A heteroscedastic structural errors-in-variables model with equation error

A.G. Patriota[a,*], H. Bolfarine[a], M. de Castro[b]

[a] *Universidade de São Paulo, Instituto de Matemática e Estatística, São Paulo-SP, Brasil*
[b] *Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, São Carlos-SP, Brasil*

Abstract

It is not uncommon with astrophysical and epidemiological data sets that the variances of the observations are accessible from an analytical treatment of the data collection process. Moreover, in a regression model, heteroscedastic measurement errors and equation errors are common situations when modelling such data. This article deals with the limiting distribution of the maximum likelihood and method-of-moments estimators for the line parameters of the regression model. We use the delta method to achieve it, making it possible to build joint confidence regions and hypothesis testing. This technique produces closed expressions for the asymptotic covariance matrix of those estimators. In the moment approach we do not assign any distribution for the unobservable covariate while with the maximum likelihood approach, we assume a normal distribution. We also conduct simulation studies of rejections rates for Wald-type statistics in order to verify the test size and power. Practical applications are reported for a data set produced by the *Chandra* observatory and also from the WHO MONICA Project on cardiovascular disease.

Key words: Asymptotic theory, equation error, heteroscedasticity, measurement error models.

*Corresponding author

Email addresses: patriota@ime.usp.br (A.G. Patriota), hbolfar@ime.usp.br (H. Bolfarine), mcastro@icmc.usp.br (M. de Castro)

Preprint submitted to Elsevier *February 9, 2012*

1. Introduction

Heteroscedastic structural errors-in-variables models have been applied in astrophysics (Akritas and Bershady, 1996; Kelly, 2007; Kelly et al., 2008), epidemiology (Kulathinal et al., 2002; de Castro et al., 2008) and analytical chemistry (Cheng and Riu, 2006) to avoid bias in parameter estimation. It is essential to avoid this type of bias for interpretation of the results and hypothesis testing (for more details see Fuller, 1987, and references therein). The most commonly used method in astronomy is, perhaps, the BCES estimator (for bivariate correlated errors and intrinsic scatter) proposed by Akritas and Bershady (1996). The term *intrinsic scatter* refers to the equation error in astronomy terminology. They were the first to consider a heteroscedastic structural model without considering any distribution for the random quantities of the model. The authors focussed on the method-of-moments and the large sample theory to estimate the model parameters and their asymptotic covariance matrix. Recently, Kulathinal et al. (2002) studied the same model supposing normal distributions for the model random quantities and proposed an EM (Expectation and Maximization) algorithm to find the maximum likelihood (ML) estimates for the model parameters in the presence of measurement errors, heteroscedasticity and equation error (with independent errors). In the same paper, the authors suggested a way to derive only the asymptotic variance of the slope estimator. The model considered in Kulathinal et al. (2002), which is assumed throughout this paper, is given by

$$\begin{aligned} Y_i &= y_i + \eta_{yi}, \\ X_i &= x_i + \eta_{xi}, \end{aligned} \tag{1.1}$$

with $y_i|x_i \stackrel{ind}{\sim} \mathcal{N}(\beta_0 + \beta_1 x_i; \sigma^2)$ (where "$\stackrel{ind}{\sim}$" means "independently distributed as"), i.e., the model (1.1) has an equation error and σ^2 is the equation error variance. This equation error is justified by the influence of other factors than x_i on the variation in y_i (Cheng and Riu, 2006). The errors, η_{yi} and η_{xi} in (1.1) are independent of x_i and y_i and are distributed as

$$\begin{pmatrix} \eta_{yi} \\ \eta_{xi} \end{pmatrix} \stackrel{ind}{\sim} \mathcal{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} ; \begin{bmatrix} \tau_{yi} & 0 \\ 0 & \tau_{xi} \end{bmatrix} \right),$$

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

$$\begin{pmatrix} Y_i \\ X_i \end{pmatrix} \stackrel{ind}{\sim} \mathcal{N}_2 \left(\begin{pmatrix} \beta_0 + \beta_1 \mu_x \\ \mu_x \end{pmatrix} ; \begin{bmatrix} \beta_1^2 \sigma_x^2 + \tau_{yi} + \sigma^2 & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \sigma_x^2 + \tau_{xi} \end{bmatrix} \right). \tag{1.2}$$

On one hand, Kulathinal et al. (2002) proposed an EM algorithm to obtain maximum likelihood estimates for β_0, β_1, μ_x, σ^2 and σ_x^2. On the other hand, we can also resort to the method of moments (MM). As can be seen in Section 2, it is not necessary to assign any distribution for x_i in order to obtain the MM estimators for the location parameters and their limiting covariance matrix. It is sufficient to assume the existence of some moments for x_i to achieve such results.

