# Improved estimation in a general multivariate elliptical model

### Tatiane F. N. Melo[a], Silvia L. P. Ferrari[b] and Alexandre G. Patriota[b]

[a]*Federal University of Goiás*
[b]*University of São Paulo*

**Abstract.** The problem of reducing the bias of maximum likelihood estimator in a general multivariate elliptical regression model is considered. The model is very flexible and allows the mean vector and the dispersion matrix to have parameters in common. Many frequently used models are special cases of this general formulation, namely: errors-in-variables models, nonlinear mixed-effects models, heteroscedastic nonlinear models, among others. In any of these models, the vector of the errors may have any multivariate elliptical distribution. We obtain the second-order bias of the maximum likelihood estimator, a bias-corrected estimator, and a bias-reduced estimator. Simulation results indicate the effectiveness of the bias correction and bias reduction schemes.

## 1 Introduction

It is well known that, under some standard regularity conditions, maximum-likelihood estimators (MLEs) are consistent and asymptotically normally distributed. Hence, their biases converge to zero when the sample size increases. However, for finite sample sizes, the MLEs are in general biased and bias correction plays an important role in the point estimation theory.

A general expression for the term of order $O(n^{-1})$ in the expansion of the bias of MLEs was given by Cox and Snell (1968). This term is often called second-order bias and can be useful in actual problems. For instance, a very high second-order bias indicates that other than maximum-likelihood estimation procedures should be used. Also, corrected estimators can be formulated by subtracting the estimated second-order biases from the respective MLEs. It is expected that these corrected estimators have smaller biases than the uncorrected ones, especially in small samples.

Cox and Snell's formulae for second-order biases of MLEs were applied in many models. Cordeiro and McCullagh (1991) use these formulae in generalized linear models; Cordeiro and Klein (1994) compute them for ARMA models; Cordeiro et al. (2000) apply them for symmetric nonlinear regression models;

Vasconcellos and Cordeiro (2000) obtain them for multivariate nonlinear Student $t$ regression models. More recently, Cysneiros, Cordeiro and Cysneiros (2010) study the univariate heteroscedastic symmetric nonlinear regression models (which are an extension of Cordeiro et al., 2000) and Patriota and Lemonte (2009) obtain a general matrix formula for the bias correction in a multivariate normal model where the mean and the covariance matrix have parameters in common.

An alternative approach to bias correction was suggested by Firth (1993). The idea is to adjust the estimating function so that the estimate becomes less biased. This approach can be viewed as a "preventive" method, since it modifies the original score function, prior to obtaining the parameter estimates. In this paper, estimates obtained from Cox and Snell's approach and Firth's method will be called bias-corrected estimates and bias-reduced estimates, respectively. Firth showed that in generalized linear models with canonical link function the preventive method is equivalent to maximizing a penalized likelihood that is easily implemented via an iterative adjustment of the data. The bias reduction proposed by Firth has received considerable attention in the statistical literature. For models for binary data, see Mehrabi and Matthews (1995); for censored data with exponential lifetimes, see Pettitt, Kelly and Gao (1998). In Bull, Mak and Greenwood (2002) bias reduction is obtained for the multinomial logistic regression model. In Kosmidis and Firth (2009) a family of bias-reducing adjustments was developed for a general class of univariate and multivariate generalized nonlinear models. The bias reduction in cumulative link models for ordinal data was studied in Kosmidis (2014). Additionally, Kosmidis and Firth (2011) showed how to obtain the bias-reducing penalized maximum likelihood estimator by using the equivalent Poisson log-linear model for the parameters of a multinomial logistic regression.

It is well-known and was noted by Firth (1993) and Kosmidis and Firth (2009) that the reduction in bias may sometimes be accompanied by inflation of variance, possibly yielding an estimator whose mean squared error is bigger than that of the original one. Nevertheless, published empirical studies such as those mentioned above show that, in some frequently used models, bias-reduced and bias-corrected estimators can perform better than the unadjusted maximum likelihood estimators, especially when the sample size is small.

Our goal in this paper is to obtain bias correction and bias reduction to the maximum likelihood estimators for the general multivariate elliptical model. We extend the work of Patriota and Lemonte (2009) to the elliptical class of distributions defined in Lemonte and Patriota (2011). We focus on analytical methods only, because simulations for a general multivariate normal model suggests that analytical bias corrections outperforms the computationally intensive bootstrap methods (Lemonte, 2011).

In order to illustrate the ampleness of the general multivariate elliptical model, we mention some of its submodels: multiple linear regression, heteroscedastic multivariate nonlinear regressions, nonlinear mixed-effects models (Patriota,

2011), heteroscedastic errors-in-variables models (Patriota, Bolfarine and de Castro, 2009; Patriota and Lemonte, 2009), structural equation models, multivariate normal regression model with general parametrization (Lemonte, 2011), simultaneous equation models and mixtures of them. It is important to note that the usual normality assumption of the error is relaxed and replaced by the assumption of elliptical errors. The elliptical family of distributions includes many important distributions such as multivariate normal, Student $t$, power exponential, contaminated normal, Pearson II, Pearson VII, and logistic, with heavier or lighter tails than the normal distribution; see Fang, Kotz and Ng (1990).

The paper is organized as follows. Section 2 presents the notation and general results for bias correction and bias reduction. Section 3 presents the model and our main results, namely the general expression for the second-order bias of MLEs, in the general multivariate elliptical model. Section 4 applies our results in four important special cases: heteroscedastic nonlinear (linear) model, nonlinear mixed-effects models, multivariate errors-in-variables models and log-symmetric regression models. Simulations are presented in Section 5. Applications that use real data are presented and discussed in Section 6. Finally, Section 7 concludes the paper. Technical details are collected in one Appendix.

## 2 Bias correction and bias reduction

Let $\theta$ be the $p$-vector of unknown parameters and $\theta_r$ its $r$th element. Also, let $U(\theta)$ be the score function and $U_r(\theta) = U_r$ its $r$th element. We use the following tensor notation for the cumulants of the log-likelihood derivatives introduced by Lawley (1956):

$$\kappa_{rs} = E\left(\frac{\partial U_r}{\partial \theta_s}\right), \qquad \kappa_{r,s} = E(U_r U_s), \qquad \kappa_{rs,t} = E\left(\frac{\partial U_r}{\partial \theta_s} U_t\right),$$

$$\kappa_{rst} = E\left(\frac{\partial^2 U_r}{\partial \theta_s \, \partial \theta_t}\right), \qquad \kappa_{rs}^{(t)} = \frac{\partial \kappa_{rs}}{\partial \theta_t}, \qquad \kappa_{r,s,t} = E(U_r U_s U_t),$$

and so on. The indices $r$, $s$ and $t$ vary from 1 to $p$. The typical $(r, s)$th element of the Fisher information matrix $K(\theta)$ is $\kappa_{r,s}$ and we denote by $\kappa^{r,s}$ the corresponding element of $K(\theta)^{-1}$. All $\kappa$'s refer to a total over the sample and are, in general, of order $n$. Under standard regular conditions, we have that $\kappa_{rs} = -\kappa_{r,s}$, $\kappa_{rs,t} = \kappa_{rs}^{(t)} - \kappa_{rst}$ and $\kappa_{r,s,t} = 2\kappa_{rst} - \kappa_{rs}^{(t)} - \kappa_{rt}^{(s)} - \kappa_{st}^{(r)}$. These identities will be used to facilitate some algebraic operations.

Let $B_{\widehat{\theta}}(\theta)$ be the second-order bias vector of $\widehat{\theta}$ whose $j$th element is $B_{\widehat{\theta}_j}(\theta)$, $j = 1, 2, \ldots, p$. It follows from the general expression for the multiparameter second-order biases of MLEs given by Cox and Snell (1968) that

$$B_{\widehat{\theta}_j}(\theta) = \sum_{r,s,t=1}^{p} \kappa^{j,r} \kappa^{s,t} \left\{\frac{1}{2}\kappa_{rst} + \kappa_{rs,t}\right\}. \tag{2.1}$$

The bias corrected MLE is defined as

$$\widehat{\theta}_{\mathrm{BC}} = \widehat{\theta} - B_{\widehat{\theta}}(\widehat{\theta}).$$

The bias-corrected estimator $\widehat{\theta}_{\mathrm{BC}}$ is expected to have smaller bias than the uncorrected estimator, $\widehat{\theta}$.

Firth (1993) proposed an alternative method to partially remove the bias of MLEs. The method replaces the score function by its modified version

$$U^*(\theta) = U(\theta) - K(\theta)B_{\widehat{\theta}}(\theta),$$

and a modified estimate, $\widehat{\theta}_{\mathrm{BR}}$, is given as a solution to $U^*(\theta) = 0$. It is noticeable that, unlike Cox and Snell's approach, Firth's bias reduction method does not depend on the finiteness of $\widehat{\theta}$.

## 3 Model and main results

We shall follow the same notation presented in Lemonte and Patriota (2011). The elliptical model as defined in Fang, Kotz and Ng (1990) follows. A $q \times 1$ random vector $Y$ has a multivariate elliptical distribution with location parameter $\mu$ and a definite positive scale matrix $\Sigma$ if its density function is

$$f_Y(y) = |\Sigma|^{-1/2} g\big((y - \mu)^\top \Sigma^{-1}(y - \mu)\big), \tag{3.1}$$

where $g : [0, \infty) \to (0, \infty)$ is called the density generating function, and it is such that $\int_0^\infty u^{\frac{q}{2}-1} g(u)\, du < \infty$. We will denote $Y \sim \mathrm{El}_q(\mu, \Sigma, g) \equiv \mathrm{El}_q(\mu, \Sigma)$. It is possible to show that the characteristic function is $\psi(t) = \mathrm{E}(\exp(it^\top Y)) = \exp(it^\top \mu)\varphi(t^\top \Sigma t)$, where $t \in \mathbb{R}^q$ and $\varphi : [0, \infty) \to \mathbb{R}$. Then, if $\varphi$ is twice differentiable at zero, we have that $\mathrm{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \xi\Sigma$, where $\xi = \varphi'(0)$. We assume that the density generating function $g$ does not have any unknown parameter, which implies that $\xi$ is a known constant. From (3.1), when $\mu = 0$ and $\Sigma = I_q$, where $I_q$ is a $q \times q$ identity matrix, we obtain the spherical family of densities. A comprehensive exposition of the elliptical multivariate class of distributions can be found in Fang, Kotz and Ng (1990). Table 1 presents the density generating functions of some multivariate elliptical distributions.

