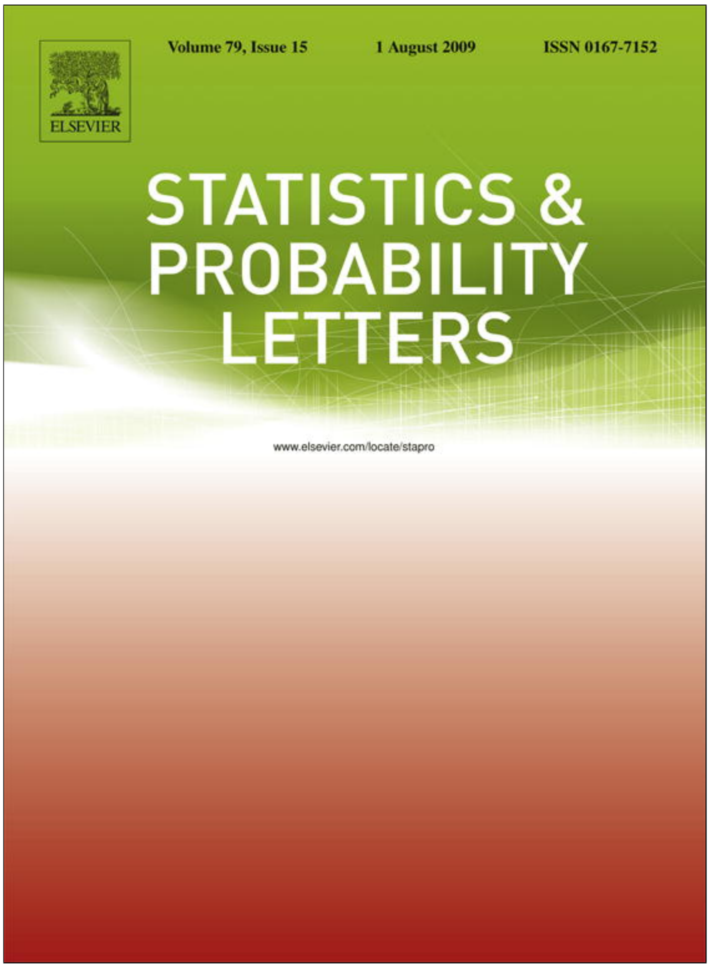
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Statistics and Probability Letters 79 (2009) 1655–1662

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Bias correction in a multivariate normal regression model with general parameterization

Alexandre G. Patriota *, Artur J. Lemonte

Departamento de Estatística, Universidade de São Paulo, Rua do Matão, 1010, São Paulo/SP, 05508-090, Brazil

ARTICLE INFO

Article history:
Received 16 January 2009
Received in revised form 7 April 2009
Accepted 20 April 2009
Available online 3 May 2009

ABSTRACT

This paper derives the second-order biases of maximum likelihood estimates from a multivariate normal model where the mean vector and the covariance matrix have parameters in common. We show that the second order bias can always be obtained by means of ordinary weighted least-squares regressions. We conduct simulation studies which indicate that the bias correction scheme yields nearly unbiased estimators.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Applications of multivariate normal models are commonly found in the literature. In the majority of problems, the estimation procedure in such multivariate normal models makes use of asymptotic theory (e.g., nonlinear multivariate regressions and errors-in-variables models, among others). Then, in the practical applications, the asymptotic distribution of the MLE is often used as an approximation to its exact distribution, since it considerably simplifies the inferential process. In general, under some regularity conditions, the MLEs are consistent and efficient, i.e., asymptotically, their biases converge to zero and their variance–covariance matrices approach the inverse of the Fisher information. Moreover, under such regularity conditions, the MLEs are asymptotically normally distributed. Although the MLEs have these important features, they may be strongly biased for small or even moderate sample sizes when more complex models are considered, since the bias of a MLE is typically of order $\mathcal{O}(n^{-1})$, whereas the asymptotic standard errors are of order $\mathcal{O}(n^{-1/2})$. Thus, a bias correction can play an important role to improve the estimation of the model parameters.

An important area of research in statistics is the study of the finite-sample behavior of MLEs. Bias adjustment has been extensively studied in the statistical literature. For example, Cook et al. (1986) relate bias to the position of the explanatory variables in the sample space; Cordeiro and McCullagh (1991) give general matrix formulae for bias correction in generalized linear models; Cordeiro and Vasconcellos (1997) obtained general matrix formulae for bias correction in multivariate nonlinear regression models with normal errors, while Vasconcellos and Cordeiro (1997) obtained general formulae for bias in multivariate nonlinear heteroscedastic regression. Also, Cordeiro and Vasconcellos (1999) obtained second order biases of the maximum likelihood estimators in von Mises regression models, while Cordeiro et al. (2000) obtained bias correction for symmetric nonlinear regression models. Vasconcellos and Cordeiro (2000) obtained bias correction for multivariate nonlinear Student t regression models, while Cordeiro and Botter (2001) derive general formulae for the second-order biases in overdispersed generalized linear models. More recently, Cordeiro and Toyama (2008) derive general formulae for the second-order biases of maximum likelihood estimates of the parameters in generalized nonlinear models with dispersion covariates.

In this paper we study a multivariate normal model with general parameterization and derive the second-order biases of the maximum likelihood estimates. Here, the general parameterization means a sort of unification of several important models which can be constructed using the multivariate normal model. For instance, the multivariate nonlinear regressions

* Corresponding author.
E-mail addresses: patriota.alexandre@gmail.com (A.G. Patriota), arturlemonte@gmail.com (A.J. Lemonte).

0167-7152/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2009.04.018

Author's personal copy

studied by Cordeiro and Vasconcellos (1997) and their heteroscedastic version (Vasconcellos and Cordeiro, 1997) are just particular cases of our proposal. In this paper we propose a model in which the mean $\boldsymbol{\mu}$ and the variance $\boldsymbol{\Sigma}$ of the observed variables are indexed by the same vector of parameters, say $\boldsymbol{\theta}$. The existing works on bias correction assume that the mean and variance do not share any parameters. However, in errors-in-variables models, for example, this assumption is not realistic. Indeed, that assumption makes the computation of the bias formulae less complicated, but it restricts the applicability of the approach to a special class of models. In view of that, the main goal of this article is to extend the bias correction to a wide class of multivariate models which has not yet been considered in the statistical literature.

The outline of the paper is as follows. Section 2 presents the main model and computes the score function and Fisher's information matrix. In Section 3, we present matrix formulae for the second-order biases of the MLEs for the general model. In Section 4, we present some useful examples of the proposed formulation. Monte Carlo simulation results are presented and discussed in Section 5. The numerical results show that the bias correction we derive is effective in small samples; it delivers estimators that are nearly unbiased and display superior finite-sample behavior. Finally, Section 6 concludes the paper.

2. Model specification

We consider the situation in which n independent multivariate random variables $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are observed and the number of responses measured in each observation is q. We also admit that auxiliary covariates can be observed, say $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. The multivariate regression model can then be represented as

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta}) + \boldsymbol{u}_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where $\boldsymbol{Y}_i$ is a $q \times 1$ vector of dependent variables, $\boldsymbol{\mu}_i(\boldsymbol{\theta}) \equiv \boldsymbol{\mu}_i(\boldsymbol{\theta}, \boldsymbol{X}_i)$ is a mean function (the shape is assumed known) which is three times continuously differentiable with respect to each element of $\boldsymbol{\theta}$ and $\boldsymbol{X}_i$ is an $m \times 1$ vector of known explanatory variables associated with the ith observed response $\boldsymbol{Y}_i$. Also, $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters of interest (where $p < n$ and it is fixed). Additionally, as the foundation for estimation by maximum likelihood and hypothesis testing, we assume that the independent random errors $\boldsymbol{u}_i$'s follow a multivariate normal $\mathcal{N}_q(\boldsymbol{0}, \boldsymbol{\Sigma}_i(\boldsymbol{\theta}))$ distribution, where $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ is a $q \times q$ positive definite covariance matrix and the entries of $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ are assumed three times continuously differentiable in each element of $\boldsymbol{\theta}$. We are assuming, in addition, that the functions $\boldsymbol{\mu}_i(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ are defined in a convenient way since $\boldsymbol{\theta}$ should be identifiable in model (1).

