Improved maximum likelihood estimators in a heteroskedastic errors-in-variables model

Alexandre G. Patriota, Artur J. Lemonte,
Heleno Bolfarine

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

Abstract

This paper develops a bias correction scheme for a multivariate heteroskedastic errors-in-variables model. The applicability of this model is justified in areas such as astrophysics, epidemiology and analytical chemistry, where the variables are subject to measurement errors and the variances vary with the observations. We conduct Monte Carlo simulations to investigate the performance of the corrected estimators. The numerical results show that the bias correction scheme yields nearly unbiased estimates. We also give an application to a real data set.

Key words: Bias correction, errors-in-variables model, maximum likelihood estimation, heteroskedastic model.

1 Introduction

Heteroskedastic errors-in-variables (or measurement error) models have been extensively studied in the statistical literature and widely applied in astrophysics (to explain relationships between black hole masses and some variates of luminosities), epidemiology (to model the cardiovascular event with its risk factors), analytical chemistry (to compare different types of measurement instruments). The applicability of this model abound mainly in the astronomy literature where all quantities are subject to measurement errors (Akritas and Bershady, 1996).

It is well-known that, when the measurement errors are ignored in the estimation process, the maximum-likelihood estimators (MLEs) become inconsistent. More precisely, the estimation of the slope parameters is attenuated (Fuller, 1987). When variables are subject to measurement errors, a special inference treatment must be carried out for the model parameters in order to avoid inconsistent estimators. Usually, a measurement equation is added to the model to capture the measurement error effect and then the MLEs from this approach are consistent, efficient and asymptotically normally distributed. A careful and deep exposition on the inferential process in errors-in-variables models can be seen in Fuller (1987) and the references therein.

Although consistent, asymptotically efficient and asymptotically normally distributed, the MLEs are oftentimes biased and point inference can be damaged. This is not a serious problem for relatively large sample sizes, since bias is typically of order $\mathcal{O}(n^{-1})$, while the asymptotic standard errors are of order $\mathcal{O}(n^{-1/2})$. However, for small or even moderate values of the sample size n, bias can constitute a problem. Bias adjustment has been extensively studied in the statistical literature. For example, Cook *et al.* (1986), Cordeiro (1993), Cordeiro and Vasconcellos (1997), Vasconcellos and Cordeiro (1997) and, more recently, Cordeiro (2008). Additionally, Patriota and Lemonte (2009) obtained general matrix formulae for the second-order biases of the maximum-likelihood estimators in a very general model which includes all previous works aforementioned. The model presented by the authors considers that the mean vector and the variance-covariance matrix of the observed variable have parameters in common. This approach includes the heteroskedastic measurement error model that we are going to study in this paper.

The main goal of this article is define bias-corrected estimators using the general second-order bias expression derived in Patriota and Lemonte (2009) assuming that model defined by (1) and (2) holds. Additionally, we compare the performance of bias-corrected estimators with the MLEs in small samples via Monte Carlo simulations. The numerical results show that the bias correction is effective in small samples and leads to estimates that are nearly unbiased and display superior finite-sample behavior.

The rest of the paper is as follows. Section 2 presents the multivariate heteroskedastic errors-in-variables model. Using general results from Patriota and Lemonte (2009), we derive in Section 3 the second-order biases of the MLEs of the parameters. The result is used to define bias-corrected estimates. In Section 4 the $O(n^{-1})$ biases of the estimates $\widehat{\boldsymbol{\mu}}_i$ and $\widehat{\boldsymbol{\Sigma}}_i$ are given. Monte Carlo simulation results are presented and discussed in Section 5. Section 6 gives an application. Finally, concluding remarks are offered in Section 7.

2 The model

The multivariate model assumed throughout this paper is

$$\boldsymbol{y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{x}_i + \boldsymbol{q}_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{y}_i$ is a $(v \times 1)$ latent response vector, $\boldsymbol{x}_i$ is a $(m \times 1)$ latent vector of covariates, $\boldsymbol{\beta}_0$ is a $(v \times 1)$ vector of intercepts, $\boldsymbol{\beta}_1$ is a $(v \times m)$ matrix which elements are inclinations and $\boldsymbol{q}_i$ is the equation error having multivariate normal distribution with mean zero and covariance-variance matrix $\boldsymbol{\Sigma}_{\boldsymbol{q}}$. The variables $\boldsymbol{y}_i$ and $\boldsymbol{x}_i$ are not directly observed, instead surrogate variables $\boldsymbol{Y}_i$ and $\boldsymbol{X}_i$ are measured with the following additive structure:

$$\boldsymbol{Y}_i = \boldsymbol{y}_i + \boldsymbol{\eta}_{\boldsymbol{y}_i} \quad \text{and} \quad \boldsymbol{X}_i = \boldsymbol{x}_i + \boldsymbol{\eta}_{\boldsymbol{x}_i}. \tag{2}$$

The errors $\boldsymbol{\eta}_{\boldsymbol{y}_i}$ and $\boldsymbol{\eta}_{\boldsymbol{x}_i}$ are assumed to follow a normal distribution given by

$$\begin{pmatrix} \boldsymbol{\eta}_{\boldsymbol{y}_i} \\ \boldsymbol{\eta}_{\boldsymbol{x}_i} \end{pmatrix} \overset{ind}{\sim} \mathcal{N}_{v+m}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\tau}_{\boldsymbol{y}_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\tau}_{\boldsymbol{x}_i} \end{pmatrix}\right],$$

where "$\stackrel{ind}{\sim}$" means "independently distributed as" and the covariance-variance matrices $\boldsymbol{\tau}_{\boldsymbol{y}_i}$ and $\boldsymbol{\tau}_{\boldsymbol{x}_i}$ are assumed to be known for all $i = 1, \ldots, n$. These matrices may be attained, for example, through an analytical treatment of the data collection mechanism, replications, machine precision, etc.

