol{u}^\top\boldsymbol{A}_s\boldsymbol{a}_{rtv}$$
$$+ 2\boldsymbol{a}_r^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{stv} - 2\boldsymbol{u}^\top\boldsymbol{A}_r\boldsymbol{a}_{stv} - 2\boldsymbol{u}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{rstv}\},$$

and all the quantities in the above expressions were defined in Section 2. Note that $\mathrm{E}(\boldsymbol{u}) = \boldsymbol{0}$ and $\mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \boldsymbol{\Sigma}$, and so the joint cumulants of log-likelihood derivatives become

$$\kappa_{tv} = \mathrm{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\theta_t\partial\theta_v}\right) = -\frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_t\boldsymbol{C}_v + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{tv} + \boldsymbol{A}_{tv}\boldsymbol{\Sigma}\} - \boldsymbol{a}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_v,$$
$$\kappa_{stv} = \mathrm{E}\left(\frac{\partial^3 \ell(\boldsymbol{\theta})}{\partial\theta_s\partial\theta_t\partial\theta_v}\right)$$
$$= -\frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_{st}\boldsymbol{C}_v + \boldsymbol{A}_t\boldsymbol{C}_{sv} + \boldsymbol{A}_s\boldsymbol{C}_{tv} + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{stv} + \boldsymbol{A}_{stv}\boldsymbol{\Sigma}\}$$
$$- \boldsymbol{a}_s^\top\boldsymbol{A}_v\boldsymbol{a}_t - \boldsymbol{a}_{st}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_v - \boldsymbol{a}_t^\top\boldsymbol{A}_s\boldsymbol{a}_v - \boldsymbol{a}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{sv} - \boldsymbol{a}_s^\top\boldsymbol{A}_t\boldsymbol{a}_v$$
$$- \boldsymbol{a}_s^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tv},$$
$$\kappa_{rstv} = \mathrm{E}\left(\frac{\partial^4 \ell(\boldsymbol{\theta})}{\partial\theta_r\theta_s\partial\theta_t\partial\theta_v}\right)$$
$$= -\frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_{rst}\boldsymbol{C}_v + \boldsymbol{A}_{st}\boldsymbol{C}_{rv} + \boldsymbol{A}_{rt}\boldsymbol{C}_{sv} + \boldsymbol{A}_t\boldsymbol{C}_{rsv} + \boldsymbol{A}_{rs}\boldsymbol{C}_{tv} + \boldsymbol{A}_s\boldsymbol{C}_{rtv}$$
$$+ \boldsymbol{A}_r\boldsymbol{C}_{stv} + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{rstv} + \boldsymbol{A}_{rstv}\boldsymbol{\Sigma}\} - \boldsymbol{a}_r^\top\boldsymbol{A}_{tv}\boldsymbol{a}_s - \boldsymbol{a}_r^\top\boldsymbol{A}_{sv}\boldsymbol{a}_t - \boldsymbol{a}_s^\top\boldsymbol{A}_{rv}\boldsymbol{a}_t$$
$$- \boldsymbol{a}_v^\top\boldsymbol{A}_{rs}\boldsymbol{a}_t - \boldsymbol{a}_v^\top\boldsymbol{A}_{rt}\boldsymbol{a}_s - \boldsymbol{a}_r^\top\boldsymbol{A}_{st}\boldsymbol{a}_v - \boldsymbol{a}_r^\top\boldsymbol{A}_v\boldsymbol{a}_{st} - \boldsymbol{a}_v^\top\boldsymbol{A}_r\boldsymbol{a}_{st} - \boldsymbol{a}_s^\top\boldsymbol{A}_v\boldsymbol{a}_{rt}$$
$$- \boldsymbol{a}_v^\top\boldsymbol{A}_s\boldsymbol{a}_{rt} - \boldsymbol{a}_t^\top\boldsymbol{A}_v\boldsymbol{a}_{rs} - \boldsymbol{a}_v^\top\boldsymbol{A}_t\boldsymbol{a}_{rs} - \boldsymbol{a}_t^\top\boldsymbol{A}_s\boldsymbol{a}_{rv} - \boldsymbol{a}_s^\top\boldsymbol{A}_t\boldsymbol{a}_{rv} - \boldsymbol{a}_t^\top\boldsymbol{A}_r\boldsymbol{a}_{sv}$$