Let $Y_1, Y_2, \ldots, Y_n$ be $n$ independent random vectors, where $Y_i$ has dimension $q_i \in \mathbb{N}$, for $i = 1, 2, \ldots, n$. The general multivariate elliptical model (Lemonte and Patriota, 2011) assumes that

$$Y_i = \mu_i(\theta, x_i) + e_i, \qquad i = 1, \ldots, n,$$

with $e_i \overset{\mathrm{ind}}{\sim} \mathrm{El}_{q_i}(0, \Sigma_i(\theta, w_i))$, where "$\overset{\mathrm{ind}}{\sim}$" means "independently distributed as", $x_i$ and $w_i$ are $m_i \times 1$ and $k_i \times 1$ nonstochastic vectors of auxiliary variables, respectively, associated with the $i$th observed response $Y_i$, which may have components in common. Then,

$$Y_i \overset{\mathrm{ind}}{\sim} \mathrm{El}_{q_i}(\mu_i, \Sigma_i), \qquad i = 1, \ldots, n, \tag{3.2}$$

**Table 1** *Generating functions of some multivariate elliptical distributions*

| Distribution | Generating function $g(u)$ |
|---|---|
| Normal | $\frac{1}{(\sqrt{2\pi})^q} e^{-u/2}$ |
| Cauchy | $\frac{\Gamma(\frac{1+q}{2})}{\Gamma(\frac{1}{2})} \pi^{-q/2} (1+u)^{-(1+q)/2}$ |
| Student $t$ | $\frac{\Gamma(\frac{\nu+q}{2})}{\Gamma(\frac{\nu}{2})} \pi^{-q/2} \nu^{-q/2} (1+\frac{u}{\nu})^{-(\nu+q)/2}, \nu > 0$ |
| Power exponential | $\frac{\lambda\Gamma(\frac{q}{2})}{\Gamma(\frac{q}{2\lambda})} 2^{-q/(2\lambda)} \pi^{-q/2} e^{-u^\lambda/2}, \lambda > 0$ |

where $\mu_i = \mu_i(\theta, x_i)$ is the location parameter and $\Sigma_i = \Sigma_i(\theta, w_i)$ is the definite positive scale matrix. Both $\mu_i$ and $\Sigma_i$ have known functional forms and are twice differentiable with respect to each element of $\theta$. Additionally, $\theta$ is a $p$-vector of unknown parameters (where $p < n$ and it is fixed). Since $\theta$ must be identifiable in model (3.2), the functions $\mu_i$ and $\Sigma_i$ must be defined to accomplish such restriction.

Several important statistical models are special cases of the general formulation (3.2), for example, linear and nonlinear regression models, homoscedastic or heteroscedastic measurement error models, and mixed-effects models with normal errors. It is noteworthy that the normality assumption for the errors may be relaxed and replaced by any distribution within the class of elliptical distributions, such as the Student $t$ and the power exponential distributions. The general formulation allows a wide range of different specifications for the location and the scale parameters, coupled with a large collection of distributions for the errors. Section 4 presents four important particular cases of the main model (3.2) that show the applicability of the general formulation.

For the sake of simplifying the notation, let $z_i = Y_i - \mu_i$ and $u_i = z_i^\top \Sigma_i^{-1} z_i$. The log-likelihood function associated with (3.2), is given by

$$\ell(\theta) = \sum_{i=1}^{n} \ell_i(\theta), \qquad (3.3)$$

where $\ell_i(\theta) = -\frac{1}{2} \log |\Sigma_i| + \log g(u_i)$. It is assumed that $g(\cdot)$, $\mu_i$ and $\Sigma_i$ are such that $\ell(\theta)$ is a regular log-likelihood function (Cox and Hinkley, 1974, Ch. 9) with respect to $\theta$. To obtain the score function and the Fisher information matrix, we need to derive $\ell(\theta)$ with respect to the unknown parameters and to compute some moments of such derivatives. We assume that such derivatives exist. Thus, we define

$$a_{i(r)} = \frac{\partial \mu_i}{\partial \theta_r}, \qquad a_{i(sr)} = \frac{\partial^2 \mu_i}{\partial \theta_s \, \partial \theta_r}, \qquad C_{i(r)} = \frac{\partial \Sigma_i}{\partial \theta_r}, \qquad C_{i(sr)} = \frac{\partial^2 \Sigma_i}{\partial \theta_s \, \partial \theta_r}$$

and

$$A_{i(r)} = -\Sigma_i^{-1} C_{i(r)} \Sigma_i^{-1},$$

for $r, s = 1, \ldots, p$. We make use of matrix differentiation methods (Magnus and Neudecker, 2007) to compute the derivatives of the log-likelihood function. The score vector and the Fisher information matrix for $\theta$ can be shortly written as

$$U(\theta) = F^\top H s \quad \text{and} \quad K(\theta) = F^\top \widetilde{H} F, \tag{3.4}$$

respectively, with $F = (F_1^\top, \ldots, F_n^\top)^\top$, $H = \text{block-diag}\{H_1, \ldots, H_n\}$, $s = (s_1^\top, \ldots, s_n^\top)^\top$, $\widetilde{H} = HMH$ and $M = \text{block-diag}\{M_1^\top, \ldots, M_n^\top\}$, wherein

$$F_i = \begin{pmatrix} D_i \\ V_i \end{pmatrix}, \qquad H_i = \begin{bmatrix} \Sigma_i & 0 \\ 0 & 2\Sigma_i \otimes \Sigma_i \end{bmatrix}^{-1}, \qquad s_i = \begin{bmatrix} v_i z_i \\ -\text{vec}(\Sigma_i - v_i z_i z_i^\top) \end{bmatrix},$$

where the "vec" operator transforms a matrix into a vector by stacking the columns of the matrix, $D_i = (a_{i(1)}, \ldots, a_{i(p)})$, $V_i = (\text{vec}(C_{i(1)}), \ldots, \text{vec}(C_{i(p)}))$, $v_i = -2W_g(u_i)$ and $W_g(u) = d \log g(u)/du$. Here, we assume that $F$ has rank $p$ (i.e., $\mu_i$ and $\Sigma_i$ must be defined to hold such condition). The symbol "$\otimes$" indicates the Kronecker product. Following Lange, Little and Taylor (1989) we have, for the $q$-variate Student $t$ distribution with $\nu$ degrees of freedom, $t_q(\mu, \Sigma, \nu)$, that $W_g(u) = -(\nu + q)/\{2(\nu + u)\}$. Following Gómez, Gómez-Villegas and Martín (1998) we have, for the $q$-variate power exponential $PE_q(\mu, \delta, \lambda)$ with shape parameter $\lambda > 0$ and $u \neq 0$, that $W_g(u) = -\lambda u^{\lambda-1}/2$, $\lambda \neq 1/2$. In addition, we have

$$M_i = \begin{bmatrix} \dfrac{4\psi_{i(2,1)}}{q_i} \Sigma_i & 0 \\ 0 & 2c_i \Sigma_i \otimes \Sigma_i \end{bmatrix} + (c_i - 1) \begin{bmatrix} 0 & 0 \\ 0 & \text{vec}(\Sigma_i) \text{vec}(\Sigma_i)^\top \end{bmatrix},$$

where $c_i = 4\psi_{i(2,2)}/\{q_i(q_i + 2)\}$, $\psi_{i(2,1)} = E(W_g^2(r_i)r_i)$ and $\psi_{i(2,2)} = E(W_g^2(r_i)r_i^2)$, with $r_i = ||L_i||^2$, $L_i \sim \text{El}_{q_i}(0, I_{q_i})$. Here, we assume that $g(u)$ is such that $\psi_{i(2,1)}$ and $\psi_{i(2,2)}$ exist and are finite for all $i = 1, \ldots, n$. One can verify these results by using standard differentiation techniques and some standard matrix operations.

The values of $\psi_{i(l,k)}$ are obtained from solving the following one-dimensional integrals (Lange, Little and Taylor, 1989):

$$\psi_{i(l,k)} = \int_0^\infty W_g(s^2)^l g(s^2) r^{q_i+2k-1} c_{q_i} \, ds, \tag{3.5}$$

where $c_{q_i} = 2\pi^{\frac{q_i}{2}} / \Gamma(\frac{q_i}{2})$ is the surface area of the unit sphere in $\mathbb{R}^{q_i}$ and $\Gamma(a)$ is the well-known gamma function. One can find these quantities for many distributions simply solving (3.5) algebraically or numerically. Table 2 shows these quantities for the normal, Cauchy, Student $t$ and power exponential distributions.

It is important to remark that the $\psi_{i(l,k)}$'s may involve unknown quantities (for instance, the degrees of freedom $\nu$ of the Student $t$ distribution and the shape parameter $\lambda$ of the power exponential distribution). One may want to estimate these quantities via maximum likelihood estimation. Here, we consider these as known quantities for the purpose of keeping the robustness property of some distributions.

**Table 2** *Functions $\psi_{i(2,1)}$, $\psi_{i(2,2)}$, $\psi_{i(3,2)}$, $\psi_{i(3,3)}$ for normal, Cauchy, Student t and power exponential (P.E.) distributions*

| | $\psi_{i(2,1)}$ | $\psi_{i(2,2)}$ | $\psi_{i(3,2)}$ | $\psi_{i(3,3)}$ |
|---|---|---|---|---|
| Normal | $\frac{q_i}{4}$ | $\frac{q_i(q_i+2)}{4}$ | $-\frac{q_i(q_i+2)}{8}$ | $-\frac{q_i(q_i+2)(q_i+4)}{8}$ $\quad q_i \geq 1$ |
| Cauchy | $\frac{q_i(q_i+1)}{4(q_i+3)}$ | $\frac{q_i(q_i+2)(q_i+1)}{4(q_i+3)}$ | $-\frac{q_i(q_i+2)(q_i+1)^2}{8(q_i+3)(q_i+5)}$ | $-\frac{q_i(q_i+2)(q_i+4)(q_i+1)^2}{8(q_i+3)(q_i+5)}$ $\quad q_i \geq 1$ |
| Student $t$ | $\frac{q_i(q_i+\nu)}{4(q_i+\nu+2)}$ | $\frac{q_i(q_i+2)(q_i+\nu)}{4(q_i+\nu+2)}$ | $-\frac{q_i(q_i+2)(q_i+\nu)^2}{8(q_i+2+\nu)(q_i+4+\nu)}$ | $-\frac{q_i(q_i+2)(q_i+4)(q_i+\nu)^2}{8(q_i+2+\nu)(q_i+4+\nu)}$ $\quad q_i \geq 1$ |
| P.E. | $\frac{\lambda^2\Gamma(\frac{4\lambda-1}{2\lambda})}{2^{1/\lambda}\Gamma(\frac{1}{2\lambda})}$ | $\frac{2\lambda+1}{4}$ | $-\frac{\lambda^3\Gamma(\frac{6\lambda-1}{2\lambda})}{2^{1/\lambda}\Gamma(\frac{1}{2\lambda})}$ | $-\frac{(2\lambda+1)(4\lambda+1)}{8}$ $\quad q_i = 1, \lambda > \frac{1}{4}$ |
| P.E. | $\frac{\lambda^2\Gamma(\frac{q_i-2}{2\lambda}+2)}{2^{1/\lambda}\Gamma(\frac{q_i}{2\lambda})}$ | $\frac{q_i(2\lambda+q_i)}{4}$ | $-\frac{\lambda^3\Gamma(\frac{q_i-2}{2\lambda}+3)}{2^{1/\lambda}\Gamma(\frac{q_i}{2\lambda})}$ | $-\frac{q_i(2\lambda+q_i)(4\lambda+q_i)}{8}$ $\quad q_i \geq 2, \lambda > 0$ |

Lucas (1997) shows that the protection against "large" observations is only valid when the degrees of freedom parameter is kept fixed for the Student $t$ distribution. Therefore, the issue of estimating these quantities is beyond of the main scope of this paper. In practice, one can use model selection procedures to choose the most appropriate values of such unknown parameters.

Notice that, in the Fisher information matrix $K(\theta)$, the matrix $M$ carries all the information about the adopted distribution, while $F$ and $H$ contain the information about the adopted model. Also, $K(\theta)$ has a qquadratic form that can be computed through simple matrix operations. Under the normal case, $v_i = 1$, $M = H^{-1}$ and hence $\widetilde{H} = H$.