Model (2) has equation errors for all lines, i.e., $\boldsymbol{y}_i$ and $\boldsymbol{x}_i$ are not perfectly related. These equation errors are justified by the influence of other factors than $\boldsymbol{x}_i$ in the variation of $\boldsymbol{y}_i$. It is very reasonable to consider equation errors in (1) to capture extra variability, since the variances $\boldsymbol{\tau}_{\boldsymbol{y}_i}$ are fixed and whether some other factor affects the variation of $\boldsymbol{y}_i$, the estimation of the line parameters will be clearly affected. Supposing that $\boldsymbol{x}_i \stackrel{iid}{\sim} \mathcal{N}_m(\boldsymbol{\mu}_{\boldsymbol{x}}, \boldsymbol{\Sigma}_{\boldsymbol{x}})$, where "$\stackrel{iid}{\sim}$" means "independent and identically distributed as", and considering that the model errors ($\boldsymbol{q}_i$, $\boldsymbol{\eta}_{\boldsymbol{y}_i}$ and $\boldsymbol{\eta}_{\boldsymbol{x}_i}$) and $\boldsymbol{x}_i$ are independent, we have that the joint distribution of the observed variables can be expressed as

$$\begin{pmatrix} \boldsymbol{Y}_i \\ \boldsymbol{X}_i \end{pmatrix} \stackrel{ind}{\sim} \mathcal{N}_{v+m}\left[\begin{pmatrix} \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\boldsymbol{\mu}_{\boldsymbol{x}} \\ \boldsymbol{\mu}_{\boldsymbol{x}} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\beta}_1\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}_1^\top + \boldsymbol{\Sigma}_{\boldsymbol{q}} + \boldsymbol{\tau}_{\boldsymbol{y}_i} & \boldsymbol{\beta}_1\boldsymbol{\Sigma}_{\boldsymbol{x}} \\ \boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}_1^\top & \boldsymbol{\Sigma}_{\boldsymbol{x}} + \boldsymbol{\tau}_{\boldsymbol{x}_i} \end{pmatrix}\right]. \quad (3)$$

Note that in (3), the mean vector and the covariance-variance matrix of observed variables have the matrix $\boldsymbol{\beta}_1$ in common, i.e., they share mv parameters. Kulathinal *et al.* (2002) study the univariate case (when $v = 1$ and $m = 1$) and propose an EM (Expectation and Maximization) algorithm to obtain MLEs for model parameters. In addition, they derived the asymptotic variance of the MLE of the inclination parameter making it possible to build hypotheses testing of it. Also, de Castro *et al.* (2008) derive the observed and expected Fisher information and conduct some simulation studies to investigate the behavior of the likelihood ratio, score, Wald and $C(\alpha)$ statistics for testing hypothesis of the parameters and Patriota *et al.* (2009) study the asymptotic properties of method-of-moments estimators in the univariate model proposed by Kulathinal *et al.* (2002). Model (2) is a multivariate version of the model proposed by Kulathinal *et al.* (2002).

3 Second-order bias of $\widehat{\boldsymbol{\theta}}$

In order to follow the same scheme adopted by Patriota and Lemonte (2009), define the vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}_0^\top, \mathrm{vec}(\boldsymbol{\beta}_1)^\top, \boldsymbol{\mu}_{\boldsymbol{x}}^\top, \mathrm{vech}(\boldsymbol{\Sigma}_{\boldsymbol{x}})^\top, \mathrm{vech}(\boldsymbol{\Sigma}_{\boldsymbol{q}})^\top)^\top$, where $\mathrm{vec}(\cdot)$ is the vec operator, which transforms a matrix into a vector by stacking the columns of the matrix and $\mathrm{vech}(\cdot)$ is the vech operator, which transforms a symmetric matrix into a vector by stacking the on or above diagonal elements. Also, consider $\boldsymbol{Z}_i = (\boldsymbol{Y}_i^\top, \boldsymbol{X}_i^\top)^\top$ and the mean and covariance-variance function as

$$\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\boldsymbol{\mu}_{\boldsymbol{x}} \\ \boldsymbol{\mu}_{\boldsymbol{x}} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\beta}_1\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}_1^\top + \boldsymbol{\Sigma} + \boldsymbol{\tau}_{\boldsymbol{y}_i} & \boldsymbol{\beta}_1\boldsymbol{\Sigma}_{\boldsymbol{x}} \\ \boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}_1^\top & \boldsymbol{\Sigma}_{\boldsymbol{x}} + \boldsymbol{\tau}_{\boldsymbol{x}_i} \end{pmatrix},$$

respectively.

Moreover, to simplify notation, define the quantities $\boldsymbol{Z} = \mathrm{vec}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)$, $\boldsymbol{\mu} = \mathrm{vec}(\boldsymbol{\mu}_1(\boldsymbol{\theta}), \ldots, \boldsymbol{\mu}_n(\boldsymbol{\theta}))$, $\boldsymbol{\Sigma} = \mathrm{diag}\{\boldsymbol{\Sigma}_1(\boldsymbol{\theta}), \ldots, \boldsymbol{\Sigma}_n(\boldsymbol{\theta})\}$ and $\boldsymbol{u} = \boldsymbol{Z} - \boldsymbol{\mu}$. The log-likelihood function for the vector parameter $\boldsymbol{\theta}$ from a random sample, except for constants, can be expressed as

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\boldsymbol{u}^\top\}. \tag{4}$$

Additionally, for the purpose of computing the score function, the Fisher information and the second-order biases, also define