$$- \boldsymbol{a}_r^\top \boldsymbol{A}_t \boldsymbol{a}_{sv} - \boldsymbol{a}_s^\top \boldsymbol{A}_r \boldsymbol{a}_{tv} - \boldsymbol{a}_r^\top \boldsymbol{A}_s \boldsymbol{a}_{tv} - \boldsymbol{a}_r^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{stv} - \boldsymbol{a}_s^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{rtv}$$
$$- \boldsymbol{a}_t^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{rsv} - \boldsymbol{a}_v^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{rst} - \boldsymbol{a}_{rs}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{tv} - \boldsymbol{a}_{rt}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{sv} - \boldsymbol{a}_{st}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{rv}.$$

In addition, it follows that

$$\begin{aligned}
\kappa_{vs}^{(t)} &= \frac{\partial \kappa_{vs}}{\partial \theta_t} = -\frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_{tv}\boldsymbol{C}_s + \boldsymbol{A}_{vs}\boldsymbol{C}_t + \boldsymbol{A}_v\boldsymbol{C}_{ts} + \boldsymbol{A}_t\boldsymbol{C}_{vs} + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{vs} + \boldsymbol{A}_{tvs}\boldsymbol{\Sigma})\\
&- \boldsymbol{a}_s\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tv} - \boldsymbol{a}_v\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{ts} - \boldsymbol{a}_v\boldsymbol{A}_t\boldsymbol{a}_s,\\
\kappa_{tvr}^{(s)} &= \frac{\partial \kappa_{tvr}}{\partial \theta_s} = -\frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_{stv}\boldsymbol{C}_r + \boldsymbol{A}_{tv}\boldsymbol{C}_{sr} + \boldsymbol{A}_{sv}\boldsymbol{C}_{tr} + \boldsymbol{A}_{st}\boldsymbol{C}_{vr} + \boldsymbol{A}_v\boldsymbol{C}_{str} + \boldsymbol{A}_t\boldsymbol{C}_{svr}\\
&+ \boldsymbol{A}_s\boldsymbol{C}_{tvr} + \boldsymbol{A}_{tvr}\boldsymbol{C}_s + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{stvr} + \boldsymbol{A}_{sttvr}\boldsymbol{\Sigma}) - \boldsymbol{a}_v^\top\boldsymbol{A}_r\boldsymbol{a}_{st} - \boldsymbol{a}_t^\top\boldsymbol{A}_r\boldsymbol{a}_{sv}\\
&- \boldsymbol{a}_r^\top\boldsymbol{A}_t\boldsymbol{a}_{sv} - \boldsymbol{a}_v^\top\boldsymbol{A}_t\boldsymbol{a}_{sr} - \boldsymbol{a}_r^\top\boldsymbol{A}_v\boldsymbol{a}_{st} - \boldsymbol{a}_t^\top\boldsymbol{A}_v\boldsymbol{a}_{sr} - \boldsymbol{a}_t^\top\boldsymbol{A}_s\boldsymbol{a}_{vr} - \boldsymbol{a}_v^\top\boldsymbol{A}_s\boldsymbol{a}_{tr}\\
&- \boldsymbol{a}_r^\top\boldsymbol{A}_s\boldsymbol{a}_{tv} - \boldsymbol{a}_t^\top\boldsymbol{A}_{sr}\boldsymbol{a}_v - \boldsymbol{a}_v^\top\boldsymbol{A}_{st}\boldsymbol{a}_r - \boldsymbol{a}_t^\top\boldsymbol{A}_{sv}\boldsymbol{a}_r - \boldsymbol{a}_{st}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{vr} - \boldsymbol{a}_{sv}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tr}\\
&- \boldsymbol{a}_{sr}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tv} - \boldsymbol{a}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{svr} - \boldsymbol{a}_v^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{str} - \boldsymbol{a}_r^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{stv},\\
\kappa_{tr}^{(vs)} &= \frac{\partial^2 \kappa_{tr}}{\partial \theta_v \partial \theta_s} = -\frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_{vst}\boldsymbol{C}_r + \boldsymbol{A}_{vtr}\boldsymbol{C}_s + \boldsymbol{A}_{str}\boldsymbol{C}_v + \boldsymbol{A}_{st}\boldsymbol{C}_{vr} + \boldsymbol{A}_{tr}\boldsymbol{C}_{vs} + \boldsymbol{A}_{vt}\boldsymbol{C}_{sr}\\
&+ \boldsymbol{A}_{vs}\boldsymbol{C}_{tr} + \boldsymbol{A}_t\boldsymbol{C}_{vsr} + \boldsymbol{A}_s\boldsymbol{C}_{vtr} + \boldsymbol{A}_v\boldsymbol{C}_{str} + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{vstr} + \boldsymbol{A}_{vstr}\boldsymbol{\Sigma})\\
&- \boldsymbol{a}_r^\top\boldsymbol{A}_v\boldsymbol{a}_{st} - \boldsymbol{a}_t^\top\boldsymbol{A}_v\boldsymbol{a}_{sr} - \boldsymbol{a}_{vt}^\top\boldsymbol{A}_s\boldsymbol{a}_r - \boldsymbol{a}_t^\top\boldsymbol{A}_s\boldsymbol{a}_{vr} - \boldsymbol{a}_t^\top\boldsymbol{A}_{vs}\boldsymbol{a}_r\\
&- \boldsymbol{a}_{vr}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{st} - \boldsymbol{a}_{vt}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{sr} - \boldsymbol{a}_r^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{vst} - \boldsymbol{a}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{vsr}.
\end{aligned}$$