The class of models presented above includes many important regression models. For example, in an errors-in-variables model, we observe two variables, namely Y_i and X_i whose relationship is given by

$$Y_i = \alpha + \beta x_i + e_i \quad \text{and} \quad X_i = x_i + u_i, \tag{2}$$

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 and, additionally, x_i, e_i and u_i are mutually uncorrelated, with $i = 1, 2, \ldots, n$. Then, denoting $\boldsymbol{Y}_i = (Y_i, X_i)^\top$ and $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_x^2, \sigma^2)^\top$ we have that $\boldsymbol{Y}_i \sim \mathcal{N}_2(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, where

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

This is a simple linear regression in which the covariate is subject to measurement errors. This is a good example where the usual approach (assuming that $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ do not share any parameter) is not applicable. Measurement error models have been largely used in epidemiology (Kulathinal et al., 2002; de Castro et al., 2008; Patriota et al., 2009), astrophysics (Akritas and Bershady, 1996; Kelly, 2007; Kelly et al., 2008; Patriota et al., 2009) and analytical chemistry (Cheng and Riu, 2006) to avoid inconsistent estimators (see Fuller, 1987, for further details). Other special cases of model (1) are: multivariate heteroscedastic nonlinear errors-in-variables models, multivariate nonlinear heteroscedastic models, univariate nonlinear models, mixed models, and so on. As can be seen, model (1) can encompass a wide class of models.

To simplify the notation, define $\boldsymbol{Y} = \text{vec}(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n)$, $\boldsymbol{\mu} = \text{vec}(\boldsymbol{\mu}_1(\boldsymbol{\theta}), \boldsymbol{\mu}_2(\boldsymbol{\theta}), \ldots, \boldsymbol{\mu}_n(\boldsymbol{\theta}))$, $\boldsymbol{\Sigma} = \text{diag}\{\boldsymbol{\Sigma}_1(\boldsymbol{\theta}), \boldsymbol{\Sigma}_2(\boldsymbol{\theta}), \ldots, \boldsymbol{\Sigma}_n(\boldsymbol{\theta})\}$ and $\boldsymbol{u} = \boldsymbol{Y} - \boldsymbol{\mu}$, where $\text{vec}(\cdot)$ is the vec operator, which transforms a matrix into a vector by stacking the columns of the matrix one underneath the other. Then, the log-likelihood function associated with (1), apart from an unimportant constant, is

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\text{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\boldsymbol{u}^\top\}. \tag{3}$$

We make some assumptions (Cox and Hinkley, 1974, Ch. 9) on the behavior 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 existence and uniqueness of the MLE of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$. Define the following quantities:

$$\boldsymbol{a}_r = \frac{\partial\boldsymbol{\mu}}{\partial\theta_r}, \qquad \boldsymbol{a}_{sr} = \frac{\partial^2\boldsymbol{\mu}}{\partial\theta_s\partial\theta_r}, \qquad \boldsymbol{C}_r = \frac{\partial\boldsymbol{\Sigma}}{\partial\theta_r}, \qquad \boldsymbol{C}_{sr} = \frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_s\partial\theta_r} \quad \text{and} \quad \boldsymbol{A}_r = \frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\theta_r} = -\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_r\boldsymbol{\Sigma}^{-1},$$

where $r, s = 1, 2, \ldots, p$. To compute the derivatives of $\ell(\boldsymbol{\theta})$ we make use of methods in matrix differential calculus, as described in Magnus and Neudecker (1988). Let

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

Author's personal copy

where $\widetilde{\boldsymbol{D}} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p)$, $\widetilde{\boldsymbol{V}} = (\mathrm{vec}(\boldsymbol{C}_1), \ldots, \mathrm{vec}(\boldsymbol{C}_p))$, $\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}$ and $\otimes$ is the Kronecker product. We assume that $\widetilde{\boldsymbol{F}}$ has rank p. The score function for $\boldsymbol{\theta}$ can be written as $\boldsymbol{U}_{\boldsymbol{\theta}} = \widetilde{\boldsymbol{F}}^\top \widetilde{\boldsymbol{H}} \widetilde{\boldsymbol{u}}$. Also, the expected Fisher information for $\boldsymbol{\theta}$ is

$$\boldsymbol{K}_{\boldsymbol{\theta}} = \widetilde{\boldsymbol{F}}^\top \widetilde{\boldsymbol{H}} \widetilde{\boldsymbol{F}}. \tag{4}$$

The MLE $\widehat{\boldsymbol{\theta}}$ satisfies the equation $\boldsymbol{U}_{\boldsymbol{\theta}} = \boldsymbol{0}$. After some matrix manipulations, the Fisher scoring method can be used to estimate $\boldsymbol{\theta}$ by iteratively solving the equation

$$(\widetilde{\boldsymbol{F}}^{(m)\top} \widetilde{\boldsymbol{H}}^{(m)} \widetilde{\boldsymbol{F}}^{(m)}) \boldsymbol{\theta}^{(m+1)} = \widetilde{\boldsymbol{F}}^{(m)\top} \widetilde{\boldsymbol{H}}^{(m)} \widetilde{\boldsymbol{u}}^{*(m)}, \quad m = 0, 1, 2, \ldots, \tag{5}$$

where $\widetilde{\boldsymbol{u}}^{*(m)} = \widetilde{\boldsymbol{F}}^{(m)} \boldsymbol{\theta}^{(m)} + \widetilde{\boldsymbol{u}}^{(m)}$. Each loop, through the iterative scheme (5), consists of an iterative re-weighted least squares algorithm to optimize the log-likelihood (3). Using Eq. (5) and any software (MAPLE, MATLAB, Ox, R, SAS) with a weighted linear regression routine, one can compute the MLE $\widehat{\boldsymbol{\theta}}$ iteratively. It is also noteworthy that the MLE in even much more complex models, such as multivariate heteroscedastic nonlinear errors-in-variables models, may be attained via iterative formula (5).

3. Biases of estimates of θ

We shall use the following tensor notation for the mixed cumulants of the log-likelihood derivatives in which the indices r, s and t range from 1 to p: $\kappa_{sr} = \mathbb{E}(\partial^2 \ell(\boldsymbol{\theta})/\partial\theta_s \partial\theta_r)$, $\kappa_{s,r} = \mathbb{E}((\partial \ell(\boldsymbol{\theta})/\partial\theta_s)(\partial \ell(\boldsymbol{\theta})/\partial\theta_r))$, $\kappa_{tsr} = \mathbb{E}(\partial^3 \ell(\boldsymbol{\theta})/\partial\theta_t \theta_s \partial\theta_r)$, $\kappa_{sr}^{(t)} = \partial\kappa_{sr}/\partial\theta_t$ and so on. This notation is taken from Lawley (1956). Not all κ's are functionally independent; e.g., $\kappa_{s,r} = -\kappa_{sr}$, which is the typical element of the information matrix $\boldsymbol{K}_{\boldsymbol{\theta}}$, assumed to be nonsingular. All κ's refer to a total over the sample and are, in general, of order n. Finally, let $\kappa^{s,r}$ denote the corresponding element of $\boldsymbol{K}_{\boldsymbol{\theta}}^{-1}$.