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

and

$$\begin{gathered}\boldsymbol{F}^{(r)}_{\boldsymbol{\beta}_0} = \frac{\partial\boldsymbol{\beta}_0}{\partial\theta_r}, \quad \boldsymbol{F}^{(s)}_{\boldsymbol{\beta}_1} = \frac{\partial\boldsymbol{\beta}_1}{\partial\theta_s}, \quad \boldsymbol{F}^{(s)}_{\boldsymbol{\mu}_{\boldsymbol{x}}} = \frac{\partial\boldsymbol{\mu}_{\boldsymbol{x}}}{\partial\theta_s}, \\ \boldsymbol{F}^{(s)}_{\boldsymbol{\Sigma}_{\boldsymbol{x}}} = \frac{\partial\boldsymbol{\Sigma}_{\boldsymbol{x}}}{\partial\theta_s} \quad \text{and} \quad \boldsymbol{F}^{(s)}_{\boldsymbol{\Sigma}_{\boldsymbol{q}}} = \frac{\partial\boldsymbol{\Sigma}_{\boldsymbol{q}}}{\partial\theta_s},\end{gathered} \tag{5}$$

with $r, s = 1, 2, \ldots, p$, where p is the dimension of $\boldsymbol{\theta}$. The quantities (5) are vectors or matrices of zeros with a unit in the position referring to the s^{th} element of $\boldsymbol{\theta}$. Let $\widetilde{\boldsymbol{D}} = (\boldsymbol{a}_{\boldsymbol{\beta}_0}, \boldsymbol{a}_{\boldsymbol{\beta}_1}, \boldsymbol{a}_{\boldsymbol{\mu}_{\boldsymbol{x}}}, \boldsymbol{0}, \boldsymbol{0})$ and $\widetilde{\boldsymbol{V}} = (\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\beta}_1}, \boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\Sigma}_{\boldsymbol{x}}}, \boldsymbol{C}_{\boldsymbol{\Sigma}_{\boldsymbol{q}}})$, with $\boldsymbol{a}_{\boldsymbol{\beta}_0} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_v)$, $\boldsymbol{a}_{\boldsymbol{\beta}_1} = (\boldsymbol{a}_{v+1}, \ldots, \boldsymbol{a}_{v(m+1)})$, $\boldsymbol{a}_{\boldsymbol{\mu}_{\boldsymbol{x}}} = (\boldsymbol{a}_{v(m+1)+1}, \ldots, \boldsymbol{a}_{v(m+1)+m})$, $\boldsymbol{C}_{\boldsymbol{\beta}_1} = \big(\mathrm{vec}(\boldsymbol{C}_{v+1}), \ldots, \mathrm{vec}(\boldsymbol{C}_{v(m+1)})\big)$, $\boldsymbol{C}_{\boldsymbol{\Sigma}_{\boldsymbol{x}}} = \big(\mathrm{vec}(\boldsymbol{C}_{(v+1)(m+1)}), \ldots, \mathrm{vec}(\boldsymbol{C}_{p'})\big)$ and $\boldsymbol{C}_{\boldsymbol{\Sigma}_{\boldsymbol{q}}} = \big(\mathrm{vec}(\boldsymbol{C}_{p'+1}), \ldots, \mathrm{vec}(\boldsymbol{C}_p)\big)$, where $p' = v(m+1) + m + m(m+1)/2$.

The first derivative of (4) with respect to the r^{th} element of $\boldsymbol{\theta}$ is

$$U_r = \frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_r(\boldsymbol{\Sigma} - \boldsymbol{u}\boldsymbol{u}^\top)\} + \mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_r\boldsymbol{u}^\top\}; \tag{6}$$

the expectation of the derivative of (6) with respect to the s^{th} element of $\boldsymbol{\theta}$ is given by

$$\kappa_{sr} = \frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_r\boldsymbol{C}_s\} - \boldsymbol{a}_s^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_r.$$

Note that, under general regularity conditions (Cox and Hinkley, 1974, Ch. 9), $-\kappa_{sr}$ is the $(s,r)^{\text{th}}$ element of the expected Fisher information. The score function and the expected Fisher information are given, respectively, by $\boldsymbol{U}_{\boldsymbol{\theta}} = \widetilde{\boldsymbol{D}}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{u} - \frac{1}{2}\widetilde{\boldsymbol{V}}^\top\widetilde{\boldsymbol{\Sigma}}^{-1}\mathrm{vec}(\boldsymbol{\Sigma} - \boldsymbol{u}\boldsymbol{u}^\top)$ and $\boldsymbol{K}_{\boldsymbol{\theta}} = \widetilde{\boldsymbol{D}}^\top\boldsymbol{\Sigma}^{-1}\widetilde{\boldsymbol{D}} + \frac{1}{2}\widetilde{\boldsymbol{V}}^\top\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{V}}$, with $\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}$ and $\otimes$ is the Kronecker product. Defining

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

we can write the score function and the Fisher information in a short form as

$$\boldsymbol{U}_{\boldsymbol{\theta}} = \widetilde{\boldsymbol{F}}^\top\widetilde{\boldsymbol{H}}\widetilde{\boldsymbol{u}} \quad \text{and} \quad \boldsymbol{K}_{\boldsymbol{\theta}} = \widetilde{\boldsymbol{F}}^\top\widetilde{\boldsymbol{H}}\widetilde{\boldsymbol{F}}.$$

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

$$\boldsymbol{\theta}^{(m+1)} = (\widetilde{\boldsymbol{F}}^{(m)\top}\widetilde{\boldsymbol{H}}^{(m)}\widetilde{\boldsymbol{F}}^{(m)})^{-1}\widetilde{\boldsymbol{F}}^{(m)\top}\widetilde{\boldsymbol{H}}^{(m)}\widetilde{\boldsymbol{u}}^{*(m)}, \quad m = 0, 1, 2, \ldots, \tag{7}$$

where $\widetilde{\boldsymbol{u}}^{*(m)} = \widetilde{\boldsymbol{F}}^{(m)}\boldsymbol{\theta}^{(m)} + \widetilde{\boldsymbol{u}}^{(m)}$. Each loop, through the iterative scheme (7), consists of an iterative re-weighted least squares algorithm to optimize the log-likelihood (4). Using equation (7) and any software (`MAPLE`, `MATLAB`, `Ox`, `R`, `SAS`) with a weighted linear regression routine one can compute the MLE, $\widehat{\boldsymbol{\theta}}$, iteratively. Initial approximation $\boldsymbol{\theta}^{(0)}$ for the iterative algorithm is used to evaluate $\widetilde{\boldsymbol{F}}^{(0)}$, $\widetilde{\boldsymbol{H}}^{(0)}$ and $\widetilde{\boldsymbol{u}}^{*(0)}$ from which these equations can be used to obtain the next estimate $\boldsymbol{\theta}^{(1)}$. This new value can update $\widetilde{\boldsymbol{F}}$, $\widetilde{\boldsymbol{H}}$ and $\widetilde{\boldsymbol{u}}^{*}$ and so the iterations continue until convergence is achieved.