The Fisher scoring method can be used to estimate $\theta$ by iteratively solving the equation

$$\left(F^{(m)\top}\widetilde{H}^{(m)}F^{(m)}\right)\theta^{(m+1)} = F^{(m)\top}\widetilde{H}^{(m)}s^{*(m)}, \qquad m = 0, 1, \ldots, \qquad (3.6)$$

where the quantities with the upper index "$(m)$" are evaluated at $\widehat{\theta}$, $m$ is the iteration counter and

$$s^{*(m)} = F^{(m)}\theta^{(m)} + H^{-1(m)}M^{-1(m)}s^{(m)}.$$

Each loop, through the iterative scheme (3.6), consists of an iterative re-weighted least squares algorithm to optimize the log-likelihood (3.3). Thus, (3.4) and (3.6) agree with the corresponding equations derived in Patriota and Lemonte (2009). Observe that, despite the complexity and generality of the postulated model, expressions (3.4) and (3.6) are very simple and friendly.

Now, we can give the main result of the paper.

**Theorem 3.1.** *The second-order bias vector $B_{\widehat{\theta}}(\theta)$ under model (3.2) is given by*

$$B_{\widehat{\theta}}(\theta) = \left(F^\top\widetilde{H}F\right)^{-1}F^\top\widetilde{H}\xi, \qquad (3.7)$$

*where $\xi = (\Phi_1, \ldots, \Phi_p)\,\mathrm{vec}((F^\top\widetilde{H}F)^{-1})$, $\Phi_r = (\Phi_{1(r)}^\top, \ldots, \Phi_{n(r)}^\top)^\top$, and $\Phi_{i(r)}$ is given in the Appendix.*

**Proof.** See the Appendix. □

We shall emphasize that Theorem 3.1 is only valid under the assumption that no extra unknown parameter is estimated along with $\theta$. For instance, for the Student t distribution, $\nu$ is assumed to be fixed. If extra unknown parameters were to be considered, cumulants of the log-likelihood function with respect to these parameters would appear in (3.7).

In many models the location vector and the scale matrix do not have parameters in common, that is, $\mu_i = \mu_i(\theta_1, x_i)$ and $\Sigma_i = \Sigma_i(\theta_2, w_i)$, where $\theta = (\theta_1^\top, \theta_2^\top)^\top$. Therefore, $F = \text{block-diag}\{F_{\theta_1}, F_{\theta_2}\}$ and the parameter vectors $\theta_1$ and $\theta_2$ will be orthogonal (Cox and Reid, 1987). This happens in mixed models, nonlinear models, among others. However, in errors-in-variables and factor analysis models orthogonality does not hold. Model (3.2) is general enough to encompass a large number of models even those that do not have orthogonal parameters.

**Corollary 3.1.** *When $\mu_i = \mu_i(\theta_1, x_i)$ and $\Sigma_i = \Sigma_i(\theta_2, w_i)$, where $\theta = (\theta_1^\top, \theta_2^\top)^\top$ the second-order bias vector of $\widehat\theta_1$ and $\widehat\theta_2$ are given by*

$$B_{\widehat\theta_1}(\theta) = \left(F_{\theta_1}^\top \widetilde H_1 F_{\theta_1}\right)^{-1} F_{\theta_1}^\top \widetilde H_1 \xi_1$$

*and*

$$B_{\widehat\theta_2}(\theta) = \left(F_{\theta_2}^\top \widetilde H_2 F_{\theta_2}\right)^{-1} F_{\theta_2}^\top \widetilde H_2 \xi_2,$$

*respectively. The quantities $F_{\theta_1}, F_{\theta_2}, \widetilde H_1, \widetilde H_2, \xi_1$ and $\xi_2$ are defined in the Appendix.*

**Proof.** See the Appendix. □

Formula (3.7) says that, for any particular model of the general multivariate elliptical class of models (3.2), it is always possible to express the bias of $\widehat\theta$ as the solution of an weighted least-squares regression. Also, if $z_i \sim N_{q_i}(0, \Sigma_i)$ then $c_i = -\widetilde\omega_i = 1$, $\eta_{1i} = 0$, $\eta_{2i} = -2$, $\widetilde H = H$,

$$J_{i(r)} = \begin{pmatrix} 0 \\ 2(I_{q_i} \otimes a_{i(r)}) D_i \end{pmatrix},$$

and formula (3.7) reduces to the one obtained by Patriota and Lemonte (2009).

Theorem 3.1 implies that all one needs to compute bias-corrected and bias-reduced MLEs in the general elliptical model is: (i) the first and second derivatives of the location vector $\mu_i$ and the scale matrix $\Sigma_i$ with respect to all the parameters; (2) the derivatives $W_g(u)$; (3) some moments involving the chosen elliptical distribution (these moments are given in Table 2 for some elliptical distributions). With these quantities, the matrices in (3.7) can be computed and the bias vector can be computed through an weighted least-squares regression.

## 4 Special models

In this section, we present four important particular cases of the main model (3.2). All special cases presented in Patriota and Lemonte (2009) are also special cases of the general multivariate elliptical model defined in this paper.

### 4.1 Heteroscedastic nonlinear models

Consider the univariate heteroscedastic nonlinear model defined by

$$Y_i = f(x_i, \alpha) + e_i, \qquad i = 1, 2, \ldots, n,$$

where $Y_i$ is the response, $x_i$ is a column vector of explanatory variables, $\alpha$ is a column vector $p_1 \times 1$ of unknown parameters and $f$ is a nonlinear function of $\alpha$. Assume that $e_1, e_2, \ldots, e_n$ are independent, with $e_i \sim \text{El}(0, \sigma_i^2)$. Here $\sigma_i^2 = \sigma_i^2(\gamma) = h(\omega_i^\top \gamma)$, where $\gamma$ is a $p_2 \times 1$ vector of unknown parameters. Then

$$Y_i \overset{\text{ind}}{\sim} \text{El}(f(x_i, \alpha), \sigma_i^2),$$

which is a special case of (3.2) with $\theta = (\alpha^\top, \gamma^\top)^\top$, $\mu_i = f(x_i, \alpha)$ and $\Sigma_i = \sigma_i^2$. Here El stands for $\text{El}_1$. Notice that for the heteroscedastic linear model $f(x_i, \alpha) = x_i^\top \alpha$.

The second-order bias vector $B_{\hat\theta}(\theta)$ comes from (3.7), which depends on derivatives of $f(x_i, \alpha)$ and $\sigma_i^2$ with respect to the parameter vector $\theta$. Also, it depends on the quantities $\psi_{i(2,1)}, \psi_{i(2,2)}, \psi_{i(3,2)}, \psi_{i(3,3)}$ (see Table 2) and $W_g(u_i)$ containing information about the adopted distribution.

### 4.2 Nonlinear mixed-effects model

One of the most important examples is the nonlinear mixed-effects model introduced by Lange, Little and Taylor (1989) and studied under the assumption of a Student $t$ distribution. Let

$$Y_i = \mu_i(x_i, \alpha) + Z_i b_i + u_i,$$

where $Y_i$ is the $q_i \times 1$ vector response, $\mu_i$ is a $q_i$-dimensional nonlinear function of $\alpha$, $x_i$ is a vector of nonstochastic covariates, $Z_i$ is a matrix of known constants, $\alpha$ is a $p_1 \times 1$ vector of unknown parameters and $b_i$ is an $r \times 1$ vector of unobserved random regression coefficients. Assume that,

$$\begin{pmatrix} b_i \\ u_i \end{pmatrix} \sim \text{El}_{r+q_i} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_b(\gamma_1) & 0 \\ 0 & R_i(\gamma_2) \end{bmatrix} \right),$$

where $\gamma_1$ is a $p_2$-dimensional vector of unknown parameters and $\gamma_2$ is a $p_3 \times 1$ vector of unknown parameters. Furthermore, the vectors $(b_1, u_1)^\top$, $(b_2, u_2)^\top$, $\ldots, (b_n, u_n)^\top$ are independent. Therefore, the marginal distribution of the observed vector is

$$Y_i \sim \text{El}_{q_i}(\mu_i(x_i, \alpha); \Sigma_i(Z_i, \gamma)), \tag{4.1}$$

where $\gamma = (\gamma_1^\top, \gamma_2^\top)^\top$ and $\Sigma_i(Z_i, \gamma) = Z_i \Sigma_b(\gamma_1) Z_i^\top + R_i(\gamma_2)$. Equation (4.1) is a special case of (3.2) with $\theta = (\alpha^\top, \gamma^\top)^\top$, $\mu_i = \mu_i(x_i, \alpha)$ and $\Sigma_i = \Sigma_i(Z_i, \gamma)$. From (3.7), one can compute the bias vector $B_{\hat{\theta}}(\theta)$.

## 4.3 Errors-in-variables model

Consider the model

$$x_{1i} = \beta_0 + \beta_1 x_{2i} + q_i, \qquad i = 1, \ldots, n,$$

where $x_{1i}$ is a $v \times 1$ latent response vector, $x_{2i}$ is a $m \times 1$ latent vector of covariates, $\beta_0$ is a $v \times 1$ vector of intercepts, $\beta_1$ is a $v \times m$ matrix of slopes, and $q_i$ is the equation error having a multivariate elliptical distribution with location vector zero and scale matrix $\Sigma_q$. The variables $x_{1i}$ and $x_{2i}$ are not directly observed, instead surrogate variables $X_{1i}$ and $X_{2i}$ are measured with the following additive structure:

$$X_{1i} = x_{1i} + \delta_{x_{1i}} \quad \text{and} \quad X_{2i} = x_{2i} + \delta_{x_{2i}}. \tag{4.2}$$

The random quantities $x_{2i}$, $q_i$, $\delta_{x_{1i}}$ and $\delta_{x_{2i}}$ are assumed to follow an elliptical distribution given by

$$\begin{pmatrix} x_{2i} \\ q_i \\ \delta_{x_{1i}} \\ \delta_{x_{2i}} \end{pmatrix} \overset{\text{ind}}{\sim} \text{El}_{2v+2m} \left[ \begin{pmatrix} \mu_{x_2} \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{x_2} & 0 & 0 & 0 \\ 0 & \Sigma_q & 0 & 0 \\ 0 & 0 & \tau_{x_{1i}} & 0 \\ 0 & 0 & 0 & \tau_{x_{2i}} \end{pmatrix} \right],$$

where the matrices $\tau_{xi}$ and $\tau_{zi}$ are known for all $i = 1, \ldots, n$. These "known" matrices may be attained, for example, through an analytical treatment of the data collection mechanism, replications, machine precision, etc. (Kulathinal, Kuulasmaa and Gasbarra (2002)).

Therefore, the observed vector $Y_i = (X_{1i}^\top, X_{2i}^\top)^\top$ has marginal distribution given by

$$Y_i \overset{\text{ind}}{\sim} \text{El}_{v+m}\big(\mu(\theta), \Sigma_i(\theta)\big) \tag{4.3}$$

with

$$\mu(\theta) = \begin{pmatrix} \beta_0 + \beta_1 \mu_{x_2} \\ \mu_{x_2} \end{pmatrix} \quad \text{and} \quad \Sigma_i(\theta) = \begin{pmatrix} \beta_1 \Sigma_{x_2} \beta_1^\top + \Sigma_q + \tau_{x_{1i}} & \beta_1 \Sigma_{x_2} \\ \Sigma_{x_2} \beta_1^\top & \Sigma_{x_2} + \tau_{x_{2i}} \end{pmatrix},$$

where $\theta = (\beta_0^\top, \text{vec}(\beta_1)^\top, \mu_{x_2}^\top, \text{vech}(\Sigma_{x_2})^\top, \text{vech}(\Sigma_q)^\top)^\top$, "vech" operator transforms a symmetric matrix into a vector by stacking into columns its diagonal and superior diagonal elements. The mean vector $(\theta)$ and the covariance-variance matrix $\Sigma_i(\theta)$ of observed variables have the matrix $\beta_1$ in common, that is, they share $mv$ parameters. Kulathinal, Kuulasmaa and Gasbarra (2002) study the linear univariate case ($v = 1$, $m = 1$).