After some lengthy algebra, the quantities κ_{sr}, κ_{tsr} and $\kappa_{ts}^{(r)}$ $(r, s, t = 1, 2, \ldots, p)$ are given, respectively, by $\kappa_{sr} = \frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_r \boldsymbol{C}_s\} - \boldsymbol{a}_s^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_r$,

$$\begin{aligned}\kappa_{tsr} = {} & \mathrm{tr}\{(\boldsymbol{A}_r \boldsymbol{\Sigma} \boldsymbol{A}_s + \boldsymbol{A}_s \boldsymbol{\Sigma} \boldsymbol{A}_r)\boldsymbol{C}_t\} + \frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_s \boldsymbol{C}_{tr} + \boldsymbol{A}_r \boldsymbol{C}_{ts} + \boldsymbol{A}_t \boldsymbol{C}_{sr}\} \\ & - (\boldsymbol{a}_t^\top \boldsymbol{A}_s \boldsymbol{a}_r + \boldsymbol{a}_s^\top \boldsymbol{A}_t \boldsymbol{a}_r + \boldsymbol{a}_s^\top \boldsymbol{A}_r \boldsymbol{a}_t + \boldsymbol{a}_t^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{sr} + \boldsymbol{a}_{ts}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_r + \boldsymbol{a}_s^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{tr})\end{aligned} \tag{6}$$

and

$$\kappa_{ts}^{(r)} = \frac{1}{2}\mathrm{tr}\{(\boldsymbol{A}_r \boldsymbol{\Sigma} \boldsymbol{A}_s + \boldsymbol{A}_s \boldsymbol{\Sigma} \boldsymbol{A}_r)\boldsymbol{C}_t + \boldsymbol{A}_t \boldsymbol{C}_{rs} + \boldsymbol{A}_s \boldsymbol{C}_{rt}\} - (\boldsymbol{a}_{rt}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_s + \boldsymbol{a}_t^\top \boldsymbol{A}_r \boldsymbol{a}_s + \boldsymbol{a}_t^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{rs}). \tag{7}$$

Let $B(\widehat{\theta}_a)$ be the n^{-1} bias of $\widehat{\theta}_a$, $a = 1, 2, \ldots, p$. It follows from the general expression for the multiparameter n^{-1} biases of MLEs given by Cox and Snell (1968) that

$$B(\widehat{\theta}_a) = \sum_{t,s,r}{}' \kappa^{a,t} \kappa^{s,r} \left\{ \frac{1}{2}\kappa_{tsr} - \kappa_{ts,r} \right\},$$

where $\sum'$ denotes the summation over all combinations of the parameters $\theta_1, \ldots, \theta_p$. Following Cordeiro and Klein (1994), we write the above equation in matrix notation to obtain n^{-1} bias vector $\boldsymbol{B}(\widehat{\boldsymbol{\theta}})$ of $\widehat{\boldsymbol{\theta}}$ in the form $\boldsymbol{B}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{K}_{\boldsymbol{\theta}}^{-1} \boldsymbol{W} \,\mathrm{vec}(\boldsymbol{K}_{\boldsymbol{\theta}}^{-1})$, where $\boldsymbol{W} = (\boldsymbol{W}^{(1)}, \ldots, \boldsymbol{W}^{(p)})$ is a $p \times p^2$ partitioned matrix, each $\boldsymbol{W}^{(r)}$ referring to the rth component of $\boldsymbol{\theta}$ being a $p \times p$ matrix with typical (t, s)th element given by $w_{ts}^{(r)} = \frac{1}{2}\kappa_{tsr} + \kappa_{ts,r} = \kappa_{ts}^{(r)} - \frac{1}{2}\kappa_{tsr}$. Notice that from (6) and (7) we have that

$$w_{ts}^{(r)} = \frac{1}{4}\mathrm{tr}\{\boldsymbol{A}_t \boldsymbol{C}_{sr} + \boldsymbol{A}_s \boldsymbol{C}_{tr} - \boldsymbol{A}_r \boldsymbol{C}_{ts}\} - \frac{1}{2}(\boldsymbol{a}_t^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{sr} + \boldsymbol{a}_s^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{tr} - \boldsymbol{a}_r^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{ts}) + \frac{1}{2}(\boldsymbol{a}_s^\top \boldsymbol{A}_t \boldsymbol{a}_r + \boldsymbol{a}_t^\top \boldsymbol{A}_s \boldsymbol{a}_r - \boldsymbol{a}_t^\top \boldsymbol{A}_r \boldsymbol{a}_s). \tag{8}$$

Since $\boldsymbol{K}_{\boldsymbol{\theta}}$ is a symmetric matrix and we are interested in the multiplication result of $\boldsymbol{W}\,\mathrm{vec}(\boldsymbol{K}_{\boldsymbol{\theta}}^{-1})$, many terms of (8) cancel. Indeed, note that the tth element of $\boldsymbol{W}\,\mathrm{vec}(\boldsymbol{K}_{\boldsymbol{\theta}}^{-1})$ is given by $w_{t1}^{(1)}\kappa^{1,1} + (w_{t2}^{(1)} + w_{t1}^{(2)})\kappa^{1,2} + \cdots + (w_{tr}^{(s)} + w_{ts}^{(r)})\kappa^{s,r} + \cdots + (w_{tp}^{(p-1)} + w_{t(p-1)}^{(p)})\kappa^{p-1,p} + w_{tp}^{(p)}\kappa^{p,p}$ and $w_{tr}^{(s)} + w_{ts}^{(r)} = \frac{1}{2}\mathrm{tr}(\boldsymbol{A}_t \boldsymbol{C}_{sr}) - \boldsymbol{a}_t^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{sr} + \boldsymbol{a}_s^\top \boldsymbol{A}_t \boldsymbol{a}_r$. Therefore, we can replace the element $w_{ts}^{(r)}$ by $\frac{1}{4}\mathrm{tr}(\boldsymbol{A}_t \boldsymbol{C}_{sr}) - \frac{1}{2}\boldsymbol{a}_t^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_{sr} + \frac{1}{2}\boldsymbol{a}_s^\top \boldsymbol{A}_t \boldsymbol{a}_r$ and $\boldsymbol{W}^{(r)}$ may be written in an equivalent way as $\boldsymbol{W}^{(r)} = \widetilde{\boldsymbol{F}}^\top \widetilde{\boldsymbol{H}} \boldsymbol{\Phi}_r$, $r = 1, \ldots, p$, where $\boldsymbol{\Phi}_r = -\frac{1}{2}(\boldsymbol{G}_r + \boldsymbol{J}_r)$ with

$$\boldsymbol{G}_r = \begin{bmatrix} \boldsymbol{a}_{1r} & \cdots & \boldsymbol{a}_{pr} \\ \mathrm{vec}(\boldsymbol{C}_{1r}) & \cdots & \mathrm{vec}(\boldsymbol{C}_{pr}) \end{bmatrix} \quad \text{and} \quad \boldsymbol{J}_r = \begin{bmatrix} \boldsymbol{0} \\ 2(\boldsymbol{I}_{nq} \otimes \boldsymbol{a}_r)\widetilde{\boldsymbol{D}} \end{bmatrix},$$

where $\boldsymbol{I}_m$ denotes an $m \times m$ identity matrix. That is, the matrix $\boldsymbol{W}$ can be written as $\boldsymbol{W} = \widetilde{\boldsymbol{F}}^\top \widetilde{\boldsymbol{H}}(\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p)$. Then, we arrive at the following theorem.

Theorem 1. *The n^{-1} bias vector $\boldsymbol{B}(\widehat{\boldsymbol{\theta}})$ of $\widehat{\boldsymbol{\theta}}$ is given by*

$$\boldsymbol{B}(\widehat{\boldsymbol{\theta}}) = (\widetilde{\boldsymbol{F}}^\top \widetilde{\boldsymbol{H}} \widetilde{\boldsymbol{F}})^{-1} \widetilde{\boldsymbol{F}}^\top \widetilde{\boldsymbol{H}} \widetilde{\boldsymbol{\xi}}, \tag{9}$$

where $\widetilde{\boldsymbol{\xi}} = (\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p)\mathrm{vec}((\widetilde{\boldsymbol{F}}^\top \widetilde{\boldsymbol{H}} \widetilde{\boldsymbol{F}})^{-1})$.