The likelihood function associated with (1.2) is very complicated to deal with in the sense of finding its global maximum. An iterative procedure is needed (Kulathinal et al., 2002). Problems regarding iterative procedures for obtaining the ML estimates, mainly for small sample sizes and flat or not concave likelihood functions are well known. Sometimes the algorithm does not converge or a relative maximum is found instead of the global one. Although the ML estimators have the best asymptotic (large samples) covariance matrix, good alternatives may be found in small sample situations. The moment estimators can be used as an alternative approach and, in addition, its asymptotic properties hold in more general cases (when x_i has a non-normal distribution). Furthermore, it is possible to compute the moment estimators and their asymptotic covariance matrix using a simple calculator.

Considering the same model (1.1), Cheng and Riu (2006) derived an approximation for the asymptotic covariance matrix of the moment estimators based on estimating equations. Nevertheless, the solutions from the estimating equations are slightly different from the moment estimators proposed in this paper. The sample variance is defined with n in its denominator while we consider it to be $n-1$. However, the estimators are close for large sample sizes. The observed Fisher information of the parameters in model (1.2) is computed in de Castro et al. (2008). These authors also perform hypothesis testing using the likelihood ratio, score, Wald and $C(\alpha)$ statistics.

As can be seen in the simulation studies of this paper, it is important to have the limiting covariance matrix to improve the accuracy of the statistical inferences for the regression parameters. We conduct simulation studies due to the lack of general theoretical results regarding their performance in small and moderate sample sizes.

We use the delta method to find the limiting covariance matrix of the ML and MM estimators. Although Cheng and Riu (2006) gave a general way to estimate the asymptotic covariance matrix of the MM estimators without specifying the errors' distributions, under normality large samples are required to have reliable hypothesis testing. This is a somewhat expected behavior because their approach is more general and the normal case is just a particular case. Hence, when normality is verified for the model errors, we advise to use the limiting covariance matrices for the MM and ML approaches derived in this paper; in addition they lead to closed form expressions. When it comes to comparing the rejection rates under the null hypothesis, the MM approach with the limiting covariance matrix seems to be better (for small and moderate sample sizes), the ML approach appears to be in the middle and Cheng and Riu's approach seems to be the one with worst performance. We borrow the data sets analyzed in Kulathinal et al. (2002) in order to compare the results in our paper with the others and we also apply the model and methods in two astrophysical data sets obtained by the *Chandra* X-ray observatory.

Basically, the main goal of this paper is to refine the asymptotic distribution of the estimation approaches by finding the exact asymptotic covariance matrix of the estimators. It is important to emphasize that joint confidence regions were not studied before for the model considered in Kulathinal et al. (2002).

This paper is organized as follows. In Section 2 we derived the exact expression for the asymptotic covariance matrix of the MM estimators using the delta method. We must remark that the computation of this asymptotic covariance matrix does not depend on the distribution of x_i. Such limiting covariance seems not available in the literature. Hence, precise joint hypothesis testing and joint confidence regions can be defined. In Section 3, a closed form expression is obtained for the asymptotic covariance matrix of the ML estimators of (β_0, β_1), using the delta method and the invariance property of the maximum likelihood estimators. Results of simulation studies are reported in Section 4. Applications to the WHO MONICA project and to the *Chandra* observatory data are considered in Section 5 and we end the paper with conclusions and remarks in Section 6.