Finally, from the general matrix result for the Bartlett correction factor in Cordeiro [33], we immediately obtain the (t, v)th elements of the matrices $\boldsymbol{E}^{(rs)}$, $\boldsymbol{P}^{(r)}$, and $\boldsymbol{Q}^{(s)}$ given by

$$\begin{aligned}
P_{tv}^{(r)} &= \kappa_{tvr} = -\frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_{tv}\boldsymbol{C}_r + \boldsymbol{A}_v\boldsymbol{C}_{tr} + \boldsymbol{A}_t\boldsymbol{C}_{vr} + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{tvr} + \boldsymbol{A}_{tvr}\boldsymbol{\Sigma}\} - \boldsymbol{a}_t^\top\boldsymbol{A}_r\boldsymbol{a}_v\\
&\quad - \boldsymbol{a}_v^\top\boldsymbol{A}_t\boldsymbol{a}_r - \boldsymbol{a}_t^\top\boldsymbol{A}_v\boldsymbol{a}_r - \boldsymbol{a}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{vr} - \boldsymbol{a}_v^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tr} - \boldsymbol{a}_r^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tv},\\
Q_{tv}^{(s)} &= \kappa_{vs}^{(t)} = -\frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_{tv}\boldsymbol{C}_s + \boldsymbol{A}_{vs}\boldsymbol{C}_t + \boldsymbol{A}_v\boldsymbol{C}_{ts} + \boldsymbol{A}_t\boldsymbol{C}_{vs} + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{tvs} + \boldsymbol{A}_{tvs}\boldsymbol{\Sigma}\}\\
&\quad - \boldsymbol{a}_v^\top\boldsymbol{A}_t\boldsymbol{a}_s - \boldsymbol{a}_s^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tv} - \boldsymbol{a}_v^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{ts},
\end{aligned}$$