Author's personal copy

In order to interpret formulae (9) it is helpful to exploit the relationship between the n^{-1} bias of $\widehat{\boldsymbol{\theta}}$ and a linear regression. The bias vector $\boldsymbol{B}(\widehat{\boldsymbol{\theta}})$ is simply the set coefficients from the ordinary weighted least-squares regression of $\widetilde{\boldsymbol{\xi}}$ on the columns of $\widetilde{\boldsymbol{F}}$, using weights in $\widetilde{\boldsymbol{H}}$. As expression (9) makes clear, for any particular model of the class of models presented in Section 2, it is always possible to express the bias of $\widehat{\boldsymbol{\theta}}$ as the solution of an ordinary weighted least-squares regression. Eq. (9) is easily handled algebraically for any type of nonlinear model, since it only involves simple operations on matrices and vectors. This equation, in conjunction with a computer algebra system such as MAPLE (Abell and Braselton, 1994), will yield $\boldsymbol{B}(\widehat{\boldsymbol{\theta}})$ algebraically with minimal effort. Also, we can compute the bias $\boldsymbol{B}(\widehat{\boldsymbol{\theta}})$ numerically via software with numerical linear algebra facilities such as Ox (Doornik, 2006) and R (R Development Core Team, 2008). [Note that we have described a procedure to attain a corrected estimator in a general formulation that covers a wide class of models. In the next section, we shall present some special cases to shed light on the applicability of our general formulation.]

Therefore, we are able to compute the n^{-1} biases of the MLEs for the general model (1) using formula (9). On the right-hand side of expression (9), which is of order n^{-1}, consistent estimates of the parameter $\boldsymbol{\theta}$ can be inserted to define the corrected MLE $\widetilde{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{B}}(\widehat{\boldsymbol{\theta}})$, where $\widehat{\boldsymbol{B}}(\cdot)$ denotes the value of $\boldsymbol{B}(\cdot)$ at $\widehat{\boldsymbol{\theta}}$. The bias-corrected estimate $\widetilde{\boldsymbol{\theta}}$ is expected to have better sampling properties than the uncorrected estimator, $\widehat{\boldsymbol{\theta}}$. For example, in Cordeiro et al. (2008) some simulation studies are presented, showing that the BCEs have a distribution that seems to be closer to the normal distribution than the MLEs. We also present some simulations in Section 5 that indicate that $\widetilde{\boldsymbol{\theta}}$ has smaller bias than its corresponding MLE. On the other hand, we cannot say that the bias corrected estimates always offer some improvement over the MLEs, since in some situations they can have larger mean squared errors.

It is worth emphasizing that there are other methods to bias-correcting MLEs. In regular parametric problems, (Firth, 1993) developed the so-called "preventive" method, which also allows for the removal of the second-order bias. His method consists of modifying the original score function to remove the first-order term from the asymptotic bias of these estimates. In exponential families with canonical parameterization, his correction scheme consists in penalizing the likelihood by the Jeffreys invariant priors. This is a preventive approach to bias adjustment, which has its merits, but the connections between our results and his work are not pursued in this paper since they could be developed in future research. Additionally, we should also stress that it is possible to avoid cumbersome and tedious algebra on cumulant calculations by using Efron's bootstrap (Efron and Tibshirani, 1993). We use the analytical approach here since this leads to a nice formula and we do not need extensive numerical procedures. Moreover, the application of the analytical bias approximation seems to generally be the most feasible procedure to use and it continues to receive attention in the literature.

4. Special models

It is useful to consider some examples to illustrate the applicability of the results in the previous section and clarify the notation used. Other important special models could also be easily handled, since formula (9) only requires simple operations on matrices and vectors.

First, consider a univariate nonlinear model ($q = 1$) in which $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$. Note that this model is a particular case of model (1) with $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$ and $\boldsymbol{\mu} = (\mu_1(\boldsymbol{\beta}), \ldots, \mu_n(\boldsymbol{\beta}))^\top$, where the components of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unrelated and vary independently. Let $p-1$ be the dimension of $\boldsymbol{\beta}$. Here, $\widetilde{\boldsymbol{D}} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{p-1}, \boldsymbol{0})$ and $\widetilde{\boldsymbol{V}} = (\boldsymbol{0}, \mathrm{vec}(\boldsymbol{C}_p))$. Also, $\widetilde{\boldsymbol{F}} = \mathrm{diag}\{\widetilde{\boldsymbol{D}}^{(1)}, \widetilde{\boldsymbol{V}}^{(2)}\}$, where $\widetilde{\boldsymbol{D}}^{(1)} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{p-1})$ and $\widetilde{\boldsymbol{V}}^{(2)} = \mathrm{vec}(\boldsymbol{C}_p)$. Then, from (4), the expected Fisher information for $\boldsymbol{\theta}$ can be written as $\boldsymbol{K}_{\boldsymbol{\theta}} = \widetilde{\boldsymbol{F}}^\top \widetilde{\boldsymbol{H}} \widetilde{\boldsymbol{F}} = \mathrm{diag}\{\boldsymbol{K}_{\boldsymbol{\beta}}, K_{\sigma^2}\}$, where $\boldsymbol{K}_{\boldsymbol{\beta}} = \widetilde{\boldsymbol{D}}^{(1)\top}\widetilde{\boldsymbol{D}}^{(1)}/\sigma^2$ is Fisher's information for $\boldsymbol{\beta}$ and $K_{\sigma^2} = n/2\sigma^4$ is the information relative to σ^2. Since $\boldsymbol{K}_{\boldsymbol{\theta}}$ is block-diagonal, $\boldsymbol{\beta}$ and σ are globally orthogonal (Cox and Reid, 1987). From (9), note that

$$(\widetilde{\boldsymbol{F}}^\top \widetilde{\boldsymbol{H}} \widetilde{\boldsymbol{F}})^{-1} \widetilde{\boldsymbol{F}}^\top \widetilde{\boldsymbol{H}} = \begin{bmatrix} (\widetilde{\boldsymbol{D}}^{(1)\top}\widetilde{\boldsymbol{D}}^{(1)})^{-1}\widetilde{\boldsymbol{D}}^{(1)\top} & \boldsymbol{0} \\ \boldsymbol{0} & \dfrac{1}{n}\widetilde{\boldsymbol{V}}^{(2)\top}(\boldsymbol{I}_n \otimes \boldsymbol{I}_n) \end{bmatrix}.$$

Also,

$$\widetilde{\boldsymbol{\xi}} = \begin{pmatrix} \widetilde{\xi}_1 \\ \widetilde{\xi}_2 \end{pmatrix} = \begin{bmatrix} -\dfrac{\sigma^2}{2}\ddot{\boldsymbol{G}}\, \mathrm{vec}\{(\widetilde{\boldsymbol{D}}^{(1)\top}\widetilde{\boldsymbol{D}}^{(1)})^{-1}\} \\ -\displaystyle\sum_{k=1}^{p-1} (\boldsymbol{I}_n \otimes \boldsymbol{a}_k)\widetilde{\boldsymbol{D}}^{(1)}\boldsymbol{K}_{\beta k}^{-1} \end{bmatrix},$$

where $\ddot{\boldsymbol{G}} = (\boldsymbol{a}_{\beta 1}, \ldots, \boldsymbol{a}_{\beta(p-1)})$ with $\boldsymbol{a}_{\beta k} = (\boldsymbol{a}_{1k}, \ldots, \boldsymbol{a}_{(p-1)k})$ and $\boldsymbol{K}_{\beta k}^{-1}$ is the kth column of $\boldsymbol{K}_{\boldsymbol{\beta}}^{-1}$. Then,

$$\boldsymbol{B}(\widehat{\boldsymbol{\theta}}) = \begin{pmatrix} \boldsymbol{B}(\widehat{\boldsymbol{\beta}}) \\ B(\widehat{\sigma}^2) \end{pmatrix} = \begin{bmatrix} (\widetilde{\boldsymbol{D}}^{(1)\top}\widetilde{\boldsymbol{D}}^{(1)})^{-1}\widetilde{\boldsymbol{D}}^{(1)\top}\widetilde{\xi}_1 \\ \dfrac{1}{n}\widetilde{\boldsymbol{V}}^{(2)\top}(\boldsymbol{I}_n \otimes \boldsymbol{I}_n)\widetilde{\xi}_2 \end{bmatrix}.$$