2. Moments approach

Moment estimators for the heteroscedastic structural linear errors-in-variables model with equation error are presented in Kulathinal et al. (2002). It is shown that the estimators have explicit solution and can be easily computed. Cheng and Riu (2006) proposed a general estimating function that generates the moment estimators by its minimization and, based on M-estimation theory (Huber, 1964), they suggested an estimate for the asymptotic covariance matrix of those estimators. Therefore, it is possible to use the estimated covariance matrix derived in their paper for testing hypotheses concerning the vector (β_0, β_1) using Wald-type statistics. However, under normality of the errors, the estimated covariance matrix proposed by Cheng and Riu (2006) seems to be not very accurate, as we can see in our simulation study (Section 4). In this section we find the asymptotic distribution of the MM estimator of (β_0, β_1) using the delta method and considering a normal distribution for the measurement errors. Although we do not assume any distribution for the covariate x_i, it is possible to find the asymptotic covariance matrix of the estimators of the regression parameters. First of all, to derive the main results of this section, we have to define the central moments of the distribution of x_i. Let $\mathbb{E}(x_i) = \mu_x$ and $\mathbb{E}[(x_i - \mu_x)^r] = \nu_r$ in such a way that $|\mu_x|, |\nu_k| < \infty$ for all $k = 2, 3, 4$, that is, we are supposing that the first four moments of x_i do exist. We regard x_i, η_{xi} and η_{yi} as mutually not correlated. Thereafter we must assume that the following assumptions hold:

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{\tau_{yi}}{n} = \tau_{y*} \ , \quad \lim_{n\to\infty} \sum_{i=1}^{n} \frac{\tau_{xi}}{n} = \tau_{x*},$$
$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{\tau_{xi}\tau_{yi}}{n} = \tau_{xy*}, \quad \lim_{n\to\infty} \sum_{i=1}^{n} \frac{\tau_{yi}^2}{n} = \tau_{y**} \text{ and } \lim_{n\to\infty} \sum_{i=1}^{n} \frac{\tau_{xi}^2}{n} = \tau_{x**}. \tag{2.1}$$

These conditions are needed to guarantee convergence of the estimators and the existence of their limiting covariance matrix. To compute the MM estimators we must define the quantities

$$\bar{Y} = \sum_{i=1}^{n} \frac{Y_i}{n}, \qquad \bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}, \qquad M_Y = \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{n-1},$$
$$M_X = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1} \qquad \text{and} \qquad M_{XY} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})Y_i}{n-1}.$$

Thus, it can be shown that $\bar{Y}$, $\bar{X}$, M_Y, M_X and M_{XY} converge in probability to $\beta_0+\beta_1\mu_x$, μ_x, $\beta_1^2\nu_2+\sigma^2+\tau_{y*}$, $\nu_2+\tau_{x*}$ and $\beta_1\nu_2$, respectively. The MM estimators are obtained by equating sample and population moments and solving with respect to the unknown parameters. As a result, they are given by

$$\widehat{\beta}_{0_{ME}} = \bar{Y} - \frac{M_{XY}}{M_X - \bar{\tau}_x}\bar{X}, \qquad \widehat{\beta}_{1_{ME}} = \frac{M_{XY}}{M_X - \bar{\tau}_x}, \qquad \widehat{\mu}_{x_{ME}} = \bar{X},$$
$$\widehat{\nu}_{2_{ME}} = M_X - \bar{\tau}_x \quad \text{and} \quad \widehat{\sigma}^2_{ME} = M_Y - \frac{M_{XY}^2}{M_X - \bar{\tau}_x} - \bar{\tau}_y, \tag{2.2}$$

where $\bar{\tau}_x = \sum_{i=1}^{n} \tau_{xi}/n$ and $\bar{\tau}_y = \sum_{i=1}^{n} \tau_{yi}/n$. It is easy to check that the estimators (2.2) are consistent. There are some restrictions that need to hold; namely, $M_X > \bar{\tau}_x$ and $M_Y > M_{XY}^2/(M_X - \bar{\tau}_x) + \bar{\tau}_y$ to avoid inadmissible results.