and

$$\begin{aligned}
E_{tv}^{(rs)} &= \frac{1}{4}\kappa_{tvrs} - \kappa_{tvr}^{(s)} + \kappa_{tr}^{(vs)},\\
&= -\frac{1}{8}\mathrm{tr}\{\boldsymbol{A}_{tvr}\boldsymbol{C}_s + 4\boldsymbol{A}_{str}\boldsymbol{C}_v + \boldsymbol{A}_t\boldsymbol{C}_{vrs} + \boldsymbol{A}_v\boldsymbol{C}_{trs}\\
&\quad + \boldsymbol{A}_r\boldsymbol{C}_{tvs} + \boldsymbol{A}_{tv}\boldsymbol{C}_{rs} + \boldsymbol{A}_{vr}\boldsymbol{C}_{ts} + 5\boldsymbol{A}_{tr}\boldsymbol{C}_{vs} + \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{tvrs}\\
&\quad + \boldsymbol{A}_{tvrs}\boldsymbol{\Sigma}\} - \frac{1}{4}(\boldsymbol{a}_t^\top\boldsymbol{A}_{vs}\boldsymbol{a}_r + \boldsymbol{a}_t^\top\boldsymbol{A}_{vr}\boldsymbol{a}_s - 3\boldsymbol{a}_t^\top\boldsymbol{A}_{rs}\boldsymbol{a}_v - 3\boldsymbol{a}_v^\top\boldsymbol{A}_{ts}\boldsymbol{a}_r\\
&\quad + \boldsymbol{a}_s^\top\boldsymbol{A}_{tv}\boldsymbol{a}_r + \boldsymbol{a}_s^\top\boldsymbol{A}_{tr}\boldsymbol{a}_v + \boldsymbol{a}_t^\top\boldsymbol{A}_s\boldsymbol{a}_{vr} - 3\boldsymbol{a}_t^\top\boldsymbol{A}_r\boldsymbol{a}_{vs} + \boldsymbol{a}_t^\top\boldsymbol{A}_v\boldsymbol{a}_{rs}\\
&\quad + \boldsymbol{a}_s^\top\boldsymbol{A}_t\boldsymbol{a}_{vr} + \boldsymbol{a}_s^\top\boldsymbol{A}_v\boldsymbol{a}_{tr} + \boldsymbol{a}_s^\top\boldsymbol{A}_r\boldsymbol{a}_{tv} + \boldsymbol{a}_r^\top\boldsymbol{A}_v\boldsymbol{a}_{ts} - 3\boldsymbol{a}_r^\top\boldsymbol{A}_t\boldsymbol{a}_{vs}\\
&\quad + \boldsymbol{a}_r^\top\boldsymbol{A}_s\boldsymbol{a}_{tv} - 3\boldsymbol{a}_v^\top\boldsymbol{A}_s\boldsymbol{a}_{tr} - 3\boldsymbol{a}_v^\top\boldsymbol{A}_r\boldsymbol{a}_{ts} - 3\boldsymbol{a}_v^\top\boldsymbol{A}_t\boldsymbol{a}_{rs} + \boldsymbol{a}_{tv}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{rs}\\
&\quad - 3\boldsymbol{a}_{tr}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{vs} + \boldsymbol{a}_{vr}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{ts} + \boldsymbol{a}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{vrs} - 3\boldsymbol{a}_v^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{trs}\\
&\quad + \boldsymbol{a}_r^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tvs} + \boldsymbol{a}_s^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tvr}\}.
\end{aligned}$$

Therefore, the result follows.

A.2 Proof of Proposition 3.2

Let $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta})$, where

$$\ell_i(\boldsymbol{\theta}) = -\frac{1}{2}\log\{\det(\boldsymbol{\Sigma}_i)\} - \frac{1}{2}\boldsymbol{u}_i^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{u}_i.$$

Note that $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$ are independent for $i \neq j$, and so we initially have that

$$\mathrm{E}_{\omega_1}[\boldsymbol{U}(\boldsymbol{\omega}_1)(\ell(\boldsymbol{\omega}_1) - \ell(\boldsymbol{\omega}_2))] = \sum_{i=1}^{n} \boldsymbol{F}_i(\boldsymbol{\omega}_1)^\top \boldsymbol{H}_i(\boldsymbol{\omega}_1)\mathrm{E}_{\omega_1}\big[\boldsymbol{u}_i^*(\boldsymbol{\omega}_1)(\ell_i(\boldsymbol{\omega}_1) - \ell_i(\boldsymbol{\omega}_2))\big],$$

$$\mathrm{E}_{\omega_1}[\boldsymbol{U}(\boldsymbol{\omega}_1)\boldsymbol{U}(\boldsymbol{\omega}_2)^\top] = \sum_{i=1}^{n} \boldsymbol{F}_i(\boldsymbol{\omega}_1)^\top \boldsymbol{H}_i(\boldsymbol{\omega}_1)\mathrm{E}_{\omega_1}\big[\boldsymbol{u}_i^*(\boldsymbol{\omega}_1)\boldsymbol{u}_i^*(\boldsymbol{\omega}_2)^\top\big]\boldsymbol{H}_i(\boldsymbol{\omega}_2)\boldsymbol{F}_i(\boldsymbol{\omega}_2),$$

where $\boldsymbol{F}_i$, $\boldsymbol{H}_i$ and $\boldsymbol{u}_i^*$ were defined in Section 2. Thus, the problem is reduced to obtain the expectations

$$\mathrm{E}_{\omega_1}\big[\boldsymbol{u}_i^*(\boldsymbol{\omega}_1)(\ell_i(\boldsymbol{\omega}_1) - \ell_i(\boldsymbol{\omega}_2))\big] \quad \text{and} \quad \mathrm{E}_{\omega_1}\big[\boldsymbol{u}_i^*(\boldsymbol{\omega}_1)\boldsymbol{u}_i^*(\boldsymbol{\omega}_2)^\top\big].$$