Note that $\boldsymbol{B}(\widehat{\boldsymbol{\beta}}) = (\widetilde{\boldsymbol{D}}^{(1)\top}\widetilde{\boldsymbol{D}}^{(1)})^{-1}\widetilde{\boldsymbol{D}}^{(1)\top}\widetilde{\boldsymbol{\xi}}_1$ agrees with the result due to Cook et al. (1986, Eq. (3)). Additionally, we obtain the following simple form originally first given by Beale (1960): $B(\widehat{\sigma}^2) = -(p-1)\sigma^2/n$; note that

$$\widetilde{\boldsymbol{V}}^{(2)\top}(\boldsymbol{I}_n \otimes \boldsymbol{I}_n)\sum_{k=1}^{p-1}(\boldsymbol{I}_n \otimes \boldsymbol{a}_k)\widetilde{\boldsymbol{D}}^{(1)}\boldsymbol{K}_{\beta k}^{-1} = \sum_{k=1}^{p-1}\mathrm{vec}(\boldsymbol{C}_p)^\top(\boldsymbol{I}_n \otimes \boldsymbol{a}_k)\widetilde{\boldsymbol{D}}^{(1)}\boldsymbol{K}_{\beta k}^{-1}$$

Author's personal copy

$$= \sum_{k=1}^{p-1} \operatorname{tr}\{\boldsymbol{a}_k \boldsymbol{K}_{\beta k}^{-1} \widetilde{\boldsymbol{D}}^{(1)\top}\} = \operatorname{tr}\{\widetilde{\boldsymbol{D}}^{(1)} \boldsymbol{K}_\beta^{-1} \widetilde{\boldsymbol{D}}^{(1)\top}\} = (p-1)\sigma^2.$$

As a second application, consider the multivariate nonlinear heteroscedastic regression model studied by Vasconcellos and Cordeiro (1997). Note that this model is a particular case of model (1), with $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\sigma}^\top)^\top$, $\boldsymbol{\mu} = \operatorname{vec}(\boldsymbol{\mu}_1(\boldsymbol{\beta}), \ldots, \boldsymbol{\mu}_n(\boldsymbol{\beta}))$ and $\boldsymbol{\Sigma} = \operatorname{diag}\{\boldsymbol{\Sigma}_1(\boldsymbol{\sigma}), \ldots, \boldsymbol{\Sigma}_n(\boldsymbol{\sigma})\}$. Therefore, the components of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unrelated and vary independently. Let p_1 and $p_2 = p - p_1$ be the dimensions of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$, respectively. Here, $\widetilde{\boldsymbol{D}} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{p_1}, \boldsymbol{0})$ and $\widetilde{\boldsymbol{V}} = (\boldsymbol{0}, \operatorname{vec}(\boldsymbol{C}_{p_1+1}), \ldots, \operatorname{vec}(\boldsymbol{C}_p))$. Let $\widetilde{\boldsymbol{D}}^{(1)} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{p_1})$ and $\widetilde{\boldsymbol{V}}^{(2)} = (\operatorname{vec}(\boldsymbol{C}_{p_1+1}), \operatorname{vec}(\boldsymbol{C}_{p_1+2}) \ldots, \operatorname{vec}(\boldsymbol{C}_p))$, then $\widetilde{\boldsymbol{F}} = \operatorname{diag}\{\widetilde{\boldsymbol{D}}^{(1)}, \widetilde{\boldsymbol{V}}^{(2)}\}$. From (4), the expected Fisher information for $\boldsymbol{\theta}$ can be written as $\boldsymbol{K}_\theta = \widetilde{\boldsymbol{F}}^\top \widetilde{\boldsymbol{H}} \widetilde{\boldsymbol{F}} = \operatorname{diag}\{\boldsymbol{K}_\beta, \boldsymbol{K}_\sigma\}$, where $\boldsymbol{K}_\beta = \widetilde{\boldsymbol{D}}^{(1)\top} \boldsymbol{\Sigma}^{-1} \widetilde{\boldsymbol{D}}^{(1)}$ is Fisher's information for $\boldsymbol{\beta}$ and $\boldsymbol{K}_\sigma = \frac{1}{2} \widetilde{\boldsymbol{V}}^{(2)\top} \widetilde{\boldsymbol{\Sigma}}^{-1} \widetilde{\boldsymbol{V}}^{(2)}$ is the information relative to $\boldsymbol{\sigma}$. Since $\boldsymbol{K}_\theta$ is block-diagonal, $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ are globally orthogonal. From (9), it follows that

$$(\widetilde{\boldsymbol{F}}^\top \widetilde{\boldsymbol{H}} \widetilde{\boldsymbol{F}})^{-1} \widetilde{\boldsymbol{F}}^\top \widetilde{\boldsymbol{H}} = \begin{bmatrix} (\widetilde{\boldsymbol{D}}^{(1)\top} \boldsymbol{\Sigma}^{-1} \widetilde{\boldsymbol{D}}^{(1)})^{-1} \widetilde{\boldsymbol{D}}^{(1)\top} \boldsymbol{\Sigma}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & (\widetilde{\boldsymbol{V}}^{(2)\top} \widetilde{\boldsymbol{\Sigma}}^{-1} \widetilde{\boldsymbol{V}}^{(2)})^{-1} \widetilde{\boldsymbol{V}}^{(2)\top} \widetilde{\boldsymbol{\Sigma}}^{-1} \end{bmatrix}.$$

Also,

$$\widetilde{\boldsymbol{\xi}} = \begin{pmatrix} \widetilde{\boldsymbol{\xi}}_1 \\ \widetilde{\boldsymbol{\xi}}_2 \end{pmatrix} = \begin{bmatrix} -\frac{1}{2} \ddot{\boldsymbol{G}} \operatorname{vec}\{(\widetilde{\boldsymbol{D}}^{(1)\top} \boldsymbol{\Sigma}^{-1} \widetilde{\boldsymbol{D}}^{(1)})^{-1}\} \\ -\left(\ddot{\boldsymbol{W}} \operatorname{vec}\{(\widetilde{\boldsymbol{V}}^{(2)\top} \widetilde{\boldsymbol{\Sigma}}^{-1} \widetilde{\boldsymbol{V}}^{(2)})^{-1}\} + \sum_{k=1}^{p_1} (\boldsymbol{I}_{nq} \otimes \boldsymbol{a}_k) \widetilde{\boldsymbol{D}}^{(1)} \boldsymbol{K}_{\beta k}^{-1}\right) \end{bmatrix},$$

where $\ddot{\boldsymbol{G}} = (\boldsymbol{a}_{\beta 1}, \boldsymbol{a}_{\beta 2}, \ldots, \boldsymbol{a}_{\beta p_1})$ with $\boldsymbol{a}_{\beta k} = (\boldsymbol{a}_{1k}, \boldsymbol{a}_{2k}, \ldots, \boldsymbol{a}_{p_1 k})$ and $\ddot{\boldsymbol{W}} = (\boldsymbol{v}_{\sigma(p_1+1)}, \ldots, \boldsymbol{v}_{\sigma p})$ with $\boldsymbol{v}_{\sigma k} = (\operatorname{vec}(\boldsymbol{C}_{(p_1+1)k}), \ldots, \operatorname{vec}(\boldsymbol{C}_{pk}))$ and $\boldsymbol{K}_{\beta k}^{-1}$ is the kth column of $\boldsymbol{K}_\beta^{-1}$. Therefore,

$$\boldsymbol{B}(\widehat{\boldsymbol{\theta}}) = \begin{pmatrix} \boldsymbol{B}(\widehat{\boldsymbol{\beta}}) \\ \boldsymbol{B}(\widehat{\boldsymbol{\sigma}}) \end{pmatrix} = \begin{bmatrix} (\widetilde{\boldsymbol{D}}^{(1)\top} \boldsymbol{\Sigma}^{-1} \widetilde{\boldsymbol{D}}^{(1)})^{-1} \widetilde{\boldsymbol{D}}^{(1)\top} \boldsymbol{\Sigma}^{-1} \widetilde{\boldsymbol{\xi}}_1 \\ (\widetilde{\boldsymbol{V}}^{(2)\top} \widetilde{\boldsymbol{\Sigma}}^{-1} \widetilde{\boldsymbol{V}}^{(2)})^{-1} \widetilde{\boldsymbol{V}}^{(2)\top} \widetilde{\boldsymbol{\Sigma}}^{-1} \widetilde{\boldsymbol{\xi}}_2 \end{bmatrix}.$$