The general matrix formulae derived by Patriota and Lemonte (2009) for n^{-1} bias vector $\boldsymbol{B}(\widehat{\boldsymbol{\theta}})$ of $\widehat{\boldsymbol{\theta}}$ is given by

$$\boldsymbol{B}(\widehat{\boldsymbol{\theta}}) = (\widetilde{\boldsymbol{F}}^{\top}\widetilde{\boldsymbol{H}}\widetilde{\boldsymbol{F}})^{-1}\widetilde{\boldsymbol{F}}^{\top}\widetilde{\boldsymbol{H}}\widetilde{\boldsymbol{\xi}}, \tag{8}$$

where $\widetilde{\boldsymbol{\xi}} = (\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p)\mathrm{vec}\{(\widetilde{\boldsymbol{F}}^{\top}\widetilde{\boldsymbol{H}}\widetilde{\boldsymbol{F}})^{-1}\}$ and $\boldsymbol{\Phi}_r = -\frac{1}{2}(\boldsymbol{G}_r + \boldsymbol{J}_r)$, $r = 1, 2, \ldots, p$, 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 $\boldsymbol{I}_k$ denotes an $k \times k$ identity matrix. The bias vector $\boldsymbol{B}(\widehat{\boldsymbol{\theta}})$ is simply the set coefficients from the ordinary weighted lest-squares regression of the $\widetilde{\boldsymbol{\xi}}$ on the columns of $\widetilde{\boldsymbol{F}}$, using weights in $\widetilde{\boldsymbol{H}}$. The bias vector $\boldsymbol{B}(\widehat{\boldsymbol{\theta}})$ will be small when $\widetilde{\boldsymbol{\xi}}$ is orthogonal to the columns of $\widetilde{\boldsymbol{H}}$ and it can be large when n is small. Note that equation (8) involves simple operations on matrices and vectors and we can calculate 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, 2006) with minimal effort.

After a somewhat algebra, we have

$$\boldsymbol{a}_r = \boldsymbol{1}_n \otimes \begin{pmatrix} \boldsymbol{F}^{(r)}_{\boldsymbol{\beta}_0} \\ \boldsymbol{0} \end{pmatrix}, \quad \boldsymbol{a}_s = \boldsymbol{1}_n \otimes \begin{pmatrix} \boldsymbol{F}^{(s)}_{\boldsymbol{\beta}}\boldsymbol{\mu}_{\boldsymbol{x}} \\ \boldsymbol{0} \end{pmatrix}, \quad \boldsymbol{a}_t = \boldsymbol{1}_n \otimes \begin{pmatrix} \boldsymbol{\beta}\boldsymbol{F}^{(t)}_{\boldsymbol{\mu}_{\boldsymbol{x}}} \\ \boldsymbol{F}^{(t)}_{\boldsymbol{\mu}_{\boldsymbol{x}}} \end{pmatrix} \quad \text{and} \quad \boldsymbol{a}_u = \boldsymbol{0},$$

for $r = 1, \ldots, v$; $s = v + 1, \ldots, v(m+1)$; $t = v(m+1)+1, \ldots, v(m+1)+m$; and $u = (v+1)(m+1), \ldots, p$; where $p = v(m+1) + m + m(m+1)/2 + v(v+1)/2$. (Here, $\boldsymbol{1}_n$ denotes an $n \times 1$ vector of ones.) Moreover,

$$\boldsymbol{a}_{rs} = \boldsymbol{1}_n \otimes \begin{pmatrix} \boldsymbol{F}^{(s)}_{\boldsymbol{\beta}_1}\boldsymbol{F}^{(r)}_{\boldsymbol{\mu}_{\boldsymbol{x}}} \\ \boldsymbol{0} \end{pmatrix},$$

for all r and s,

$$\boldsymbol{C}_s = \boldsymbol{I}_n \otimes \begin{pmatrix} \boldsymbol{F}^{(s)}_{\boldsymbol{\beta}_1}\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}_1^{\top} + \boldsymbol{\beta}_1\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{F}^{(s)\top}_{\boldsymbol{\beta}_1} & \boldsymbol{F}^{(s)}_{\boldsymbol{\beta}_1}\boldsymbol{\Sigma}_{\boldsymbol{x}} \\ \boldsymbol{F}^{(s)}_{\boldsymbol{\beta}_1}\boldsymbol{\Sigma}_{\boldsymbol{x}} & \boldsymbol{0} \end{pmatrix}, \; \boldsymbol{C}_t = \boldsymbol{I}_n \otimes \begin{pmatrix} \boldsymbol{\beta}_1\boldsymbol{F}^{(t)}_{\boldsymbol{\Sigma}_{\boldsymbol{x}}}\boldsymbol{\beta}_1^{\top} & \boldsymbol{\beta}_1\boldsymbol{F}^{(t)}_{\boldsymbol{\Sigma}_{\boldsymbol{x}}} \\ \boldsymbol{F}^{(t)}_{\boldsymbol{\Sigma}_{\boldsymbol{x}}}\boldsymbol{\beta}_1^{\top} & \boldsymbol{0} \end{pmatrix}$$

and

$$\boldsymbol{C}_u = \boldsymbol{I}_n \otimes \begin{pmatrix} \boldsymbol{F}^{(u)}_{\boldsymbol{\Sigma}_{\boldsymbol{q}}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix},$$

for $s = v+1, \ldots, v(m+1)$; $t = v(m+1)+1, \ldots, v(m+1)+m$; and $u = (v+1)(m+1), \ldots, p$. Additionally,