Equation (4.3) is a special case of (3.2) with $q_i = v + m$, $\theta = (\alpha^\top, \gamma^\top)^\top$, $\mu_i = \mu_i(\theta)$ and $\Sigma_i = \Sigma_i(\theta)$. In this case, a programming language or software that

can perform operations on vectors and matrices, for example, Ox (Doornik, 2013) and R (Ihaka and Gentleman, 1996), can be used to obtain the bias vector $B_{\widehat{\theta}}(\theta)$ from (3.7).

## 4.4 Log-symmetric regression models

Let $T$ be a continuous positive random variable with probability density function

$$f_T(t; \eta, \phi, g) = \frac{1}{\sqrt{\phi}t} g\left(\log^2\left[\left(\frac{t}{\eta}\right)^{\frac{1}{\sqrt{\phi}}}\right]\right), \qquad \eta > 0, \phi > 0, \qquad (4.4)$$

where $g$ is the density generating function of a univariate elliptical distribution, and we write $T \sim LS(\eta, \phi, g)$. Vanegas and Paula (2016) called the class of distribution in (4.4) the log-symmetric class of distributions. It includes log-normal, log-Student $t$, log-power-exponential distributions, among many others, as special cases. It is easy to verify that $\log(T)$ has a univariate elliptical distribution (i.e., symmetric distribution) with location parameter $\mu = \log(\eta)$ and scale parameter $\phi$. The parameter $\eta$ is the median of $T$, and $\phi$ can be interpreted as a skewness or relative dispersion parameter.

Vanegas and Paula (2015) defined and studied semi-parametric regression models for a set $T_1, T_2, \ldots, T_n$ with $T_i \sim LS(\eta_i, \phi_i, g)$ with $\eta_i > 0$ and $\phi_i > 0$ following semi-parametric regression structures. Here we assume parametric specification for $\eta_i$ and $\phi_i$ as $\eta_i = \eta_i(x_i, \alpha)$ and $\phi_i = \phi_i(\omega_i, \gamma)$.

Hence,

$$Y_i = \log(T_i) \stackrel{\text{ind}}{\sim} \text{El}\big(\mu_i(x_i, \alpha), \phi_i(\omega_i, \gamma)\big), \qquad (4.5)$$

where $\mu_i(x_i, \alpha) = \log(\eta(x_i, \alpha))$. Therefore, (4.5) is a special case of the general elliptical model (3.2), and formula (3.7) applies.

# 5 Simulation results

In this section, we shall present the results of Monte Carlo simulation experiments in which we evaluate the finite sample performances of the original MLEs and their bias-corrected and bias-reduced versions. The simulations are based on the univariate nonlinear model without random effects (Section 4.1) and the errors-in-variables model presented in Section 4.2, when $Y_i$ follows a normal distribution, a Student $t$ distribution with $\nu$ degrees of freedom, or a power exponential distribution with shape parameter $\lambda$. For all the simulations, the number of Monte Carlo replications is 10,000 (ten thousand) and they have been performed using the Ox matrix programming language (Doornik, 2013).

First, consider the model described in (4.1) with $q_i = 1$, $Z_i = 0$, $\Sigma_i = \sigma^2$ and

$$\mu_i(\alpha) = \mu_i(x_i, \alpha) = \alpha_1 + \frac{\alpha_2}{1 + \alpha_3 x_i^{\alpha_4}}, \qquad i = 1, \ldots, n. \qquad (5.1)$$

Here the unknown parameter vector is $\theta = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \sigma^2)^\top$. The values of $x_i$ were obtained as random draws from the uniform distribution $U(0, 100)$. The sample sizes considered are $n = 10, 20, 30, 40$ and $50$. The parameter values are $\alpha_1 = 50$, $\alpha_2 = 500$, $\alpha_3 = 0.50$, $\alpha_4 = 2$ and $\sigma_i^2 = 200$. For the Student $t$ distribution, we fixed the degrees of freedom at $\nu = 4$, and for the power exponential model the shape parameter is fixed at $\lambda = 0.8$.

Tables 3–4 present the bias, and the root mean squared errors ($\sqrt{\text{MSE}}$) of the maximum likelihood estimates, the bias-corrected estimates and the bias-reduced estimates for the nonlinear model with normal and Student $t$ distributed errors, respectively. To save space, the corresponding results for the power exponential

**Table 3** *Biases and $\sqrt{\text{MSE}}$ of the maximum likelihood estimate and its adjusted versions*; *nonlinear model*; *normal distribution*

| $n$ | $\theta$ | MLE | | Bias-corrected MLE | | Bias-reduced MLE | |
|---|---|---|---|---|---|---|---|
| | | Bias | $\sqrt{\text{MSE}}$ | Bias | $\sqrt{\text{MSE}}$ | Bias | $\sqrt{\text{MSE}}$ |
| 10 | $\alpha_1$ | −0.29 | 6.69 | −0.13 | 6.67 | −0.01 | 6.67 |
| | $\alpha_2$ | 2.16 | 20.07 | 0.70 | 19.40 | −0.27 | 19.06 |
| | $\alpha_3$ | 0.01 | 0.13 | 0.00 | 0.12 | 0.00 | 0.12 |
| | $\alpha_4$ | 0.03 | 0.30 | 0.01 | 0.29 | −0.00 | 0.29 |
| | $\sigma^2$ | −80.05 | 106.44 | −32.06 | 103.32 | 9.09 | 128.72 |
| 20 | $\alpha_1$ | −0.08 | 4.07 | −0.01 | 4.07 | 0.01 | 4.07 |
| | $\alpha_2$ | 0.66 | 17.94 | −0.08 | 17.84 | −0.27 | 17.82 |
| | $\alpha_3$ | 0.00 | 0.09 | 0.00 | 0.09 | −0.00 | 0.09 |
| | $\alpha_4$ | 0.02 | 0.21 | 0.01 | 0.20 | 0.00 | 0.20 |
| | $\sigma^2$ | −40.07 | 69.73 | −8.09 | 68.95 | 0.86 | 72.02 |
| 30 | $\alpha_1$ | −0.10 | 3.11 | −0.04 | 3.10 | −0.02 | 3.10 |
| | $\alpha_2$ | 0.71 | 17.24 | −0.05 | 17.15 | −0.18 | 17.13 |
| | $\alpha_3$ | 0.00 | 0.09 | −0.00 | 0.09 | −0.00 | 0.09 |
| | $\alpha_4$ | 0.02 | 0.20 | 0.00 | 0.19 | 0.00 | 0.19 |
| | $\sigma^2$ | −26.41 | 55.26 | −3.26 | 55.11 | 0.82 | 56.32 |
| 40 | $\alpha_1$ | −0.08 | 2.69 | −0.02 | 2.69 | −0.01 | 2.69 |
| | $\alpha_2$ | 0.83 | 16.80 | 0.09 | 16.70 | 0.01 | 16.69 |
| | $\alpha_3$ | 0.00 | 0.09 | 0.00 | 0.09 | 0.00 | 0.09 |
| | $\alpha_4$ | 0.02 | 0.19 | 0.00 | 0.18 | −0.00 | 0.18 |
| | $\sigma^2$ | −20.04 | 47.26 | −2.04 | 47.13 | 0.33 | 47.74 |
| 50 | $\alpha_1$ | −0.08 | 2.39 | −0.03 | 2.38 | −0.02 | 2.38 |
| | $\alpha_2$ | 1.07 | 14.25 | 0.30 | 14.12 | 0.23 | 14.11 |
| | $\alpha_3$ | 0.00 | 0.08 | 0.00 | 0.08 | 0.00 | 0.08 |
| | $\alpha_4$ | 0.01 | 0.19 | 0.00 | 0.18 | −0.00 | 0.18 |
| | $\sigma^2$ | −15.93 | 41.41 | −1.21 | 41.30 | 0.36 | 41.67 |

**Table 4**  *Biases and $\sqrt{\text{MSE}}$ of the maximum likelihood estimate and its adjusted versions*; *nonlinear model*; *Student t distribution*

| $n$ | $\theta$ | MLE | | Bias-corrected MLE | | Bias-reduced MLE | |
|---|---|---|---|---|---|---|---|
| | | Bias | $\sqrt{\text{MSE}}$ | Bias | $\sqrt{\text{MSE}}$ | Bias | $\sqrt{\text{MSE}}$ |
| 10 | $\alpha_1$ | −0.51 | 8.66 | −0.31 | 8.63 | −0.20 | 8.56 |
| | $\alpha_2$ | 3.34 | 28.47 | 1.39 | 27.34 | 2.05 | 27.67 |
| | $\alpha_3$ | 0.01 | 0.17 | 0.00 | 0.16 | 0.01 | 0.16 |
| | $\alpha_4$ | 0.06 | 0.42 | 0.03 | 0.39 | −0.01 | 0.38 |
| | $\sigma^2$ | −93.18 | 127.60 | −54.24 | 130.73 | −17.40 | 170.35 |
| 20 | $\alpha_1$ | −0.17 | 5.03 | −0.07 | 5.02 | −0.04 | 5.01 |
| | $\alpha_2$ | 2.01 | 25.64 | 0.91 | 25.11 | 1.29 | 24.98 |
| | $\alpha_3$ | 0.01 | 0.14 | 0.01 | 0.14 | 0.01 | 0.13 |
| | $\alpha_4$ | 0.04 | 0.29 | 0.01 | 0.28 | 0.00 | 0.27 |
| | $\sigma^2$ | −41.24 | 85.51 | −12.30 | 89.41 | −4.55 | 93.08 |
| 30 | $\alpha_1$ | −0.10 | 3.81 | −0.01 | 3.80 | 0.01 | 3.82 |
| | $\alpha_2$ | 2.25 | 25.75 | 1.13 | 25.34 | 1.61 | 25.41 |
| | $\alpha_3$ | 0.01 | 0.14 | 0.01 | 0.14 | 0.01 | 0.14 |
| | $\alpha_4$ | 0.04 | 0.29 | 0.01 | 0.27 | 0.00 | 0.26 |
| | $\sigma^2$ | −27.15 | 70.02 | −6.15 | 72.64 | −1.78 | 107.53 |
| 40 | $\alpha_1$ | −0.10 | 3.27 | −0.02 | 3.26 | −0.01 | 3.26 |
| | $\alpha_2$ | 1.82 | 24.94 | 0.75 | 24.67 | 1.18 | 24.78 |
| | $\alpha_3$ | 0.01 | 0.12 | 0.00 | 0.12 | 0.01 | 0.12 |
| | $\alpha_4$ | 0.03 | 0.26 | 0.01 | 0.25 | 0.00 | 0.25 |
| | $\sigma^2$ | −20.38 | 60.43 | −4.01 | 62.21 | −1.82 | 62.98 |
| 50 | $\alpha_1$ | −0.13 | 2.86 | −0.05 | 2.85 | −0.03 | 2.85 |
| | $\alpha_2$ | 1.48 | 18.86 | 0.38 | 18.59 | 0.24 | 18.46 |
| | $\alpha_3$ | 0.01 | 0.11 | 0.00 | 0.11 | 0.00 | 0.11 |
| | $\alpha_4$ | 0.02 | 0.24 | 0.00 | 0.23 | 0.00 | 0.23 |
| | $\sigma^2$ | −15.40 | 53.99 | −1.94 | 55.56 | −0.43 | 56.11 |

model are not shown.[1] We note that the bias-corrected estimates and the bias-reduced estimates are less biased than the original MLE for all the sample sizes considered. For instance, when $n = 20$ and the errors follow a Student $t$ distribution (see Table 4) the estimated biases of $\widehat{\sigma}^2$ are −41.24 (MLE), −12.30 (bias-corrected) and −4.55 (bias-reduced). For the normal case with $n = 10$ (see Table 3), the estimated biases of $\widehat{\alpha}_2$ are 2.16 (MLE), 0.70 (bias-corrected) and −0.27 (bias-reduced). We also observe that the bias-reduced estimates are less biased than the bias-corrected estimates in most cases. As $n$ increases, the bias and the root mean squared error of all the estimators decrease, as expected. Additionally, we note that the MLE of $\alpha_2$ has $\sqrt{\text{MSE}}$ larger than those of the modified versions.