Note that $\boldsymbol{B}(\widehat{\boldsymbol{\beta}}) = (\widetilde{\boldsymbol{D}}^{(1)\top} \boldsymbol{\Sigma}^{-1} \widetilde{\boldsymbol{D}}^{(1)})^{-1} \widetilde{\boldsymbol{D}}^{(1)\top} \boldsymbol{\Sigma}^{-1} \widetilde{\boldsymbol{\xi}}_1$ agrees with the result due to Vasconcellos and Cordeiro (1997, Eq. (3.2)). Additionally, note that $\boldsymbol{B}(\widehat{\boldsymbol{\sigma}}) = (\widetilde{\boldsymbol{V}}^{(2)\top} \widetilde{\boldsymbol{\Sigma}}^{-1} \widetilde{\boldsymbol{V}}^{(2)})^{-1} \widetilde{\boldsymbol{V}}^{(2)\top} \widetilde{\boldsymbol{\Sigma}}^{-1} \widetilde{\boldsymbol{\xi}}_2$ also reduces to Vasconcellos and Cordeiro's 1997 Eq. (3.8), since $\widetilde{\boldsymbol{V}}^{(2)\top} \widetilde{\boldsymbol{\Sigma}}^{-1} \sum_{k=1}^{p_1} (\boldsymbol{I}_{nq} \otimes \boldsymbol{a}_k) \widetilde{\boldsymbol{D}}^{(1)} \boldsymbol{K}_{\beta k}^{-1} = \widetilde{\boldsymbol{V}}^{(2)\top} \widetilde{\boldsymbol{\Sigma}}^{-1} \operatorname{vec}(\boldsymbol{\Delta}_*)$, where $\boldsymbol{\Delta}_*$ is as defined by Vasconcellos and Cordeiro (1997, p. 148).

Next, unlike the two models discussed previously, we consider a model where the elements of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are related and do not vary independently. Consider the nonlinear heteroscedastic errors-in-variables model

$$Y_i = \alpha + \beta x_i + \exp(\gamma z_i) + e_i \quad \text{and} \quad X_i = x_i + u_i,$$

where $x_i \sim \mathcal{N}(\mu_x, \sigma_x^2)$ is the unobservable covariate, $u_i \sim \mathcal{N}(0, \sigma_u^2)$ and $e_i \sim \mathcal{N}(0, \sigma^2 \exp\{\eta z_i\})$ are the measurement errors with σ_u^2 known. The covariate z_i is known. In this example, the vector of parameters is $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \mu_x, \sigma_x^2, \sigma^2, \eta)^\top$ and the mean and variance functions for the ith observation (Y_i, X_i) are given by

$$\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \begin{pmatrix} \alpha + \beta\mu_x + \exp(\gamma z_i) \\ \mu_x \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \begin{pmatrix} \beta^2\sigma_x^2 + \sigma^2 \exp(\eta z_i) & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix}.$$

Then,

$$\boldsymbol{a}_1 = \boldsymbol{1}_n \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \boldsymbol{a}_2 = \boldsymbol{1}_n \otimes \begin{pmatrix} \mu_x \\ 0 \end{pmatrix}, \qquad \boldsymbol{a}_3 = \operatorname{vec}\left\{ \begin{pmatrix} z_1 \exp(\gamma z_1) \\ 0 \end{pmatrix} \cdots \begin{pmatrix} z_n \exp(\gamma z_n) \\ 0 \end{pmatrix} \right\},$$

$$\boldsymbol{a}_4 = \boldsymbol{1}_n \otimes \begin{pmatrix} \beta \\ 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{a}_5 = \boldsymbol{a}_6 = \boldsymbol{a}_7 = \boldsymbol{0},$$

where $\boldsymbol{1}_n$ denotes an $n \times 1$ vector of ones. Also, $\boldsymbol{a}_{rs} = \boldsymbol{0}$ for all r, s except for

$$\boldsymbol{a}_{24} = \boldsymbol{a}_{42} = \boldsymbol{1}_n \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{a}_{33} = \operatorname{vec}\left\{ \begin{pmatrix} z_1^2 \exp(\gamma z_1) \\ 0 \end{pmatrix} \cdots \begin{pmatrix} z_n^2 \exp(\gamma z_n) \\ 0 \end{pmatrix} \right\}.$$

Also, $\boldsymbol{C}_1 = \boldsymbol{C}_3 = \boldsymbol{C}_4 = \boldsymbol{0}$ and

$$\boldsymbol{C}_2 = \boldsymbol{I}_n \otimes \begin{pmatrix} 2\beta\sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & 0 \end{pmatrix}, \qquad \boldsymbol{C}_5 = \boldsymbol{I}_n \otimes \begin{pmatrix} \beta^2 & \beta \\ \beta & 1 \end{pmatrix}, \qquad \boldsymbol{C}_6 = \bigoplus_{i=1}^{n} \begin{pmatrix} \exp(\eta z_i) & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$\boldsymbol{C}_7 = \bigoplus_{i=1}^{n} \begin{pmatrix} z_i\sigma^2 \exp(\eta z_i) & 0 \\ 0 & 0 \end{pmatrix},$$

Author's personal copy

where $\oplus$ is the direct sum of matrices. Additionally, $\boldsymbol{C}_{rs} = \boldsymbol{0}$ for all r, s except for

$$\boldsymbol{C}_{22} = \boldsymbol{I}_n \otimes \begin{pmatrix} 2\sigma_x^2 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \boldsymbol{C}_{25} = \boldsymbol{C}_{52} = \boldsymbol{I}_n \otimes \begin{pmatrix} 2\beta & 1 \\ 1 & 0 \end{pmatrix},$$

$$\boldsymbol{C}_{67} = \boldsymbol{C}_{76} = \bigoplus_{i=1}^{n} \begin{pmatrix} z_i \exp(\eta z_i) & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{C}_{77} = \bigoplus_{i=1}^{n} \begin{pmatrix} z_i^2 \sigma^2 \exp(\eta z_i) & 0 \\ 0 & 0 \end{pmatrix}.$$

Thus, $\widetilde{\boldsymbol{D}} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{a}_4, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0})$, $\widetilde{\boldsymbol{V}} = (\boldsymbol{0}, \mathrm{vec}(\boldsymbol{C}_2), \boldsymbol{0}, \boldsymbol{0}, \mathrm{vec}(\boldsymbol{C}_5), \mathrm{vec}(\boldsymbol{C}_6), \mathrm{vec}(\boldsymbol{C}_7))$ and the matrix formula (9) can be used to compute the second-order bias for this model. Notice that, as $\mathrm{vec}(\boldsymbol{C}_2)$ is not equal to zero, the derivation of algebraic expression using matrix formula (9) becomes very tedious, since the structure of $\boldsymbol{K}_{\boldsymbol{\theta}}$ is not block-diagonal unlike the two previous examples. However, using MAPLE, for example, the derivation can be easily done. Also, the n^{-1} bias vector $\boldsymbol{B}(\widehat{\boldsymbol{\theta}})$ can be obtained numerically via software with numerical linear algebra facilities, with minimal effort, such as R and Ox.