As shown in Cheng and Riu (2006), there is another way to attain the moment estimators for β_0 and β_1. They can be viewed as "modified" least squares estimators, as described in Cheng and Van Ness (1999, Chap. 3), and may be obtained by minimizing the function

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n}\{(Y_i - \beta_0 - \beta_1 X_i)^2 - \tau_{yi} - \beta_1^2\tau_{xi}\},$$

leading to the following estimating equations:

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial\beta_0} = \sum_{i=1}^{n}\Phi_{0i}(\boldsymbol{\beta}) = 0, \qquad \frac{\partial Q(\boldsymbol{\beta})}{\partial\beta_1} = \sum_{i=1}^{n}\Phi_{1i}(\boldsymbol{\beta}) = 0, \tag{2.3}$$

where

$$\boldsymbol{\Phi}_i(\boldsymbol{\beta}) = \begin{pmatrix}\Phi_{0i}(\boldsymbol{\beta})\\ \Phi_{1i}(\boldsymbol{\beta})\end{pmatrix} = \begin{pmatrix} Y_i - \beta_0 - \beta_1 X_i \\ (Y_i - \beta_0 - \beta_1 X_i)X_i + \beta_1\tau_{xi}\end{pmatrix}$$

with $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$. Then, solving the equations (2.3) the MM estimators (2.2) are obtained with a little change. The quantities M_X, M_Y and M_{XY} are divided by n instead of $n-1$ as we have mentioned before.

As the MM estimators (2.2) present explicit solutions, we can use the delta method (Lehmann and Casella, 1998) for studying their asymptotic properties. That is, it is possible to show that, under some regularity conditions, when $\sqrt{n}(\boldsymbol{w} - \boldsymbol{\delta}) \xrightarrow{D} \mathcal{N}(\mathbf{0};\boldsymbol{\Pi})$, then $\sqrt{n}(g(\boldsymbol{w}) - g(\boldsymbol{\delta})) \xrightarrow{D} \mathcal{N}(\mathbf{0};\boldsymbol{P})$, where "$\xrightarrow{D}$" means convergence in distribution, $g(\cdot)$ is a continuous function, $\boldsymbol{P} = g'(\boldsymbol{\delta})\boldsymbol{\Pi}g'(\boldsymbol{\delta})^\top$ and $g'(\boldsymbol{\delta}) = \frac{\partial g(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}^\top} \neq \mathbf{0}$.

Hence, we have, initially, to find the asymptotic distribution of the vector $(\bar{Y}, \bar{X}, M_X, M_{XY})$ in order to derive the asymptotic covariance matrix of $(\widehat{\beta}_{0_{ME}}, \widehat{\beta}_{1_{ME}})$. One can check that the asymptotic distribution of $(\bar{Y}, \bar{X}, M_X - \bar{\tau}_x, M_{XY})$ is such that

$$\sqrt{n}\begin{pmatrix}\bar{Y} - \beta_0 - \beta_1\mu_x\\ \bar{X} - \mu_x \\ M_X - \bar{\tau}_x - \nu_2 \\ M_{XY} - \beta_1\nu_2\end{pmatrix} \xrightarrow{D} \mathcal{N}_4(\mathbf{0};\boldsymbol{L}), \tag{2.4}$$

where

$$\boldsymbol{L} = \begin{bmatrix} \beta_1^2\nu_2 + \sigma^2 + \tau_{y*} & \beta_1\nu_2 & \beta_1\nu_3 & \beta_1^2\nu_3 \\ \cdot & \nu_2 + \tau_{x*} & \nu_3 & \beta_1\nu_3 \\ \cdot & \cdot & L_{33} & L_{34} \\ \cdot & \cdot & \cdot & L_{44}\end{bmatrix}$$

with $L_{33} = 2\tau_{x**} + 4\nu_2\tau_{x*} + \nu_4 - \nu_2^2$, $L_{34} = \beta_1\nu_4 + 2\beta_1\nu_2\tau_{x*} - \beta_1\nu_2^2$ and $L_{44} = \beta_1^2\nu_4 + \beta_1^2\tau_{x*}\nu_2 + \tau_{y*}\nu_2 + \tau_{xy*} + \sigma^2\nu_2 + \tau_{x*}\sigma^2 - \beta_1^2\nu_2^2$. In order to obtain the elements of $\boldsymbol{L}$, one can rewrite the observed variables as

$$\begin{aligned} Y_i - \mu_y &= \beta_1(x_i - \mu_x) + \eta_{yi} + q_i \quad \text{and} \\ X_i - \mu_x &= (x_i - \mu_x) + \eta_{xi}, \end{aligned} \tag{2.5}$$

where $\mu_y = \beta_0 + \beta_1\mu_x$ and $q_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Thus, as x_i, η_{yi}, q_i and η_{xi} are mutually uncorrelated, the element L_{ij} can be computed using the equations in (2.5).