---

[1]All the omitted tables in this paper are presented in a supplement available from the authors upon request.

For the estimation of $\sigma^2$, $\sqrt{\text{MSE}}$ is smaller for the original MLE. In other cases, we note that the estimators have similar root mean squared errors.

We now consider the errors-in-variables model described in (4.2). The sample sizes considered are $n = 15, 25, 35$ and $50$. The parameter values are $\beta_0 = 0.70$ $\mathbf{1}_{v \times 1}$, $\beta_1 = 0.40 \, \mathbf{1}_{v \times m}$, $\mu_{x_2} = 70 \, \mathbf{1}_{m \times 1}$, $\Sigma_q = 40 \, \mathbf{I}_{v \times 1}$ and $\Sigma_{x_2} = 250 \, \mathbf{I}_{m \times 1}$. Here, $\mathbf{1}_{r \times s}$ is as $r \times s$ matrix of ones and $\mathbf{I}_{r \times s}$ is the identity matrix with dimension $r \times s$. For the Student $t$ distribution, we fixed the degrees of freedom at $v = 4$ and, for power exponential model, the shape parameter was fixed at $\lambda = 0.7$. We consider $v \in \{1, 2\}$ and $m = 1$.

In Tables 5–6, we present the MLE, the bias-corrected estimates, the bias-reduced estimates, and corresponding estimated root mean squared errors for the Student $t$ and power exponential distributions, for the errors-in-variables model. The results for the normal distribution are not shown to save space. We observe that, in absolute value, the biases of the bias-corrected estimates and bias-reduced estimates are smaller than those of the original MLE for different sample sizes. Furthermore, the bias-reduced estimates are less biased than the bias-corrected estimates in most cases. This can be seen for example, in Table 6 when $v = 1, m = 1$,

**Table 5** *Biases and* $\sqrt{\text{MSE}}$ *of the maximum likelihood estimate and its adjusted versions; errors-in-variables model;* $v = 1$ *and* $m = 1$; *Student t distribution*

| | | MLE | | Bias-corrected MLE | | Bias-reduced MLE | |
|---|---|---|---|---|---|---|---|
| $n$ | $\theta$ | Bias | $\sqrt{\text{MSE}}$ | Bias | $\sqrt{\text{MSE}}$ | Bias | $\sqrt{\text{MSE}}$ |
| 15 | $\beta_0$ | $-0.00$ | 9.90 | 0.01 | 9.90 | 0.01 | 9.89 |
| | $\beta_1$ | 0.00 | 0.14 | 0.00 | 0.14 | 0.00 | 0.14 |
| | $\mu_{x_2}$ | 0.05 | 4.82 | 0.05 | 4.82 | 0.05 | 4.82 |
| | $\Sigma_{x_2}$ | 5.18 | 129.34 | 2.91 | 128.12 | 2.66 | 127.86 |
| | $\Sigma_q$ | $-3.64$ | 19.52 | $-0.68$ | 20.72 | $-0.42$ | 20.85 |
| 25 | $\beta_0$ | $-0.02$ | 7.14 | $-0.01$ | 7.14 | $-0.01$ | 7.14 |
| | $\beta_1$ | 0.00 | 0.10 | 0.00 | 0.10 | 0.00 | 0.10 |
| | $\mu_{x_2}$ | 0.03 | 3.69 | 0.03 | 3.69 | 0.03 | 3.69 |
| | $\Sigma_{x_2}$ | 3.61 | 97.32 | 2.25 | 96.76 | 2.17 | 96.69 |
| | $\Sigma_q$ | $-2.31$ | 14.87 | $-0.47$ | 15.41 | $-0.38$ | 15.44 |
| 35 | $\beta_0$ | $-0.02$ | 5.93 | $-0.02$ | 5.93 | $-0.02$ | 5.93 |
| | $\beta_1$ | 0.00 | 0.08 | 0.00 | 0.08 | 0.00 | 0.08 |
| | $\mu_{x_2}$ | $-0.01$ | 3.12 | $-0.01$ | 3.12 | $-0.01$ | 3.12 |
| | $\Sigma_{x_2}$ | 1.94 | 79.78 | 0.98 | 79.45 | 0.94 | 79.44 |
| | $\Sigma_q$ | $-1.65$ | 12.63 | $-0.31$ | 12.96 | $-0.26$ | 12.97 |
| 50 | $\beta_0$ | $-0.01$ | 4.92 | $-0.01$ | 4.92 | $-0.01$ | 4.92 |
| | $\beta_1$ | 0.00 | 0.07 | 0.00 | 0.07 | 0.00 | 0.07 |
| | $\mu_{x_2}$ | 0.01 | 2.59 | 0.01 | 2.59 | 0.01 | 2.59 |
| | $\Sigma_{x_2}$ | 1.04 | 65.50 | 0.37 | 65.33 | 0.36 | 65.33 |
| | $\Sigma_q$ | $-1.18$ | 10.53 | $-0.24$ | 10.78 | $-0.21$ | 10.79 |

**Table 6** *Biases and $\sqrt{\text{MSE}}$ of the maximum likelihood estimate and its adjusted versions*; *errors-in-variables model*; $v = 1$ *and* $m = 1$; *power exponential distribution*

| | | MLE | | Bias-corrected MLE | | Bias-reduced MLE | |
|---|---|---|---|---|---|---|---|
| $n$ | $\theta$ | Bias | $\sqrt{\text{MSE}}$ | Bias | $\sqrt{\text{MSE}}$ | Bias | $\sqrt{\text{MSE}}$ |
| 15 | $\beta_0$ | −0.12 | 9.25 | −0.11 | 9.25 | −0.11 | 9.24 |
| | $\beta_1$ | 0.00 | 0.13 | 0.00 | 0.13 | 0.00 | 0.13 |
| | $\mu_{x_2}$ | −0.02 | 6.47 | −0.02 | 6.47 | −0.02 | 6.47 |
| | $\Sigma_{x_2}$ | −9.27 | 103.32 | 0.52 | 107.51 | 0.82 | 107.64 |
| | $\Sigma_q$ | −4.92 | 15.67 | −0.66 | 17.55 | −0.17 | 17.76 |
| 25 | $\beta_0$ | 0.02 | 6.83 | 0.03 | 6.83 | 0.03 | 6.83 |
| | $\beta_1$ | 0.00 | 0.09 | −0.00 | 0.09 | −0.00 | 0.09 |
| | $\mu_{x_2}$ | −0.02 | 4.98 | −0.02 | 4.98 | −0.02 | 4.98 |
| | $\Sigma_{x_2}$ | −5.60 | 80.20 | 0.36 | 81.95 | 0.47 | 81.99 |
| | $\Sigma_q$ | −3.04 | 12.94 | −0.36 | 13.49 | −0.18 | 13.54 |
| 35 | $\beta_0$ | 0.01 | 5.59 | 0.02 | 5.58 | 0.02 | 5.58 |
| | $\beta_1$ | −0.00 | 0.08 | −0.00 | 0.08 | −0.00 | 0.08 |
| | $\mu_{x_2}$ | −0.04 | 4.21 | −0.04 | 4.21 | −0.04 | 4.21 |
| | $\Sigma_{x_2}$ | −3.53 | 68.01 | 0.77 | 69.10 | 0.82 | 69.12 |
| | $\Sigma_q$ | −2.14 | 11.11 | −0.18 | 11.46 | −0.08 | 11.49 |
| 50 | $\beta_0$ | 0.03 | 4.67 | 0.03 | 4.67 | 0.03 | 4.67 |
| | $\beta_1$ | −0.00 | 0.06 | −0.00 | 0.06 | −0.00 | 0.06 |
| | $\mu_{x_2}$ | −0.03 | 3.52 | −0.03 | 3.52 | −0.03 | 3.52 |
| | $\Sigma_{x_2}$ | −2.83 | 56.89 | 0.18 | 57.51 | 0.21 | 57.52 |
| | $\Sigma_q$ | −1.51 | 9.21 | −0.12 | 9.41 | −0.07 | 9.42 |

$Y_i$ follows a power exponential distribution and $n = 15$. In this case, the bias of the MLE, the bias-corrected estimate and the bias-reduced estimate of $\Sigma_q$ are −4.92, −0.66 and −0.17, respectively. When $Y_i$ follows a Student $t$ distribution, $n = 15$, $v = 1$ and $m = 1$ we observe the following biases of the estimates of $\Sigma_{x_2}$: 5.18 (MLE), 2.91 (bias-corrected) and 2.66 (bias-reduced); see Table 5. We note that the root mean squared errors decrease with $n$.

For the sake of saving space, the simulation results for the normal, Student $t$ and power exponential errors-in-variable models with $v = 2$ and $m = 1$ are not presented. Overall, our findings are similar to those reached for the other models.

## 6 Applications

### 6.1 Radioimmunoassay data

Tiede and Pagano (1979) present a dataset, referred here as the radioimmunoassay data, obtained from the Nuclear Medicine Department at the Veterans Administration Hospital, Buffalo, New York. Lemonte and Patriota (2011) analyzed the data

**Table 7** *Estimates and standard errors (given in parentheses); radioimmunoassay data*

| $\theta$ | MLE | Bias-corrected MLE | Bias-reduced MLE |
|---|---|---|---|
| | Normal distribution | | |
| $\alpha_1$ | 0.44 (0.80) | 0.65 (0.99) | 1.03 (1.06) |
| $\alpha_2$ | 7.55 (0.95) | 7.34 (1.16) | 6.91 (1.25) |
| $\alpha_3$ | 0.13 (0.06) | 0.13 (0.06) | 0.13 (0.08) |
| $\alpha_4$ | 0.96 (0.24) | 0.93 (0.28) | 0.95 (0.34) |
| $\sigma^2$ | 0.31 (0.12) | 0.40 (0.15) | 0.50 (0.19) |
| | Student $t$ distribution | | |
| $\alpha_1$ | 0.90 (0.12) | 0.91 (0.13) | 0.90 (0.15) |
| $\alpha_2$ | 7.09 (0.17) | 7.08 (0.19) | 7.07 (0.22) |
| $\alpha_3$ | 0.09 (0.01) | 0.09 (0.01) | 0.09 (0.02) |
| $\alpha_4$ | 1.31 (0.08) | 1.31 (0.09) | 1.29 (0.10) |
| $\sigma^2$ | 0.02 (0.01) | 0.02 (0.01) | 0.03 (0.01) |

to illustrate the applicability of the elliptical models with general parameterization. Following Tiede and Pagano (1979), we shall consider the nonlinear regression model (5.1), with $n = 14$. The response variable is the observed radioactivity (count in thousands), the covariate corresponds to the thyrotropin dose (measured in micro-international units per milliliter) and the errors follow a normal distribution or a Student $t$ distribution with $\nu = 4$ degrees of freedom. We assume that the scale parameter is unknown for both models. In Table 7 we present the maximum likelihood estimates, the bias-corrected estimates, the bias-reduced estimates, and the corresponding estimated standard errors are given in parentheses. We note that all the estimates present smaller standard errors under the Student $t$ model than under the normal model (Table 7).