5. Simulation study

We recall that, for large samples, the biases of the MLEs are negligible. However, for small and moderate sample sizes the second-order biases may be large and can be used to improve the estimation. We shall use Monte Carlo simulation to evaluate the finite sample performance of the original MLEs and their corrected versions. All simulations were performed using R (R Development Core Team, 2008). The sample sizes considered were $n = 15, 25, 35, 50$ and 100, and the number of Monte Carlo replications was 5000.

We consider the simple errors-in-variables model as described in (2). Here, $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_x^2, \sigma^2)^\top$ and

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \boldsymbol{1}_n \otimes \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{I}_n \otimes \begin{pmatrix} \beta^2\sigma_x^2 + \sigma^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix}.$$

From the previous expressions, we have immediately that

$$\boldsymbol{a}_1 = \boldsymbol{1}_n \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \boldsymbol{a}_2 = \boldsymbol{1}_n \otimes \begin{pmatrix} \mu_x \\ 0 \end{pmatrix}, \qquad \boldsymbol{a}_3 = \boldsymbol{1}_n \otimes \begin{pmatrix} \beta \\ 1 \end{pmatrix}, \qquad \boldsymbol{a}_4 = \boldsymbol{a}_5 = \boldsymbol{0}$$

and $\boldsymbol{a}_{rs} = \boldsymbol{0}$ for all r, s except for

$$\boldsymbol{a}_{23} = \boldsymbol{a}_{32} = \boldsymbol{1}_n \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Also, $\boldsymbol{C}_1 = \boldsymbol{C}_3 = \boldsymbol{0}$ and

$$\boldsymbol{C}_2 = \boldsymbol{I}_n \otimes \begin{pmatrix} 2\beta\sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & 0 \end{pmatrix}, \qquad \boldsymbol{C}_4 = \boldsymbol{I}_n \otimes \begin{pmatrix} \beta^2 & \beta \\ \beta & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{C}_5 = \boldsymbol{I}_n \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Additionally, $\boldsymbol{C}_{rs} = \boldsymbol{0}$ for all r, s except for

$$\boldsymbol{C}_{22} = \boldsymbol{I}_n \otimes \begin{pmatrix} 2\sigma_x^2 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{C}_{24} = \boldsymbol{C}_{42} = \boldsymbol{I}_n \otimes \begin{pmatrix} 2\beta & 1 \\ 1 & 0 \end{pmatrix}.$$

Thus, $\widetilde{\boldsymbol{D}} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{0}, \boldsymbol{0})$ and $\widetilde{\boldsymbol{V}} = (\boldsymbol{0}, \mathrm{vec}(\boldsymbol{C}_2), \boldsymbol{0}, \mathrm{vec}(\boldsymbol{C}_4), \mathrm{vec}(\boldsymbol{C}_5))$. Therefore, all the quantities necessary to calculate $\boldsymbol{B}(\widehat{\boldsymbol{\theta}})$ using expression (9) are given.

In order to analyze the point estimation results, we computed, for each sample size and for each estimator: the bias (the bias of an estimator $\widehat{\theta}$ is defined as $\mathbb{E}(\widehat{\theta}) - \theta$, its estimate being obtained by estimating $\mathbb{E}(\widehat{\theta})$ through Monte Carlo simulations), the bias relative (defined as $\{\mathbb{E}(\widehat{\theta}) - \theta\}/\theta$), the standard deviation (SD) and the root mean squared error ($\sqrt{\text{MSE}}$). The true values of the regression parameters were set at $\alpha = 67$, $\beta = 0.42$, $\mu_x = 70$, $\sigma_x^2 = 247$ and $\sigma^2 = 43$. The parameter setting were chosen in order to represent the dataset (yields of corn on Marshall soil in Iowa) presented in Fuller (1987, p. 18). The known measurement error variance is $\sigma_u^2 = 57$ (which was attained through a previous experiment).

Table 1 gives the bias, relative bias, SD and $\sqrt{\text{MSE}}$ of both uncorrected and corrected estimates. The figures in this table confirm that the bias-corrected estimates are generally closer to the true parameter values than the unadjusted estimates. We observe that, in absolute value, the estimated relative biases of the bias-corrected estimator were smaller than that of the original MLE for all sample sizes considered, thus showing the effectiveness of the bias correction schemes used in the definition of these estimators.

For instance, when $n = 15$, the estimated relative bias of the estimators of $\alpha, \beta, \mu_x, \sigma_x^2$ and σ^2 average -0.0518 whereas the biases of the five corresponding bias-adjusted estimators average -0.0056; that is, the average bias (in value absolute) of the MLEs is almost ten times larger than that of the bias-corrected estimators.

We can readily see that the MLEs of σ_x^2 and σ^2 are on average far from the true parameter value, thus displaying large bias, for the different sample sizes considered, even when $n = 100$. This stresses the importance of using a bias correction. For instance, when $n = 50$, the relative biases of $\widehat{\sigma}_x^2$ and $\widehat{\sigma}^2$ (MLEs) were -0.0226 and -0.0563, respectively, while the relative biases of $\widetilde{\sigma}_x^2$ and $\widetilde{\sigma}^2$ (BCEs) were 0.0016 (sixteen times lesser) and -0.0011 (fifty times lesser), respectively. Observe that the MLEs are, on average, underestimating the true values of σ_x^2 and σ^2, since their biases are, on average, negatives. Note

Author's personal copy

Table 1
Biases, relative biases, SD and $\sqrt{\text{MSE}}$ of uncorrected and corrected estimates for an errors-in-variables model.

n	$\boldsymbol{\theta}$	MLE				BCE			
		Bias	Rel. bias	SD	$\sqrt{\text{MSE}}$	Bias	Rel. bias	SD	$\sqrt{\text{MSE}}$
15	α	−1.6080	−0.0240	12.36	12.46	1.5544	0.0232	11.19	11.29
	β	0.0230	0.0547	0.17	0.17	−0.0220	−0.0526	0.16	0.16
	μ_x	0.0980	0.0014	4.47	4.48	0.0980	0.0014	4.47	4.48
	σ_x^2	−19.6612	−0.0796	106.70	108.49	−0.7163	−0.0029	113.81	113.81
	σ^2	−7.7701	−0.1807	17.90	19.52	0.1333	0.0031	20.38	20.38
25	α	−1.3266	−0.0198	8.95	9.05	0.0603	0.0009	8.14	8.14
	β	0.0185	0.0440	0.13	0.13	−0.0012	−0.0029	0.11	0.11
	μ_x	0.0280	0.0004	3.43	3.43	0.0280	0.0004	3.43	3.43
	σ_x^2	−13.6591	−0.0553	84.64	85.73	−2.0254	−0.0082	88.02	88.05
	σ^2	−5.1514	−0.1198	14.60	15.48	−0.4472	−0.0104	15.73	15.73
35	α	−0.7839	−0.0117	7.01	7.05	0.0670	0.0010	6.68	6.68
	β	0.0112	0.0267	0.10	0.10	−0.0009	−0.0023	0.09	0.09
	μ_x	−0.0070	−0.0001	2.96	2.96	−0.0070	−0.0001	2.96	2.96
	σ_x^2	−10.4728	−0.0424	70.59	71.36	−2.0748	−0.0084	72.61	72.64
	σ^2	−3.4357	−0.0799	12.36	12.83	−0.0602	−0.0014	13.04	13.04
50	α	−0.5360	−0.0080	5.66	5.69	0.0134	0.0002	5.50	5.50
	β	0.0080	0.0190	0.08	0.08	0.0002	0.0005	0.08	0.08
	μ_x	−0.0490	−0.0007	2.45	2.45	−0.0490	−0.0007	2.45	2.45
	σ_x^2	−5.5822	−0.0226	60.50	60.76	0.3952	0.0016	61.71	61.71
	σ^2	−2.4209	−0.0563	10.48	10.75	−0.0473	−0.0011	10.89	10.89
100	α	−0.1675	−0.0025	3.88	3.83	0.0871	0.0013	3.83	3.78
	β	0.0024	0.0057	0.05	0.05	−0.0012	−0.0029	0.05	0.05
	μ_x	0.0140	0.0002	1.75	1.72	0.0140	0.0002	1.75	1.72
	σ_x^2	−3.2357	−0.0131	42.73	42.24	−0.2223	−0.0009	43.15	42.54
	σ^2	−1.2814	−0.0298	7.40	7.63	−0.0903	−0.0021	7.55	7.67

BCE: bias-corrected estimator.