The function $g(\cdot)$ which gives $\widehat{\boldsymbol{\beta}}_{ME} = (\widehat{\beta}_{0_{ME}}, \widehat{\beta}_{1_{ME}})^\top$ from $(\bar{Y}, \bar{X}, M_X - \bar{\tau}_x, M_{XY})$ can be written as

$$\widehat{\boldsymbol{\beta}}_{ME} = g(\bar{Y}, \bar{X}, M_X - \bar{\tau}_x, M_{XY}) = \begin{pmatrix} \bar{Y} - \dfrac{M_{XY}}{M_X - \bar{\tau}_x}\bar{X} \\ \\ \dfrac{M_{XY}}{M_X - \bar{\tau}_x} \end{pmatrix}$$

and, defining $\boldsymbol{w} = (w_1, w_2, w_3, w_4)^\top$ and $\boldsymbol{\delta} = (\beta_0 + \beta_1\mu_x,\ \mu_x,\ \nu_2,\ \beta_1\nu_2)^\top$, its derivative is given by

$$g'(\boldsymbol{\delta}) = \left.\frac{\partial g(\boldsymbol{w})}{\partial \boldsymbol{w}^\top}\right|_{\boldsymbol{w}=\boldsymbol{\delta}} = \begin{bmatrix} 1 & -\beta_1 & \dfrac{\beta_1\mu_x}{\nu_2} & -\dfrac{\mu_x}{\nu_2} \\ \\ 0 & 0 & -\dfrac{\beta_1}{\nu_2} & \dfrac{1}{\nu_2} \end{bmatrix}.$$

Then, using the delta method we have that $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{ME} - \boldsymbol{\beta}) \xrightarrow{D} \mathcal{N}_2(\mathbf{0}_2; \boldsymbol{\Psi}(\boldsymbol{\theta}))$, where $\boldsymbol{\theta} = (\beta_0, \beta_1, \mu_x, \nu_2, \sigma^2)^\top$ and

$$\boldsymbol{\Psi}(\boldsymbol{\theta}) = g'(\boldsymbol{\delta})\boldsymbol{L}g'(\boldsymbol{\delta})^\top = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{12} & \psi_{22} \end{bmatrix}.$$

After somewhat tedious algebra we arrive at

$$\psi_{11} = \frac{\nu_2^2(\beta_1^2\tau_{x*} + \sigma^2 + \tau_{y*}) + 2\beta_1^2\mu_x^2(\tau_{x**} - \nu_2^2) + \mu_x^2\pi}{\nu_2^2},$$

$$\psi_{12} = -\frac{2\beta_1^2\mu_x(\tau_{x**} - \nu_2^2) + \mu_x\pi}{\nu_2^2} \quad \text{and} \quad \psi_{22} = \frac{2\beta_1^2(\tau_{x**} - \nu_2^2) + \pi}{\nu_2^2},$$

where $\pi = \beta_1^2\nu_2\tau_{x*} + \sigma^2\nu_2 + \tau_{xy*} + \sigma^2\tau_{x*} + \nu_2\tau_{y*} + 2\beta_1^2\nu_2^2$. Note that, these expressions do not depend on ν_3 and ν_4, since they cancel out. For this reason, the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{ME}$ is robust against the distribution of x_i. This asymptotic covariance matrix is still valid when $\sigma^2 = 0$. If $\beta_0 = 0$, we can use the moment estimator for β_1 and its respective asymptotic variance to infer values about β_1. Furthermore, if the covariate is not subject to measurement errors, one can simply take $\tau_{x*} = \tau_{x**} = \tau_{xy*} = \tau_{xi} = 0$ for all $i = 1, \ldots, n$ in the estimators and in the covariance matrix above.