For all parameters, the original MLEs are very close to the bias-corrected MLE and the bias-reduced MLE when the Student $t$ model is used. However, under the normal model, significant differences in the estimates of $\alpha_1$ are noted. The estimates for $\alpha_1$ are 0.44 (MLE), 0.65 (bias-corrected MLE) and 1.03 (bias-reduced MLE).

## 6.2 Fluorescent lamp data

Rosillo and Chivelet (2009) present a dataset referred here as the fluorescent lamp data. The authors analyze the lifetime of fluorescent lamps in photovoltaic systems using an analytical model whose goal is to assist in improving ballast design and extending the lifetime of fluorescent lamps. Following Rosillo and Chivelet (2009), we shall consider the nonlinear regression model (4.1) with $q_i = 1$, $Z_i = 0$, $\Sigma_i = \sigma^2$, $\theta = (\alpha^\top, \sigma^2)^\top = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \sigma^2)^\top$ and

$$\mu_i(\alpha) = \frac{1}{1 + \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i2}^2}, \qquad i = 1, \ldots, 14,$$

**Table 8**  *Estimates and standard errors (given in parentheses); fluorescent lamp data*

| $\theta$ | MLE | Bias-corrected MLE | Bias-reduced MLE |
|---|---|---|---|
| | Normal distribution | | |
| $\alpha_0$ | 29.49 (5.21) | 28.54 (5.66) | 28.25 (5.84) |
| $\alpha_1$ | 9.99 (4.69) | 9.68 (5.21) | 9.62 (5.42) |
| $\alpha_2$ | −56.33 (10.10) | −54.45 (10.93) | −53.86 (11.26) |
| $\alpha_3$ | 26.53 (4.89) | 25.61 (5.28) | 25.31 (5.43) |
| $\sigma^2$ | $1.40 \times 10^{-2}$ ($5.00 \times 10^{-3}$) | $1.80 \times 10^{-2}$ ($7.00 \times 10^{-3}$) | $1.90 \times 10^{-2}$ ($7.00 \times 10^{-3}$) |
| | Student $t$ distribution | | |
| $\alpha_0$ | 30.66 (4.64) | 29.94 (5.05) | 29.85 (5.20) |
| $\alpha_1$ | 8.48 (4.00) | 8.24 (4.42) | 8.46 (4.57) |
| $\alpha_2$ | −58.20 (8.94) | −56.79 (9.71) | −56.67 (10.00) |
| $\alpha_3$ | 27.27 (4.30) | 26.58 (4.66) | 26.55 (4.80) |
| $\sigma^2$ | $7.30 \times 10^{-3}$ ($3.60 \times 10^{-3}$) | $9.20 \times 10^{-3}$ ($4.60 \times 10^{-3}$) | $9.80 \times 10^{-3}$ ($4.90 \times 10^{-3}$) |

where the response variable is the observed lifetime/advertised lifetime ($Y$), the covariates correspond to a measure of gas discharge ($x_1$) and the observed voltage/advertised voltage (measure of performance of lamp and ballast – $x_2$) and the errors are assumed to follow a normal distribution. Here we also assume a Student $t$ distribution with $\nu = 4$ degrees of freedom for the errors.

In Table 8, we present the maximum likelihood estimates, the bias-corrected estimates, the bias-reduced estimates, and the corresponding estimated standard errors. As in the previous application, the estimates present smaller standard errors under the Student $t$ model than under the normal model.

The original MLEs for $\alpha_0$ and $\alpha_3$ are bigger than the corresponding corrected and reduced versions by approximately one unit (normal and Student $t$ models). The largest differences are among the estimates of $\alpha_2$; for example, for the normal model we have −56.33 (MLE), −54.45 (bias-corrected MLE) and −53.86 (bias-reduced MLE).

We now use the Akaike Information Criterion (AIC, Akaike, 1974), the Schwarz Bayesian criterion (BIC, Schwarz, 1978) and the finite sample AIC (AIC$_C$, Hurvich and Tsai, 1989) to evaluate the quality of the normal and Student $t$ fits. For the normal model we have AIC $= -9.98$, BIC $= -6.79$ and AIC$_C = -2.48$. For the $t$ model we have AIC $= -11.24$, BIC $= -8.04$ and AIC$_C = -3.74$. Therefore, the $t$ model presents the best fit for this dataset, since the values of the AIC, BIC and AIC$_C$ are smaller.

Let

$$\widehat{D} = \sum_{j=1}^{14} (\widehat{Y} - \widehat{Y}_{(j)})^\top (\widehat{Y} - \widehat{Y}_{(j)}),$$

where $\widehat{Y}$ and $\widehat{Y}_{(j)}$ are the vectors of predicted values computed from the model fit for the whole sample and the sample without the $j$th observation, respectively. The quantity $\widehat{D}$ measures the total effect of deleting one observation in the predicted values. For a fixed sample size, it tends to be high if a single observation can highly influence the prediction of new observations. We have $\widehat{D} = 0.119, 0.120,$ and $0.123$ (normal model) and $\widehat{D} = 0.101, 0.100,$ and $0.095$ (Student $t$ model) when using the MLE, the bias-corrected estimate, and the bias-reduced estimate, respectively. Notice that $\widehat{D}$ is smaller for the Student $t$ model regardless of the estimate used. This is evidence that the Student $t$ model is more suitable than the normal model for predicting lifetime of fluorescent lamps in this study.

### 6.3 WHO MONICA data

We now turn to a dataset from the WHO MONICA Project that was considered in Kulathinal, Kuulasmaa and Gasbarra (2002). This dataset was first analyzed under normal distributions for the marginals of the random errors (Kulathinal, Kuulasmaa and Gasbarra, 2002; Patriota, Bolfarine and de Castro, 2009). Thereafter, it was studied under a scale mixture of normal distributions for the marginals of the random errors (Cao, Lin and Zhu, 2012). The approach used in the present paper is different from the others because here we consider a joint elliptical distribution for the vector of random errors. The other authors assumed that the distributions of the errors were independent, while we assume that they are uncorrelated but not independent. For our proposal, the errors will only be independent under normality.

The dataset considered here corresponds to the data collected for men ($n = 38$). As describe in Kulathinal, Kuulasmaa and Gasbarra (2002), the data are trends of the annual change in the event rate ($y$) and trends of the risk scores ($x$). The risk score is defined as a linear combination of smoking status, systolic blood pressure, body mass index, and total cholesterol level. A follow-up study using proportional hazards models was employed to derive its coefficients, and provides the observed risk score and its estimated variance. Therefore, the observed response variable, $X_1$, is the average annual change in event rate (%) and the observed covariate, $X_2$, is the observed risk score (%). We use the heteroscedastic model (4.2) with $v = m = 1$ and zero covariance between the errors $\delta_{x_{1i}}$ and $\delta_{x_{2i}}$.

Table 9 gives the MLE and the bias-corrected/reduced estimates (standard errors are given in parentheses). We considered the full sample ($n = 38$) and randomly chosen sub-samples of $n = 10, 20$ and $30$ observations.

The original MLEs for $\beta_0$, $\beta_1$ and $\mu_{x_2}$ are practically the same as their bias-corrected and bias-reduced versions for all sample sizes. The largest differences are among the estimates of $\Sigma_q$; for example, for $n = 10$ we have 6.17 (MLE),

**Table 9** *Estimates and standard errors* (*given in parentheses*); *WHO MONICA data*

| $n$ | $\theta$ | MLE | Bias-corrected MLE | Bias-reduced MLE |
|----|----------|-----|--------------------|------------------|
| 10 | $\beta_0$ | −2.58 (1.34) | −2.58 (1.44) | −2.45 (1.47) |
|    | $\beta_1$ | 0.05 (0.60) | 0.05 (0.63) | 0.07 (0.64) |
|    | $\mu_{x_2}$ | −1.54 (0.58) | −1.54 (0.61) | −1.53 (0.62) |
|    | $\Sigma_{x_2}$ | 2.89 (1.50) | 3.22 (1.65) | 3.29 (1.69) |
|    | $\Sigma_q$ | 6.17 (3.99) | 8.14 (4.93) | 8.81 (5.25) |
| 20 | $\beta_0$ | −2.68 (0.65) | −2.69 (0.68) | −2.69 (0.69) |
|    | $\beta_1$ | 0.48 (0.30) | 0.47 (0.31) | 0.43 (0.31) |
|    | $\mu_{x_2}$ | −1.29 (0.44) | −1.29 (0.46) | −1.29 (0.46) |
|    | $\Sigma_{x_2}$ | 3.53 (1.25) | 3.73 (1.31) | 3.76 (1.32) |
|    | $\Sigma_q$ | 3.00 (1.66) | 3.59 (1.87) | 3.73 (1.92) |
| 30 | $\beta_0$ | −2.22 (0.54) | −2.22 (0.55) | −2.20 (0.55) |
|    | $\beta_1$ | 0.43 (0.24) | 0.43 (0.25) | 0.42 (0.25) |
|    | $\mu_{x_2}$ | −0.77 (0.42) | −0.77 (0.42) | −0.77 (0.42) |
|    | $\Sigma_{x_2}$ | 4.71 (1.34) | 4.88 (1.39) | 4.89 (1.39) |
|    | $\Sigma_q$ | 4.36 (1.86) | 4.89 (2.01) | 4.88 (2.01) |
| 38 | $\beta_0$ | −2.08 (0.53) | −2.08 (0.54) | −2.08 (0.54) |
|    | $\beta_1$ | 0.47 (0.23) | 0.47 (0.24) | 0.46 (0.24) |
|    | $\mu_{x_2}$ | −1.09 (0.36) | −1.09 (0.36) | −1.09 (0.36) |
|    | $\Sigma_{x_2}$ | 4.32 (1.10) | 4.44 (1.13) | 4.45 (1.13) |
|    | $\Sigma_q$ | 4.89 (1.78) | 5.34 (1.89) | 5.30 (1.88) |

8.14 (bias-corrected MLE) and 8.81 (bias-reduced MLE). In general, as expected, larger sample sizes correspond to smaller standard errors.

## 7  Concluding remarks

We studied bias correction and bias reduction for a multivariate elliptical model with a general parameterization that unifies several important models (e.g., linear and nonlinear regressions models, linear and nonlinear mixed models, errors-in-variables models, among many others). We extend the work of Patriota and Lemonte (2009) to the elliptical class of distributions defined in Lemonte and Patriota (2011). We express the second order bias vector of the maximum likelihood estimates as an weighted least-squares regression.