Table 2
Empirical coverage and average length at 95% of confidence.

n	Coverage		Average length	
	MLE	BCE	MLE	BCE
15	0.940	0.947	0.628	0.611
25	0.944	0.950	0.449	0.446
35	0.943	0.950	0.368	0.368
50	0.944	0.950	0.300	0.300
100	0.947	0.950	0.209	0.209

also that root mean squared errors decrease with n, as expected. Additionally, we note that all estimators have similar root mean squared errors.

Although the biases presented in our simulations are small relative to the root MSEs, it is noteworthy that in some cases the biases are non-negligible when they are compared to the standard errors of MLEs and, in these cases, they can change the inferences; see, for example, Cordeiro (2008). Note that, the model considered by the author is just a particular case of our proposal.

We also conduct a simple Monte Carlo simulation study to verify the empirical coverage and average length of the asymptotic 95% confidence intervals for $\beta = 0.42$. The intervals are defined as $\widehat{\beta} \pm 1.96 \times \text{ep}(\widehat{\beta})$ (for MLE) and $\widetilde{\beta} \pm 1.96 \times \text{ep}(\widetilde{\beta})$ (for BCE), where $\text{ep}(\cdot)$ denotes the estimative of the standard error by using the corresponding estimator. Table 2 shows the results. Overall, note that the BCE is slightly better than the corresponding MLE in both empirical coverage and average length.

6. Conclusions

This paper proposed a bias correction for a multivariate normal model with quite a general parameterization. The main result centers on models where the mean and the variance share the same vector of parameters. Many models are particular cases of the proposed model, such as (non)linear regressions, errors-in-variables models, mixed models and so forth. We have shown that it is always possible to express the second order bias vector of the maximum likelihood estimates as an ordinary weighted least-squares regression. Moreover, we derived a bias-adjustment scheme that nearly eliminates the bias of the maximum likelihood estimator in small and moderate samples. Our simulation results suggest that the bias correction we have derived is very effective, even when the sample size is very small. Indeed, the bias correction mechanism proposed in this paper yields modified maximum likelihood estimators that are nearly unbiased.

Author's personal copy

Acknowledgments

We gratefully acknowledge grants from FAPESP (Brazil). The authors are also grateful to an associate editor and a referee for helpful comments and suggestions. We wish to thank Francisco Cribari–Neto for valuable comments on this manuscript.

References

Abell, M.L., Braselton, J.P., 1994. The Maple V Handbook. AP Professional, New York.
Akritas, M.G., Bershady, M.A., 1996. Linear regression for astronomical data with measurement errors and intrinsic scatter. The Astrophysical Journal 470, 706–714.
Beale, E.M.L., 1960. Confidence regions in non-linear estimation (with discussion). Journal of the Royal Statistical Society B 22, 41–88.
Cheng, C.L., Riu, J., 2006. On estimating linear relationships when both variables are subject to heteroscedastic measurement errors. Technometrics 48, 511–519.
Cook, R.D., Tsai, C., Wei, B., 1986. Bias in nonlinear regression. Biometrika 73, 615–623.
Cordeiro, G.M., 2008. Corrected maximum likelihood estimators in linear heteroscedastic regression models. Brazilian Review of Econometrics 28, 53–67.
Cordeiro, G.M., Botter, D.A., 2001. Second-order biases of maximum likelihood estimates in overdispersed generalized linear models. Statistics and Probability Letters 55, 269–280.
Cordeiro, G.M., Ferrari, S.L.P., Uribe-Opazo, M.A., Vasconcellos, K.L.P., 2000. Corrected maximum-likelihood estimation in a class of symmetric nonlinear regression models. Statistics and Probability Letters 46, 317–328.
Cordeiro, G.M., Klein, R., 1994. Bias correction in ARMA models. Statistics and Probability Letters 19, 169–176.
Cordeiro, G.M., McCullagh, P., 1991. Bias correction in generalized linear models. Journal of the Royal Statistical Society B 53, 629–643.
Cordeiro, GM., Previdelli, I., Samohyl, W., 2008. Bias corrected maximum likelihood estimators in nonlinear overdispersed models. Brazilian Journal Probability and Statistics 22, 105–118.
Cordeiro, G.M., Vasconcellos, K.L.P., 1997. Bias correction for a class of multivariate nonlinear regression models. Statistics and Probability Letters 35, 155–164.
Cordeiro, G.M., Vasconcellos, K.L.P., 1999. Second-order biases of the maximum likelihood estimates in von Mises regression models. Australian and New Zealand Journal of Statistics 41, 901–910.
Cordeiro, G.M., Toyama, M.C., 2008. Bias correction in generalized nonlinear models with dispersion covariates. Communications in Statistics, Theory and Methods 37, 2219–2225.
Cox, D.R., Hinkley, D.V., 1974. Theoretical Statistics. Chapman and Hall, London.
Cox, D.R., Reid, N., 1987. Parameter orthogonality and approximate conditional inference (with discussion). Journal of the Royal Statistical Society B 40, 1–39.
Cox, D.R., Snell, E., 1968. A general definition of residuals (with discussion). Journal of the Royal Statistical Society B 30, 248–275.
de Castro, M., Galea, M., Bolfarine, H., 2008. Hypothesis testing in an errors-in-variables model with heteroscedastic measurement errors. Statistics in Medicine 27, 5217–5234.
Doornik, J.A., 2006. An Object-Oriented Matrix Language – Ox 4, 5th ed. Timberlake Consultants Press, London.
Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman and Hall, New York.
Firth, D., 1993. Bias reduction of maximum likelihood estimates. Biometrika 80, 27–38.
Fuller, W., 1987. Measurement Error Models. Wiley, Chichester.
Kelly, B.C., 2007. Some aspects of measurement error in linear regression of astronomical data. The Astrophysical Journal 665, 1489–1506.
Kelly, B.C., Bechtold, J., Trump, J.R., Vertergaard, M., Siemiginowska, A., 2008. Observational constraints on the dependence of ratio-quiet quasar X-ray emission on black hole mass and accretion rate. Astrophysical Journal Supplement Series 176, 355–373.
Kulathinal, S.B., Kuulasmaa, K., 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.
Lawley, D.N., 1956. A general method for approximating to the distribution of likelihood ratio criteria. Biometrika 43, 295–303.
Magnus, J.R., Neudecker, H., 1988. Matrix Differential Calculus. Wiley, New York.
Patriota, A.G., Bolfarine, H., de Castro, M., 2009. A heteroscedastic errors-in-variables model with equation error. Statistical Methodology doi:10.1016/j.stamet.2009.02.003.
R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.
Vasconcellos, K.L.P., Cordeiro, G.M., 1997. Approximate bias for multivariate nonlinear heteroscedastic regressions. Brazilian Journal of Probability and Statistics 11, 141–159.
Vasconcellos, K.L.P., Cordeiro, G.M., 2000. Bias corrected estimates in multivariate Student *t* regression models. Communications in Statistics, Theory and Methods 29, 797–822.