$$\boldsymbol{C}_{rs} = \boldsymbol{I}_n \otimes \begin{pmatrix} \boldsymbol{F}_{\boldsymbol{\beta}_1}^{(s)} \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{F}_{\boldsymbol{\beta}_1}^{(r)\top} + \boldsymbol{F}_{\boldsymbol{\beta}_1}^{(r)} \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{F}_{\boldsymbol{\beta}_1}^{(s)\top} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$$

and

$$\boldsymbol{C}_{tu} = \boldsymbol{I}_n \otimes \begin{pmatrix} \boldsymbol{F}_{\boldsymbol{\beta}_1}^{(u)} \boldsymbol{F}_{\boldsymbol{\sigma}_{\boldsymbol{x}}}^{(t)} \boldsymbol{\beta}_1^\top + \boldsymbol{\beta}_1 \boldsymbol{F}_{\boldsymbol{\Sigma}_{\boldsymbol{x}}}^{(t)} \boldsymbol{F}_{\boldsymbol{\beta}_1}^{(s)\top} & \boldsymbol{F}_{\boldsymbol{\beta}_1}^{(u)} \boldsymbol{F}_{\boldsymbol{\Sigma}_{\boldsymbol{x}}}^{(t)} \\ \boldsymbol{F}_{\boldsymbol{\Sigma}_{\boldsymbol{x}}}^{(t)} \boldsymbol{F}_{\boldsymbol{\beta}_1}^{(u)\top} & \boldsymbol{0} \end{pmatrix},$$

for $r, s, u = v+1, \ldots, v(m+1)$; $t = v(m+1)+1, \ldots, v(m+1)+m$; and $\boldsymbol{C}_{rs} = \boldsymbol{0}$ otherwise.

Therefore, in the measurement error model defined by the equations (1) and (2), all quantities necessary to compute the $O(n^{-1})$ bias of $\widehat{\boldsymbol{\theta}}$ using expression (8) are given. On the right-hand side of expression (8), 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 MLE of $\boldsymbol{B}(\cdot)$, that is, the unknown parameters are replaced by their MLEs. The bias-corrected estimate $\widetilde{\boldsymbol{\theta}}$ is expected to have better sampling properties than the uncorrected estimator, $\widehat{\boldsymbol{\theta}}$. In fact, we present some simulations in Section 5 to show that $\widetilde{\boldsymbol{\theta}}$ has smaller bias than its corresponding MLE, thus suggesting that the bias corrections have the effect of shrinking the modified estimates toward to the true parameter values.

4 Biases of the MLEs $\widehat{\boldsymbol{\mu}}_i$ and $\widehat{\boldsymbol{\Sigma}}_i$

In this section, we give matrix formulae for the $O(n^{-1})$ biases of the MLEs of the ith mean $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta})$ and ith variance-covariance vector $\boldsymbol{\Sigma}_i^* = \text{vech}(\boldsymbol{\Sigma}_i(\boldsymbol{\theta}))$. Let $q_1 = v + m$ and $q_2 = q_1(q_1+1)/2$. Additionally, let $\boldsymbol{A} = [\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n]^\top$ be a $np \times p$ matrix, where $\boldsymbol{A}_i$ is a $p \times p$ matrix, then we define $\text{tr}^*(\boldsymbol{A}) = [\text{tr}(\boldsymbol{A}_1), \ldots, \text{tr}(\boldsymbol{A}_n)]^\top$.

From a Taylor series expansion of $\widehat{\boldsymbol{\mu}}_i = \boldsymbol{\mu}_i(\widehat{\boldsymbol{\theta}})$, we obtain up to an error of order $O(n^{-2})$:

$$\boldsymbol{B}(\widehat{\boldsymbol{\mu}}_i) = \boldsymbol{L}_i \boldsymbol{B}(\widehat{\boldsymbol{\theta}}) + \frac{1}{2}\text{tr}^*[\boldsymbol{M}_i \text{Cov}(\widehat{\boldsymbol{\theta}})],$$

where $\boldsymbol{L}_i$ is a $q_1 \times p$ matrix of first partial derivatives $\partial \boldsymbol{\mu}_i / \partial \theta_r$ (for $r = 1, 2, \ldots, p$), $\boldsymbol{M}_i = [\boldsymbol{M}_{i1}, \ldots, \boldsymbol{M}_{iq_1}]^\top$ is a $q_1 p \times p$ matrix of second partial derivatives, where $\boldsymbol{M}_{il}$ is a $p \times p$ matrix with elements $\partial^2 \mu_{il} / \partial \theta_r \partial \theta_s$ (for $r, s = 1, \ldots, p$ and $l = 1, 2, \ldots, q_1$), $\text{Cov}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{K}_{\boldsymbol{\theta}}^{-1}$ is the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}$ and the vector $\boldsymbol{B}(\widehat{\boldsymbol{\theta}})$ was defined before. All quantities in the above equation should be evaluated at $\widehat{\boldsymbol{\theta}}$. The asymptotic variance of $\widehat{\boldsymbol{\mu}}_i$ can also be expressed explicitly in terms of the covariance of $\widehat{\boldsymbol{\theta}}$ by

$$\text{Var}(\widehat{\boldsymbol{\mu}}_i) = \boldsymbol{L}_i \text{Cov}(\widehat{\boldsymbol{\theta}}) \boldsymbol{L}_i^\top.$$