As can be seen in our simulation results, corrected-bias estimators and reduced-bias estimators form a basis of asymptotic inferential procedures that have better performance than the corresponding procedures based on the original estimator. We further note that, in general, the bias-reduced estimates are less biased than the bias-corrected estimates. Computer packages that perform simple operations on matrices and vectors can be used to compute bias-corrected and bias-reduced estimates.

# Appendix

**Lemma A.1.** *Let $z_i \sim \mathrm{El}_{q_i}(0, \Sigma_i, g)$, and $c_i$ and $\psi_{i(2,1)}$ as previously defined. Then,*

$$E(v_i z_i) = 0,$$

$$E(v_i^2 z_i z_i^\top) = \frac{4\psi_{i(2,1)}}{q_i} \Sigma_i,$$

$$E(v_i^2 \mathrm{vec}(z_i z_i^\top) z_i^\top) = 0,$$

$$E(v_i^2 \mathrm{vec}(z_i z_i^\top) \mathrm{vec}(z_i z_i^\top)^\top) = c_i(\mathrm{vec}(\Sigma_i) \mathrm{vec}(\Sigma_i)^\top + 2\Sigma_i \otimes \Sigma_i),$$

$$E(v_i^3 \mathrm{vec}(z_i z_i^\top) \mathrm{vec}(z_i z_i^\top)^\top) = -c_i^*(\mathrm{vec}(\Sigma_i) \mathrm{vec}(\Sigma_i)^\top + 2\Sigma_i \otimes \Sigma_i),$$

$$\begin{aligned}
E(v_i^3 z_i^\top A_{i(t)} z_i z_i^\top A_{i(s)} z_i z_i^\top A_{i(r)} z_i) = -8\widetilde{\omega}_i \big( &\mathrm{tr}\{A_{i(t)}\Sigma_i\} \mathrm{tr}\{A_{i(s)}\Sigma_i\} \mathrm{tr}\{A_{i(r)}\Sigma_i\} \\
&+ 2\mathrm{tr}\{A_{i(t)}\Sigma_i\} \mathrm{tr}\{A_{i(s)}\Sigma_i A_{i(r)}\Sigma_i\} \\
&+ 2\mathrm{tr}\{A_{i(s)}\Sigma_i\} \mathrm{tr}\{A_{i(t)}\Sigma_i A_{i(r)}\Sigma_i\} \\
&+ 2\mathrm{tr}\{A_{i(r)}\Sigma_i\} \mathrm{tr}\{A_{i(t)}\Sigma_i A_{i(s)}\Sigma_i\} \\
&+ 8\mathrm{tr}\{A_{i(t)}\Sigma_i A_{i(s)}\Sigma_i A_{i(r)}\Sigma_i\} \big),
\end{aligned}$$

*where $c_i^* = 8\psi_{i(3,2)}/\{q_i(q_i + 2)\}$, $\psi_{i(3,2)} = E(W_g^3(r_i)r_i^2)$, $\widetilde{\omega}_i = \psi_{i(3,3)}/\{q_i(q_i + 2)(q_i + 4)\}$ and $\psi_{i(3,3)} = E(W_g^3(r_i)r_i^3)$.*

**Proof.** The proof can be obtained by adapting the results of Mitchell (1989) for a matrix version. □

From Lemma A.1, we can find the cumulants of the log-likelihood derivatives required to compute the second-order biases.

**Proof of Theorem 3.1.** Following Cordeiro and Klein (1994), we write (2.1) in matrix notation to obtain the second-order bias vector of $\widehat{\theta}$ in the form

$$B_{\widehat{\theta}}(\theta) = K(\theta)^{-1} W \mathrm{vec}(K(\theta)^{-1}), \tag{A.1}$$

where $W = (W^{(1)}, \ldots, W^{(p)})$ is a $p \times p^2$ partitioned matrix, each $W^{(r)}$, referring to the $r$th component of $\theta$, being a $p \times p$ matrix with typical $(t, s)$th element given by

$$w_{ts}^{(r)} = \frac{1}{2}\kappa_{tsr} + \kappa_{ts,r} = \kappa_{ts}^{(r)} - \frac{1}{2}\kappa_{tsr} = \frac{3}{4}\kappa_{ts}^{(r)} - \frac{1}{4}(\kappa_{t,s,r} + \kappa_{sr}^{(t)} + \kappa_{rt}^{(s)}).$$

Because $K(\theta)$ is symmetric and the $t$th element of $W \mathrm{vec}(K(\theta)^{-1})$ is $w_{t1}^{(1)}\kappa^{1,1} + (w_{t2}^{(1)} + w_{t1}^{(2)})\kappa^{1,2} + \cdots + (w_{tr}^{(s)} + w_{ts}^{(r)})\kappa^{s,r} + \cdots + (w_{tp}^{(p-1)} + w_{t(p-1)}^{(p)})\kappa^{p-1,p} +$

$w_{tp}^{(p)} \kappa^{p,p}$, we may write

$$
w_{ts}^{(r)} = \frac{1}{2}(w_{tr}^{(s)} + w_{ts}^{(r)}) = \frac{1}{4}(\kappa_{ts}^{(r)} + \kappa_{tr}^{(s)} - \kappa_{sr}^{(t)} - \kappa_{t,s,r}). \tag{A.2}
$$

Comparing (A.1) and (3.7), we note that for the proof of this theorem it suffices to show that $F^\top \tilde{H} \xi = W \operatorname{vec}((F^\top \tilde{H} F)^{-1})$, that is,

$$
W = F^\top H M H (\Phi_1, \ldots, \Phi_p).
$$

Notice that

$$
\kappa_{sr} = \sum_{i=1}^{n} \left\{ \frac{c_i}{2} \operatorname{tr}\{A_{i(r)} C_{i(s)}\} - \frac{4\psi_{i(2,1)}}{q_i} a_{i(s)}^\top \Sigma_i^{-1} a_{i(r)} \right.
$$
$$
\left. - \frac{(c_i - 1)}{4} \operatorname{tr}\{A_{i(s)} \Sigma_i\} \operatorname{tr}\{A_{i(r)} \Sigma_i\} \right\}. \tag{A.3}
$$

The quantities $\psi_{i(2,1)}$ and $\psi_{i(2,2)}$ do not depend on $\theta$ and hence, the derivative of (A.3) with respect to $\theta_t$ is

$$
\kappa_{sr}^{(t)} = \sum_{i=1}^{n} \left\{ \frac{c_i}{2} \operatorname{tr}\{A_{i(t)} \Sigma_i A_{i(s)} C_{i(r)} + A_{i(s)} \Sigma_i A_{i(t)} C_{i(r)} + C_{i(ts)} A_{i(r)} \right.
$$
$$
+ C_{i(tr)} A_{i(s)}\} \bigg\} - \sum_{i=1}^{n} \left\{ \frac{4\psi_{i(2,1)}}{q_i} (a_{i(ts)}^\top \Sigma_i^{-1} a_{i(r)} + a_{i(s)}^\top A_{i(t)} a_{i(r)} \right.
$$
$$
\left. + a_{i(s)}^\top \Sigma_i^{-1} a_{i(tr)}) \right\} + \sum_{i=1}^{n} \left\{ \frac{(c_i - 1)}{4} \operatorname{tr}\{A_{i(t)} C_{i(s)} + \Sigma_i^{-1} C_{i(ts)}\} \operatorname{tr}\{A_{i(r)} \Sigma_i\} \right\}
$$
$$
+ \sum_{i=1}^{n} \left\{ \frac{(c_i - 1)}{4} \operatorname{tr}\{A_{i(t)} C_{i(r)} + \Sigma_i^{-1} C_{i(tr)}\} \operatorname{tr}\{A_{i(s)} \Sigma_i\} \right\}.
$$

Therefore,

$$
\kappa_{st}^{(r)} + \kappa_{tr}^{(s)} - \kappa_{sr}^{(t)} = \sum_{i=1}^{n} \left\{ \frac{c_i}{2} \operatorname{tr}\{A_{i(r)} \Sigma_i A_{i(s)} C_{i(t)} + A_{i(s)} \Sigma_i A_{i(r)} C_{i(t)} \right.
$$
$$
\left. + 2 C_{i(rs)} A_{i(t)}\} \right\} - \sum_{i=1}^{n} \left\{ \frac{4\psi_{i(2,1)}}{q_i} (2 a_{i(t)}^\top \Sigma_i^{-1} a_{i(sr)} \right.
$$
$$
\left. + a_{i(t)}^\top A_{i(s)} a_{i(r)} + a_{i(s)}^\top A_{i(r)} a_{i(t)} - a_{i(s)}^\top A_{i(t)} a_{i(r)}) \right\}
$$
$$
+ \sum_{i=1}^{n} \left\{ \frac{(c_i - 1)}{2} \operatorname{tr}\{A_{i(r)} C_{i(s)} + \Sigma_i^{-1} C_{i(rs)}\} \operatorname{tr}\{A_{i(t)} \Sigma_i\} \right\}.
$$

Now, the only quantity that remains to obtain is $\kappa_{t,s,r} = E(U_t U_s U_r)$. Noting that $z_i$ is independent of $z_j$ for $i \neq j$, we have

$$
\kappa_{t,s,r} = \frac{1}{8} \sum_{i=1}^{n} E\Big\{ \big[ \mathrm{tr}\{ A_{i(t)}(\Sigma_i - v_i z_i z_i^\top) \} \, \mathrm{tr}\{ A_{i(s)}(\Sigma_i - v_i z_i z_i^\top) \}
$$

$$
\times \mathrm{tr}\{ A_{i(r)}(\Sigma_i - v_i z_i z_i^\top) \} \big] + 4\,\mathrm{tr}\{ A_{i(t)}(\Sigma_i - v_i z_i z_i^\top) \}
$$

$$
\times \big( v_i^2 a_{i(r)}^\top \Sigma_i^{-1} z_i z_i^\top \Sigma^{-1} a_{i(s)} \big) + 4\,\mathrm{tr}\{ A_{i(r)}(\Sigma_i - v_i z_i z_i^\top) \}
$$

$$
\times \big( v_i^2 a_{i(t)}^\top \Sigma_i^{-1} z_i z_i^\top \Sigma^{-1} a_{i(s)} \big) + 4\,\mathrm{tr}\{ A_{i(s)}(\Sigma_i - v_i z_i z_i^\top) \}
$$

$$
\times \big( v_i^2 a_{i(t)}^\top \Sigma_i^{-1} z_i z_i^\top \Sigma^{-1} a_{i(r)} \big) \Big\}.
$$