Note that

$$\begin{aligned}\boldsymbol{u}_i(\boldsymbol{\omega}_2)^\top \boldsymbol{\Sigma}_i(\boldsymbol{\omega}_2)^{-1}\boldsymbol{u}_i(\boldsymbol{\omega}_2) = \boldsymbol{u}_i(\boldsymbol{\omega}_1)^\top \boldsymbol{\Sigma}_i(\boldsymbol{\omega}_2)^{-1}\boldsymbol{u}_i(\boldsymbol{\omega}_1) \\ + 2\boldsymbol{u}_i(\boldsymbol{\omega}_1)^\top \boldsymbol{\Sigma}_i(\boldsymbol{\omega}_2)^{-1}(\boldsymbol{\mu}_i(\boldsymbol{\omega}_1) - \boldsymbol{\mu}_i(\boldsymbol{\omega}_2)) \\ + (\boldsymbol{\mu}_i(\boldsymbol{\omega}_1) - \boldsymbol{\mu}_i(\boldsymbol{\omega}_2))^\top \boldsymbol{\Sigma}_i(\boldsymbol{\omega}_2)^{-1}(\boldsymbol{\mu}_i(\boldsymbol{\omega}_1) - \boldsymbol{\mu}_i(\boldsymbol{\omega}_2)).\end{aligned}$$

Hence, after some lengthy algebra, we have that

$$\mathrm{E}_{\omega_1}\big[\boldsymbol{u}_i^*(\boldsymbol{\omega}_1)(\ell_i(\boldsymbol{\omega}_1) - \ell_i(\boldsymbol{\omega}_2))\big] = \begin{pmatrix} \boldsymbol{\Sigma}_i(\boldsymbol{\omega}_1)\boldsymbol{\Sigma}_i(\boldsymbol{\omega}_2)^{-1}\big(\boldsymbol{\mu}_i(\boldsymbol{\omega}_1) - \boldsymbol{\mu}_i(\boldsymbol{\omega}_2)\big) \\ \mathrm{vec}\big(\boldsymbol{\Sigma}_i(\boldsymbol{\omega}_1)\boldsymbol{\Sigma}_i(\boldsymbol{\omega}_2)^{-1}\boldsymbol{\Sigma}_i(\boldsymbol{\omega}_1) - \boldsymbol{\Sigma}_i(\boldsymbol{\omega}_1)\big)\end{pmatrix},$$

and

$$\mathrm{E}_{\omega_1}\big[\boldsymbol{u}_i^*(\boldsymbol{\omega}_1)\boldsymbol{u}_i^*(\boldsymbol{\omega}_2)^\top\big] = \boldsymbol{H}_i(\boldsymbol{\omega}_1)^{-1}.$$

Therefore, the result follows. It is worth mentioning that to obtain the above expressions, we make use of the following result. Let $\boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{0}, \boldsymbol{I}_n)$, where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix, $\boldsymbol{A}$ is a $n \times n$ symmetric (definite positive) matrix and $\boldsymbol{a}$ and $\boldsymbol{b}$ are $(n \times 1)$ vectors. It follows that

$$\begin{aligned}&\mathrm{E}[\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})] = \boldsymbol{0}, \quad \mathrm{E}[\boldsymbol{X}\boldsymbol{X}^\top(\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X})] = \mathrm{tr}\{\boldsymbol{A}\}\boldsymbol{I}_n + 2\boldsymbol{A}, \\ &\mathrm{E}[\boldsymbol{X}\boldsymbol{X}^\top(\boldsymbol{a}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{b})] = (\boldsymbol{a}^\top\boldsymbol{b})\boldsymbol{I}_n + \boldsymbol{a}\boldsymbol{b}^\top + \boldsymbol{b}\boldsymbol{a}^\top.\end{aligned}$$