The second-order bias of $\widehat{\boldsymbol{\Sigma}}_i^*$ is obtained by expanding $\widehat{\boldsymbol{\Sigma}}_i^* = \boldsymbol{\Sigma}_i^*(\widehat{\boldsymbol{\theta}})$ in Taylor series. Then, the $O(n^{-1})$ bias of $\widehat{\boldsymbol{\Sigma}}_i^*$ is written as:

$$\boldsymbol{B}(\widehat{\boldsymbol{\Sigma}}_i^*) = \boldsymbol{L}_i^* \boldsymbol{B}(\widehat{\boldsymbol{\theta}}) + \frac{1}{2}\text{tr}^*[\boldsymbol{M}_i^* \text{Cov}(\widehat{\boldsymbol{\theta}})],$$

where $\boldsymbol{L}_i^*$ is a $q_2 \times p$ matrix of first partial derivatives $\partial \boldsymbol{\Sigma}_i^* / \partial \theta_r$ (for $r = 1, 2, \ldots, p$), $\boldsymbol{M}_i^* = [\boldsymbol{M}_{i1}^*, \ldots, \boldsymbol{M}_{iq_2}^*]^\top$ is a $q_2 p \times p$ matrix of second partial derivatives, where $\boldsymbol{M}_{il}^*$ is a $p \times p$ matrix with elements $\partial^2 \Sigma_{il}^* / \partial \theta_r \partial \theta_s$ (for $r, s = 1, \ldots, p$ and $l = 1, 2, \ldots, q_2$).

Therefore, we are now able to define the following second-order bias-corrected estimators for $\widehat{\boldsymbol{\mu}}_i$ and $\widehat{\boldsymbol{\Sigma}}_i^*$:

$$\widetilde{\boldsymbol{\mu}}_i = \widehat{\boldsymbol{\mu}}_i - \widehat{\boldsymbol{B}}(\widehat{\boldsymbol{\mu}}_i) \quad \text{and} \quad \widetilde{\boldsymbol{\Sigma}}_i^* = \widehat{\boldsymbol{\Sigma}}_i^* - \widehat{\boldsymbol{B}}(\widehat{\boldsymbol{\Sigma}}_i^*).$$

It is clear that the $O(n^{-1})$ bias of any other function of $\boldsymbol{\theta}$, say $\boldsymbol{\Psi}(\boldsymbol{\theta})$ $(h \times 1)$, can be obtained easily by Taylor series expansion:

$$\boldsymbol{B}(\widehat{\boldsymbol{\Psi}}) = \boldsymbol{\nabla}_{\boldsymbol{\Psi}}^{(1)} \boldsymbol{B}(\widehat{\boldsymbol{\theta}}) + \frac{1}{2} \mathrm{tr}^* [\boldsymbol{\nabla}_{\boldsymbol{\Psi}}^{(2)} \mathrm{Cov}(\widehat{\boldsymbol{\theta}})],$$

where $\boldsymbol{\nabla}_{\boldsymbol{\Psi}}^{(1)}$ is a $h \times p$ matrix of first partial derivatives $\partial \boldsymbol{\Psi} / \partial \theta_r$ (for $r = 1, 2, \ldots, p$) and $\boldsymbol{\nabla}_{\boldsymbol{\Psi}}^{(2)} = [\boldsymbol{\nabla}_{\boldsymbol{\Psi}1}^{(2)}, \ldots, \boldsymbol{\nabla}_{\boldsymbol{\Psi}h}^{(2)}]^\top$ is a $hp \times p$ matrix of second partial derivatives, where $\boldsymbol{\nabla}_{\boldsymbol{\Psi}l}^{(2)}$ is a $p \times p$ matrix with elements $\partial^2 \Psi_l / \partial \theta_r \partial \theta_s$ (for $r, s = 1, \ldots, p$ and $l = 1, 2, \ldots, h$).

5 Numerical results

We shall use Monte Carlo simulation to evaluate the finite sample performance of the MLEs attained using the iterative formula (7) and of their correspondent bias-corrected versions for a heteroskedastic errors-in-variables model presented in (2) with $m = v = 1$. The sample sizes considered were $n = 40, 60, 100$ and 200, the number of Monte Carlo replications was 10,000. All simulations were performed using the `R` programming language (R Development Core Team, 2006).

We consider the simple errors-in-variables model

$$Y_i = y_i + \eta_{y_i} \quad \text{and} \quad X_i = x_i + \eta_{x_i},$$

with $y_i | x_i \overset{ind}{\sim} \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$. This model was studied by Kulathinal *et al.* (2002). The errors η_{y_i} and η_{x_i} are independent of the unobservable covariate x_i and are distributed as

$$\begin{pmatrix} \eta_{y_i} \\ \eta_{x_i} \end{pmatrix} \overset{ind}{\sim} \mathcal{N}_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{y_i} & 0 \\ 0 & \tau_{x_i} \end{pmatrix} \right],$$

where the variances τ_{y_i} and τ_{x_i} are known for all $i = 1, \ldots, n$. Supposing in addition that $x_i \overset{iid}{\sim} \mathcal{N}(\mu_x, \sigma_x^2)$, we have that the joint distribution of the observed variables can be expressed as

$$\begin{pmatrix} Y_i \\ X_i \end{pmatrix} \overset{ind}{\sim} \mathcal{N}_2 \left[\begin{pmatrix} \beta_0 + \beta_1 \mu_x \\ \mu_x \end{pmatrix}, \begin{pmatrix} \beta_1^2 \sigma_x^2 + \tau_{y_i} + \sigma^2 & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \sigma_x^2 + \tau_{x_i} \end{pmatrix} \right].$$

Define $\boldsymbol{\theta} = (\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma^2)^\top$,

$$\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \begin{pmatrix} \beta_0 + \beta_1 \mu_x \\ \mu_x \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \begin{pmatrix} \beta_1^2 \sigma_x^2 + \sigma^2 + \tau_{y_i} & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \sigma_x^2 + \tau_{x_i} \end{pmatrix}.$$

From the previous expressions, we have immediately that

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

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

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

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

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

Thus, $\widetilde{\boldsymbol{D}} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{0}, \boldsymbol{0})$ and $\widetilde{\boldsymbol{V}} = (\boldsymbol{0}, \text{vec}(\boldsymbol{C}_2), \boldsymbol{0}, \text{vec}(\boldsymbol{C}_4), \text{vec}(\boldsymbol{C}_5))$. Therefore, all the quantities necessary to calculate $\boldsymbol{B}(\widehat{\boldsymbol{\theta}})$ using expression (8) are given.