The matrix $\boldsymbol{\Psi}(\boldsymbol{\theta})$ may be estimated using the MM estimators (2.2). Moreover, τ_{x*}, τ_{y*}, τ_{x**} and τ_{xy*} must be replaced with $\bar{\tau}_x$, $\bar{\tau}_y$, $\sum_{i=1}^n \tau_{xi}^2/n$ and $\sum_{i=1}^n \tau_{xi}\tau_{yi}/n$, respectively. Therefore, letting $\widehat{\boldsymbol{\Psi}}(\widehat{\boldsymbol{\theta}}_{ME})$ be the estimated asymptotic covariance matrix, for testing $\mathrm{H}_0 : \boldsymbol{G}\boldsymbol{\beta} = \boldsymbol{d}$ we can use the Wald-type statistic

$$\xi_{ME_1} = n\left(\boldsymbol{G}\widehat{\boldsymbol{\beta}}_{ME} - \boldsymbol{d}\right)^\top \left[\boldsymbol{G}\widehat{\boldsymbol{\Psi}}(\widehat{\boldsymbol{\theta}}_{ME})\boldsymbol{G}^\top\right]^{-1}\left(\boldsymbol{G}\widehat{\boldsymbol{\beta}}_{ME} - \boldsymbol{d}\right), \qquad (2.6)$$

which converges in distribution, under H_0, to $\chi^2(s)$, where $s = rank(G)$.

On the other hand, the Wald-type statistic built using the covariance matrix that follows from the "adjusted" least squares estimating equations (2.3), see Cheng and Riu (2006), can be written as

$$\xi_{ME_2} = n\left(\boldsymbol{G}\widehat{\boldsymbol{\beta}}_{ME*} - \boldsymbol{d}\right)^\top \left[\boldsymbol{G}\boldsymbol{A}_n^{-1}\boldsymbol{B}_n\boldsymbol{A}_n^{-1}\boldsymbol{G}^\top\right]^{-1}\left(\boldsymbol{G}\widehat{\boldsymbol{\beta}}_{ME*} - \boldsymbol{d}\right), \qquad (2.7)$$

where $\widehat{\boldsymbol{\beta}}_{ME*}$ is the moment estimator for $\boldsymbol{\beta}$ using Cheng and Riu's approach,

$$\boldsymbol{A}_n = \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{\Phi}_i(\boldsymbol{\beta}) = \frac{1}{n}\begin{bmatrix} -n & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & \sum_{i=1}^n(-X_i^2 + \tau_{xi}) \end{bmatrix}$$

and

$$\boldsymbol{B}_n = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\Phi}_i(\widehat{\boldsymbol{\beta}}_{ME*})\boldsymbol{\Phi}_i(\widehat{\boldsymbol{\beta}}_{ME*})^\top = \frac{1}{n}\begin{bmatrix} \widehat{b}_{00} & \widehat{b}_{01} \\ \widehat{b}_{01} & \widehat{b}_{11} \end{bmatrix}$$

with $b_{00} = \sum_{i=1}^n(Y_i - \beta_0 - \beta_1 X_i)^2$, $b_{01} = \sum_{i=1}^n(Y_i - \beta_0 - \beta_1 X_i)[(Y_i - \beta_0 - \beta_1 X_i)X_i + \beta_1\tau_{xi}]$ and $b_{11} = \sum_{i=1}^n\left[(Y_i - \beta_0 - \beta_1 X_i)X_i + \beta_1\tau_{xi}\right]^2$. The distributions of the Wald statistics (2.6) and (2.7) differ for finite samples but, under normality of the model errors, they are asymptotically equivalent.

3. Maximum likelihood approach

Under suitable regularity conditions, the ML approach generates efficient estimators, that is, the asymptotic covariance matrix attains the inverse of the Fisher information matrix (information matrix in short) and the bias of the ML estimates converges to zero when the sample size increases. Kulathinal et al. (2002) presents the EM algorithm to find the maximum likelihood estimators. In this paper we use the same notation as in Kulathinal et al. (2002). Thus, as part of their algorithm, they considered the following bivariate normal model for the unobservable variables (x_i, y_i):