Then, by using Lemma A.1 and from (A.2), we have, after lengthy algebra, that

$$
W^{(r)} = \sum_{i=1}^{n} F_i^\top H_i M_i H_i \Phi_{i(r)}, \tag{A.4}
$$

where

$$
\Phi_{i(r)} = -\frac{1}{2}\left( H_i^{-1} M_i^{-1} B_{i(r)} H_i F_i + \frac{\partial F_i}{\partial \theta_r} \right),
$$

and

$$
B_{i(r)} = -\frac{1}{2}\begin{pmatrix} \eta_{1i} C_{i(r)} & 2\eta_{1i}\Sigma_i \otimes a_{i(r)}^\top \\ 2\eta_{2i}\Sigma_i \otimes a_{i(r)} & 2(c_i - 1)S_{1i(r)} \end{pmatrix}
$$

$$
-\frac{1}{4}\begin{pmatrix} \eta_{1i}\Sigma_i \,\mathrm{tr}\{ C_{i(r)}\Sigma_i^{-1} \} & 2\eta_{1i}a_{i(r)}\,\mathrm{vec}(\Sigma_i)^\top \\ 2\eta_{1i}\,\mathrm{vec}(\Sigma_i)a_{i(r)}^\top & 2(c_i + 8\widetilde{\omega}_i)S_{2i(r)} \end{pmatrix},
$$

with

$$
\eta_{1i} = c_i^* + 4\psi_{i(2,1)}/q_i, \qquad \eta_{2i} = c_i^* - 4\psi_{i(2,1)}/q_i,
$$

$$
S_{1i(r)} = \mathrm{vec}(\Sigma_i)\,\mathrm{vec}(C_{i(r)})^\top + \frac{1}{2}\mathrm{vec}(\Sigma_i)\,\mathrm{vec}(\Sigma_i)^\top \mathrm{tr}\{ C_{i(r)}\Sigma_i^{-1} \} \quad \text{and}
$$

$$
S_{2i(r)} = \mathrm{vec}(C_{i(r)})\,\mathrm{vec}(\Sigma_i)^\top + \mathrm{vec}(\Sigma_i)\,\mathrm{vec}(C_{i(r)})^\top + 4\Sigma_i \otimes C_{i(r)}
$$

$$
+ \left[ \Sigma_i \otimes \Sigma_i + \frac{1}{2}\mathrm{vec}(\Sigma_i)\,\mathrm{vec}(\Sigma_i)^\top \right] \mathrm{tr}\{ C_{i(r)}\Sigma_i^{-1} \}.
$$

Using (A.4) and (A.1) the theorem is proved. $\qquad\square$

**Proof of Corollary 3.1.** It follows from Theorem 3.1, eq. (3.7), when

$$
F = \mathrm{block\text{-}diag}\{ F_{\theta_1}, F_{\theta_2} \}, \qquad \widetilde{H} = \mathrm{block\text{-}diag}\{ \widetilde{H}_1, \widetilde{H}_2 \} \quad \text{and} \quad \xi = (\xi_1^\top, \xi_2^\top)^\top,
$$

where $F_{\theta_j} = [F_{\theta_j(1)}^\top, \ldots, F_{\theta_j(n)}^\top]^\top$ and $\widetilde{H}_j = \mathrm{block\text{-}diag}\{ \widetilde{H}_{j(1)}, \ldots, \widetilde{H}_{j(n)} \}$ for $j = 1, 2$, with $F_{\theta_1(i)} = \partial \mu_i / \partial \theta_1^\top$, $F_{\theta_2(i)} = \partial [\mathrm{vec}(\Sigma_i)]/\partial \theta_2^\top$, $\widetilde{H}_{1(i)} = \frac{4\psi_{i(2,1)}}{q_i}\Sigma_i^{-1}$

and $\widetilde{H}_{2(i)} = c_i (2\Sigma_i \otimes \Sigma_i)^{-1} + (c_i - 1) \operatorname{vec}(\Sigma_i^{-1}) \operatorname{vec}(\Sigma_i^{-1})^\top$. Furthermore, $\xi_1 = [\xi_{1(1)}^\top, \ldots, \xi_{1(n)}^\top]^\top$ and $\xi_2 = [\xi_{2(1)}^\top, \ldots, \xi_{2(n)}^\top]^\top$ with

$$\xi_{1(i)} = -\frac{1}{2} \dot{F}_{\theta_1(i)} \operatorname{vec}((F_{\theta_1}^\top \widetilde{H}_{(1)} F_{\theta_1})^{-1}),$$

$$\xi_{2(i)} = \frac{1}{4} M_i^* P_i^* \operatorname{vec}((F_{\theta_1}^\top \widetilde{H}_{(1)} F_{\theta_1})^{-1}) + \frac{1}{8}(M_i^* Q_i^* - 4\dot{F}_{\theta_2(i)})$$

$$\times \operatorname{vec}((F_{\theta_2}^\top \widetilde{H}_{(2)} F_{\theta_2})^{-1}).$$

Also, $\dot{F}_{\theta_1(i)} = [F_{\theta_1(i)}^1, \ldots, F_{\theta_1(i)}^{p_1}]$, $\dot{F}_{\theta_2(i)} = [F_{\theta_2(i)}^1, \ldots, F_{\theta_2(i)}^{p_2}]$, $Q_i^* = [Q_{i(1)}^*, \ldots, Q_{i(p_2)}^*]$, $P_i^* = [P_{i(1)}^*, \ldots, P_{i(p_1)}^*]$, $F_{\theta_1(i)}^r = \frac{\partial F_{\theta_1(i)}}{\partial \theta_{1(r)}}$, $F_{\theta_2(i)}^s = \frac{\partial F_{\theta_2(i)}}{\partial \theta_{2(s)}}$, where $\theta_{1(r)}$ and $\theta_{2(s)}$ are the $r$th and $s$th elements of $\theta_1$ and $\theta_2$, respectively, $r = 1, \ldots, p_1$, $s = 1, \ldots, p_2$ and

$$M_i^* = \frac{1}{c_i}\left(I_{q_i^2} - \frac{\operatorname{vec}(\Sigma_i) \operatorname{vec}(\Sigma_i)^\top}{2c_i + \operatorname{vec}(\Sigma_i)^\top \operatorname{vec}(\Sigma_i)}\right),$$

$$Q_{i(s)}^* = \left((c_i - 1)S_{1i(s)} + \frac{1}{2}(c_i + 8\widetilde{\omega}_i)S_{2i(s)}\right)(\Sigma_i^{-1} \otimes \Sigma_i^{-1})F_{\theta_2(i)},$$

$$P_{i(r)}^* = \left(2\eta_{2i}(I_{q_i} \otimes a_{i(r)}) + \eta_{1i} \operatorname{vec}(\Sigma_i)a_{i(r)}^\top \Sigma_i^{-1}\right)F_{\theta_1(i)}. \qquad \square$$

## Acknowledgement

## References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control.* **19**, 716–723. MR0423716

Bull, S. B., Mak, C. and Greenwood, C. M. T. (2002). A modified score function estimator for multinomial logistic regression in small samples. *Computational Statistical Data Analysis* **39**, 57–74. MR1895558

Cao, C.-Z., Lin, J.-G. and Zhu, X.-X. (2012). On estimation of a heteroscedastic measurement error model under heavy-tailed distributions. *Computational Statistics and Data Analysis* **56**, 438–448. MR2842351

Cordeiro, G. M., Ferrari, S. L. P., Uribe-Opazo, M. A. and Vasconcellos, K. L. P. (2000). Corrected maximum-likelihood estimation in a class of symmetric nonlinear regression models. *Statistics and Probability Letters* **46**, 317–328. MR1743990

Cordeiro, G. M. and Klein, R. (1994). Bias correction in ARMA models. *Statistics and Probability Letters* **19**, 169–176. MR1278646

Cordeiro, G. M. and McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society B* **53**, 629–643. MR1125720

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall. MR0370837

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society B* **49**, 1–39. MR0893334

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals (with discussion). *Journal of the Royal Statistical Society B* **30**, 248–275. MR0237052

Cysneiros, F. J. A., Cordeiro, G. M. and Cysneiros, A. H. M. A. (2010). Corrected maximum likelihood estimators in heteroscedastic symmetric nonlinear models. *Journal of Statistical Computation and Simulation* **80**, 451–461. MR2604169

Doornik, J. A. (2013). *Object-Oriented Matrix Programming Using Ox*. London: Timberlake Consultants Press. (ISBN 978-0-9571708-1-0).

Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. London: Chapman and Hall. MR1071174

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38. MR1225212

Gómez, E., Gómez-Villegas, M. A. and Martín, J. M. (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics. Part A: Theory and Methods* **27**, 589–600. MR1619030

Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297–307. MR1016020

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational Graphics and Statistics* **5**, 299–314.

Kosmidis, I. (2014). Improved estimation in cumulative link models. *Journal of the Royal Statistical Society* **76**, 169–196. MR3153938

Kosmidis, I. and Firth, D. (2009). Bias reduction in exponential family nonlinear models. *Biometrika* **96**, 793–804. MR2564491

Kosmidis, I. and Firth, D. (2011). Multinomial logit bias reduction via the Poisson log-linear model. *Biometrika* **98**, 755–759. MR2836421

Kulathinal, S. B., Kuulasmaa, K. and Gasbarra, D. (2002). Estimation of an errors-in-variables regression model when the variances of the measurement error vary between the observations. *Statistics in Medicine* **21**, 1089–1101.

Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1989). Robust statistical modeling using the *t* distribution. *Journal of the American Statistical Association* **84**, 881–896. MR1134486

Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, 295–303. MR0082237

Lemonte, A. J. (2011). Improved maximum-likelihood estimation in a regression model with general parametrization. *Journal of Statistical Computation and Simulation* **81**, 1027–1037. MR2820064

Lemonte, A. J. and Patriota, A. G. (2011). Multivariate elliptical models with general parameterization. *Statistical Methodology* **8**, 389–400. MR2800359

Lucas, A. (1997). Robustness of the student *t* based M-estimator. *Communications in Statistics, Theory and Methods* **26**, 1165–1182. MR1450228

Magnus, J. R. and Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd ed. Chichester: Wiley. MR0940471

Mehrabi, Y. and Matthews, J. N. S. (1995). Likelihood-based methods for bias reduction in limiting dilution assays. *Biometrics* **51**, 1543–1549.

Mitchell, A. F. S. (1989). The information matrix, skewness tensor and $\alpha$-connections for the general multivariate elliptical distribution. *Annals of the Institute of Statistical Mathematics* **41**, 289–304. MR1006491

Patriota, A. G. (2011). A note on influence in nonlinear mixed-effects elliptical models. *Computational Statistics and Data Analysis* **55**, 218–225. MR2736549

Patriota, A. G., Bolfarine, B. and de Castro, M. (2009). A heteroscedastic structural errors-in-variables model with equation error. *Statistical Methodology* **6**, 408–423.

Patriota, A. G. and Lemonte, A. J. (2009). Bias correction in a multivariate normal regression model with general parameterization. *Statistics & Probability Letters* **79**, 1655–1662. MR2547936

Pettitt, A. N., Kelly, J. M. and Gao, J. T. (1998). Bias correction for censored data with exponential lifetimes. *Statistica Sinica* **8**, 941–963. MR1651517

Rosillo, F. G. and Chivelet, N. M. (2009). Lifetime prediction of fluorescent lamps used in photovoltaic systems. *Lighting Research and Technology* **41**, 183–197.

Schwarz, G. (1978). Estimating the dimensional of a model. *Annals of Statistics* **6**, 461–464. MR0468014

Tiede, J. J. and Pagano, M. (1979). Application of robust calibration to radioimmunoassay. *Biometrics* **35**, 567–574.

Vanegas, L. H. and Paula, G. A. (2015). A semiparametric approach for joint modeling of median and skewness. *Test* **24**, 110–135. MR3314576

Vanegas, L. H. and Paula, G. A. (2016). Log–symmetric distributions: Statistical properties and parameter estimation. *Brazilian Journal of Probability and Statistics* **30**, 196–220. MR3481101

Vasconcellos, K. L. P. and Cordeiro, G. M. (2000). Bias corrected estimates in multivariate student *t* regression models. *Communications in Statistics, Theory and Methods* **29**, 797–822.

T. F. N. Melo
Federal University of Goiás
Goiás
Brazil
E-mail: tmelo@ufg.br

S. L. P. Ferrari
A. G. Patriota
University of São Paulo
São Paulo
Brazil
E-mail: silviaferrari@usp.br
        patriota@ime.usp.br