In order to analyze the point estimation results, we computed, for each sample size and for each estimator: relative bias (the relative bias of an estimator $\widehat{\theta}$ is defined as $\{\mathbb{E}(\widehat{\theta}) - \theta\}/\theta$, its estimate being obtained by estimating $\mathbb{E}(\widehat{\theta})$ by Monte Carlo) and root mean square error, i.e., $\sqrt{\text{MSE}}$, where MSE is the mean squared error estimated from the 10,000 Monte Carlo replications. For practical reasons and without loss of generality, we adopt the same setting of parameters chosen by de Castro *et al.* (2008). (The parameters are the MLEs for the model parameters using a real data set presented in the next section.) We take $\beta_0 = -2$, $\beta_1 = 0.5$, $\mu_x = -2$, $\sigma_x^2 = 4$ and $\sigma^2 = 10$. We also consider two types of heteroskedasticity as studied by Patriota *et al.* (2009), namely: (a) $\sqrt{\tau_{x_i}} \sim U(0.5, 1.5)$ and $\sqrt{\tau_{y_i}} \sim U(0.5, 4)$, where $U(a, b)$ means uniform distribution on $[a, b]$; (b) $\sqrt{\tau_{x_i}} = 0.1|x_i|$ and $\sqrt{\tau_{y_i}} = 0.1|-2+0.51x_i|$, i.e., the variances depend on the unknown covariate. We remark that the variances are considered to be known and keep fixed in all Monte Carlo simulations.

Table 1 shows simulation results for an errors-in-variables model with a uniform heteroskedasticity. The figures in this table reveal that the maximum-likelihood estimators of the parameters can be substantially biased when the sample size is small, and that the bias correction we derived in the previous section is very effective. For instance, when $n = 40$ the biases of the estimators of β_0, β_1, μ_x, σ_x^2 and σ^2 average -0.02244 whereas the biases of the corresponding bias-adjusted estimators average -0.00276; that is, the average bias (in value absolute) of the MLEs is almost ten times greater than that of the corrected estimators. In particular, the maximum-likelihood estimators of σ_x^2 and σ^2 display substantial bias, and the bias correction proves to be quite effective when applied to these estimators.

Table 2 displays simulation results for an errors-in-variables model with a nonuniform heteroskedasticity. We note that the bias-adjusted estimator again displays smaller bias than the standard maximum-likelihood estimator. This suggests that the second-order bias of MLEs should not be ignored in samples of small to moderate sizes since they can be nonnegligible. Note also that root mean square error decrease with n, as expected. Additionally, we note that all estimators have similar root mean squared errors.

It is interesting to note that the finite-sample performance of the estimator of σ_x^2 deteriorate when we pass of the model with a uniform heteroskedasticity for the model with a nonuniform heteroskedasticity (see Tables 1 and 2). For instance, when $n = 100$, the relative biases of $\widehat{\sigma}_x^2$ (MLE) were -0.0135 (uniform heteroskedasticity) and -0.0484 (nonuniform heteroskedasticity), which amounts to an increase in relative biases of nearly 3.5 times.

Table 1: Relative bias and $\sqrt{\text{MSE}}$ of uncorrected and corrected estimates with a uniform heteroskedasticity: $\sqrt{\tau_{x_i}} \sim U(0.5, 1.5)$ and $\sqrt{\tau_{y_i}} \sim U(0.5, 4)$.

		MLE		BCE	
n	$\boldsymbol{\theta}$	Rel. bias	$\sqrt{\text{MSE}}$	Rel. bias	$\sqrt{\text{MSE}}$
40	β_0	−0.0173	0.99	−0.0043	0.97
	β_1	0.0315	0.38	0.0054	0.37
	μ_x	−0.0018	0.35	−0.0018	0.35
	σ_x^2	−0.0351	1.11	−0.0045	1.13
	σ^2	−0.0895	3.31	−0.0086	3.38
60	β_0	−0.0139	0.77	−0.0061	0.76
	β_1	0.0213	0.29	0.0058	0.29
	μ_x	0.0009	0.28	0.0009	0.28
	σ_x^2	−0.0239	0.89	−0.0036	0.90
	σ^2	−0.0548	2.60	−0.0018	2.64
100	β_0	−0.0100	0.68	−0.0037	0.67
	β_1	0.0168	0.26	0.0042	0.25
	μ_x	0.0001	0.25	0.0001	0.25
	σ_x^2	−0.0135	0.80	0.0022	0.81
	σ^2	−0.0424	2.40	0.0003	2.43
200	β_0	−0.0049	0.59	−0.0006	0.59
	β_1	0.0127	0.22	0.0041	0.22
	μ_x	0.0013	0.23	0.0013	0.23
	σ_x^2	−0.0116	0.70	0.0008	0.70
	σ^2	−0.0350	2.09	−0.0014	2.11

BCE: bias-corrected estimator.

Table 2: Relative bias and $\sqrt{\text{MSE}}$ of uncorrected and corrected estimates with a nonuniform heteroskedasticity: $\sqrt{\tau_{x_i}} = 0.1|x_i|$ and $\sqrt{\tau_{y_i}} = 0.1|\beta_0 + \beta_1 x_i|$.