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \overset{iid}{\sim} \mathcal{N}_2 \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}; \begin{bmatrix} \sigma_{11} & \rho\sqrt{\sigma_{11}\sigma_{22}} \\ \cdot & \sigma_{22} \end{bmatrix} \right),$$

where $\mu_1 = \mu_x$, $\mu_2 = \beta_0 + \beta_1\mu_x$, $\sigma_{22} = \beta_1^2\sigma_x^2 + \sigma^2$, $\sigma_{11} = \sigma_x^2$ and $\rho = \beta_1\sigma_x^2/\sqrt{\sigma_{11}\sigma_{22}}$. Then, based on the EM algorithm, it is possible to maximize the log-likelihood of $\boldsymbol{\theta^*} = (\mu_1, \mu_2, \sigma_{11}, \sigma_{22}, \rho)^\top$ associated with the observed variables (X_i, Y_i) and, using the invariance property of the ML estimators, one can estimate the parameters of interest $(\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma^2)$. Note that, if $\beta_0 = 0$ the ML estimators of β_1, μ_x, σ_x^2 and σ^2 can not be found using this technique. Notice also that, if the estimate of $\sigma_{22} - \sigma_{12}^2/\sigma_{11} < 0$ the estimator for $(\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma^2)$ attained using the approach in Kulathinal et al. (2002) is not the ML estimator (σ^2 can not take negative values). We recall that it is not the purpose of this paper to find the ML estimators since they are clearly stated in Kulathinal et al. (2002) using the EM algorithm.

Let $\boldsymbol{Z}_i = (X_i, Y_i)^\top$ and $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$, then the log-likelihood of $\boldsymbol{\theta^*}$ associated with the observed vector $\boldsymbol{Z}_i$ is given by

$$\ell(\boldsymbol{\theta^*}) \propto -\frac{1}{2}\sum_{i=1}^{n} \log|\boldsymbol{S}_i| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{Z}_i - \boldsymbol{\mu})^\top \boldsymbol{S}_i^{-1}(\boldsymbol{Z}_i - \boldsymbol{\mu}), \tag{3.1}$$

where

$$\boldsymbol{S}_i = \begin{bmatrix} \sigma_{11} + \tau_{xi} & \rho\sqrt{\sigma_{11}\sigma_{22}} \\ \cdot & \sigma_{22} + \tau_{yi} \end{bmatrix}, \quad i = 1, \ldots, n.$$

The expected information matrix of $(\sigma_{11}, \sigma_{22}, \rho)$ was also derived by Kulathinal et al. (2002), which is given by

$$\boldsymbol{I}(\sigma_{11}, \sigma_{22}, \rho) = \sum_{i=1}^{n} \boldsymbol{J}_i \boldsymbol{I}_i \boldsymbol{J}_i^\top$$

where

$$\boldsymbol{I}_i = \begin{bmatrix} \dfrac{2-\rho_i^2}{4(1-\rho_i^2)(\sigma_{11}+\tau_{xi})^2} & \dfrac{-\rho_i^2}{4(1-\rho_i^2)(\sigma_{11}+\tau_{xi})(\sigma_{22}+\tau_{yi})} & \dfrac{-\rho_i}{2(1-\rho_i^2)(\sigma_{11}+\tau_{xi})} \\ \cdot & \dfrac{2-\rho_i^2}{4(1-\rho_i^2)(\sigma_{22}+\tau_{yi})^2} & \dfrac{-\rho_i}{2(1-\rho_i^2)(\sigma_{22}+\tau_{yi})} \\ \cdot & \cdot & \dfrac{1+\rho_i^2}{(1-\rho_i^2)^2} \end{bmatrix},$$

$$\boldsymbol{J}_i = \begin{bmatrix} 1 & 0 & \dfrac{\rho_i \tau_{xi}}{2\sigma_{11}(\sigma_{11}+\tau_{xi})} \\ 0 & 1 & \dfrac{\rho_i \tau_{yi}}{2\sigma_{22}(\sigma_{22}+\tau_{yi})} \\ 0 & 0 & \dfrac{\rho_i}{\rho} \end{bmatrix} \quad \text{and} \quad \rho_i = \frac{\rho\sqrt{\sigma_{11}\sigma_{22}}}{\sqrt{(\sigma_{11}+\tau_{xi})(\sigma_{22}+\tau_{yi})}}.$$

The matrix $\boldsymbol{J}_i$ is derived in Kulathinal et al. (2002) and it is slightly incorrect in its elements $(1,3)$ and $(2,3)$ (de Castro et al., 2008), the correct one being presented here.

To obtain the asymptotic variance of the ML estimator of β_1, Kulathinal et al. (2002) suggested using the Jacobian of the transformation from $(\sigma_{11}, \sigma_{22}, \rho)$ to β_1. In this section, the main result is to find the asymptotic covariance matrix for the ML estimator of (β_0, β_1) using the delta method.