		MLE		BCE	
n	$\boldsymbol{\theta}$	Rel. bias	$\sqrt{\text{MSE}}$	Rel. bias	$\sqrt{\text{MSE}}$
40	β_0	−0.0026	0.73	−0.0018	0.73
	β_1	0.0292	0.27	0.0276	0.27
	μ_x	−0.0228	0.32	−0.0228	0.32
	σ_x^2	−0.0594	0.91	−0.0354	0.92
	σ^2	−0.0540	2.26	−0.0056	2.30
60	β_0	0.0008	0.59	0.0013	0.59
	β_1	0.0203	0.22	0.0192	0.22
	μ_x	−0.0208	0.26	−0.0208	0.26
	σ_x^2	−0.0502	0.76	−0.0340	0.75
	σ^2	−0.0332	1.85	−0.0002	1.88
100	β_0	0.0013	0.51	0.0016	0.51
	β_1	0.0184	0.19	0.0176	0.19
	μ_x	−0.0198	0.23	−0.0198	0.23
	σ_x^2	−0.0484	0.65	−0.0363	0.65
	σ^2	−0.0223	1.61	0.0027	1.64
200	β_0	0.0036	0.45	0.0039	0.45
	β_1	0.0165	0.17	0.0159	0.17
	μ_x	−0.0186	0.20	−0.0186	0.20
	σ_x^2	−0.0474	0.59	−0.0377	0.58
	σ^2	−0.0204	1.41	−0.0004	1.43

BCE: bias-corrected estimator.

6 Application

We shall now present an application of the model described in Section 2 where $v = m = 1$. We analyze a epidemiological data set from the WHO MONICA (World Health Organization Multinational MONitoring of trends and determinants in CArdiovascular disease) Project. This data set was previously studied by Kulathinal *et al.* (2002) and de Castro *et al.* (2008) where the ML approach was adopted to estimate the model parameters.

The main goal of this project is to monitor trends in cardiovascular diseases (y) and relate it with known risk factors (x). The latent variables y and x are linearly related as

$$y_i = \beta_0 + \beta_1 x_i + q_i, \quad i = 1, \ldots, n.$$

As the variables y_i and x_i are not directly observable, surrogate variables Y_i and X_i are observed in their place, respectively. Such surrogate variables are attained from an analytical treatment of the data collection process. The data set are divided into two groups, namely: men ($n = 38$) and women ($n = 36$).

In what follows, we compare the MLEs with the bias-corrected estimators. Table 3 presents the MLEs, its standard deviation, its second-order biases and the corrected estimates. It can be seen that, the greater is the standard deviation of the MLE, the more distant from zero is its respectively second-order bias. As concluded in the simulation studies, the biases of the variances estimates are larger than of those produced by the line estimators. The second-order biases of the MLEs can be expressed as a percentage of the MLEs. That is, for the men data set, the second-order biases are -0.21%, 0.85%, 0.00%, -2.92% and -9.21% of the total amount of the MLEs of β_0, β_1, μ_x, σ_x^2 and σ^2, respectively. For the women data set, the second-order biases are 52.96%, 1.21%, 0.00%, -3.16% and -10.19% of the MLEs of β_0, β_1, μ_x, σ_x^2 and σ^2, respectively. It shows that the second-order biases of the MLEs are more pronounced in the women data set, mainly for the intercept estimator.

Table 3: MLEs and bias-corrected estimates.

	Parameter	MLEs	S.E.	Bias	BCEs
Men	β_0	-2.0799	0.5285	0.0044	-2.0843
	β_1	0.4690	0.2339	0.0040	0.4650
	μ_x	-1.0924	0.3550	0.0000	-1.0924
	σ_x^2	4.3163	1.0969	-0.1261	4.4423
	σ^2	4.8883	1.7790	-0.4501	5.3384
	Parameter	MLEs	S.E.	Bias	BCEs
Women	β_0	0.0321	1.1121	0.0170	0.0151
	β_1	0.6790	0.4072	0.0082	0.6708
	μ_x	-2.0677	0.3386	0.0000	-2.0677
	σ_x^2	3.6243	0.9695	-0.1146	3.7389
	σ^2	11.0809	4.2425	-1.1289	12.2098

BCE: bias-corrected estimates.

7 Conclusions

We derive a bias-adjustment scheme to eliminate the second-order biases of the maximum-likelihood estimates in a heteroskedastic multivariate errors-in-variables regression model using the general matrix formulae for the second-order bias derived by Patriota and Lemonte (2009). The simulation results presented show that the MLEs can be considerably biased. The bias correction derived in this paper is very effective, even when the sample size is large. Indeed, the bias correction mechanism adopted yields modified maximum-likelihood estimates which are nearly unbiased. Additionally, many errors-in-variables models are special cases of the proposed model and the results obtained here can be easily particularized to these submodels. We also present an application to a real data set.

Acknowledgments

This work was partially supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil) and CNPq (Conselho Nacional de Desenvolvimento Cientﬁco e Tecnolǵico, Brazil). The authors thank Dr. Kari Kuulasmaa (National Public Health Institute, Finland) for kindly supplying the data of our application.

References

Akritas, M.G., Bershady, M.A. (1996). Linear regression for astronomical data with measurement errors and intrinsic scatter. *The Astrophysical Journal.* **470**:706–714.

Cook, R.D., Tsai, C., Wei, B. (1986). Bias in nonlinear regression. *Biometrika.* **73**:615–623.

Cordeiro, G.M. (1993). Bartlett corrections and bias correction for two heteroskedastic regression models. *Communications in Statistics, Theory and Methods.* **22**:169–188.

Cordeiro, G.M. (2008). Corrected maximum likelihood estimators in linear heteroskedastic regression models. *Brazilian Review of Econometrics.* **28**:53–67.

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.

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

de Castro, M., Galea, M., Bolfarine, H. (2008). Hypothesis testing in an errors-in-variables model with heteroskedastic measurement errors. *Statistics in Medicine.* **27**:5217–5234.

Doornik, J.A. (2006). *An Object-Oriented Matrix Language – Ox 4.* Timberlake Consultants Press, London. 5th ed.

Fuller, W. (1987). *Measurement Error Models.* Wiley: Chichester.

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.

Patriota, A.G., Lemonte, A.J. (2009). Bias correction in a multivariate normal regression model with general parameterization. Working paper `arXiv:0812.3612v3`.

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 (2006). R: A Language and Environment for Statistical Computing. Vienna, Austria.

Vasconcellos, K.L.P., Cordeiro, G.M. (1997). Approximate bias for multivariate nonlinear heteroskedastic regressions. *Brazilian Journal of Probability and Statistics.* **11**:141–159.