Therefore, differentiating twice (3.1) with respect to $\boldsymbol{\theta}^*$, computing its expectation and noting that

$$\mathbb{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\mu} \partial(\sigma_{11}, \sigma_{22}, \rho)}\right) = \mathbb{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial(\sigma_{11}, \sigma_{22}, \rho)^\top \partial \boldsymbol{\mu}^\top}\right)^\top = \mathbf{0}_{2\times 3},$$

we have the expected information matrix of $\boldsymbol{\theta}^*$ which is given by

$$\boldsymbol{I}_n(\boldsymbol{\theta}^*) = \sum_{i=1}^{n} \begin{bmatrix} \boldsymbol{S}_i^{-1} & \mathbf{0}_{2\times 3} \\ \mathbf{0}_{3\times 2} & \boldsymbol{J}_i \boldsymbol{I}_i \boldsymbol{J}_i^\top \end{bmatrix}.$$

Let $\widehat{\mu}_1$, $\widehat{\mu}_2$, $\widehat{\sigma}_{11}$, $\widehat{\sigma}_{22}$ and $\widehat{\rho}$ be the ML estimators of μ_1, μ_2, σ_{11}, σ_{22} and ρ, respectively. ML estimators have asymptotically normal distribution (which is a property of the ML estimators in the structural version since regularity conditions are satisfied). Then, by the invariance property (and assuming that $\widehat{\sigma}_{22} - \widehat{\sigma}_{12}^2/\widehat{\sigma}_{11} > 0$), the ML estimator of (β_0, β_1), $\widehat{\boldsymbol{\beta}}_{ML} = (\widehat{\beta}_{0_{ML}}, \widehat{\beta}_{1_{ML}})^\top$, is given by

$$\widehat{\boldsymbol{\beta}}_{ML} = h(\widehat{\boldsymbol{\theta}^*}) = \begin{pmatrix} \widehat{\mu}_2 - \widehat{\rho}\sqrt{\dfrac{\widehat{\sigma}_{22}}{\widehat{\sigma}_{11}}}\widehat{\mu}_1 \\ \\ \widehat{\rho}\sqrt{\dfrac{\widehat{\sigma}_{22}}{\widehat{\sigma}_{11}}} \end{pmatrix}$$

and differentiating h with respect to each of its elements, we have that

$$h'(\boldsymbol{\theta}^*) = \left.\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}^\top}\right|_{\boldsymbol{x}=\boldsymbol{\theta}^*} = \begin{bmatrix} -\rho\sqrt{\dfrac{\sigma_{22}}{\sigma_{11}}} & 1 & \dfrac{\rho\mu_1\sqrt{\sigma_{22}}}{2\sigma_{11}^{3/2}} & -\dfrac{\rho\mu_1}{2\sqrt{\sigma_{11}\sigma_{22}}} & -\mu_1\sqrt{\dfrac{\sigma_{22}}{\sigma_{11}}} \\ \\ 0 & 0 & -\dfrac{\rho\sqrt{\sigma_{22}}}{2\sigma_{11}^{3/2}} & \dfrac{\rho}{2\sqrt{\sigma_{11}\sigma_{22}}} & \sqrt{\dfrac{\sigma_{22}}{\sigma_{11}}} \end{bmatrix}.$$

Hence,

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta}\right) \xrightarrow{D} \mathcal{N}_2\left(\mathbf{0}_2; h'(\boldsymbol{\theta}^*)\boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta}^*)h'(\boldsymbol{\theta}^*)^\top\right)$$

where

$$\boldsymbol{\Gamma}(\boldsymbol{\theta}^*) = \lim_{n\to\infty}\frac{1}{n}\boldsymbol{I}_n(\boldsymbol{\theta}^*).$$

It is very complicated to state conditions for the existence of $\boldsymbol{\Gamma}(\boldsymbol{\theta}^*)$. These conditions involve functions which depend on the known variances τ_{xi} and τ_{yi}. For simplicity we consider that $\boldsymbol{\Gamma}(\boldsymbol{\theta}^*)$ exists (i.e., each element is strictly less than infinity) and that it is positive definite. Hence, we are able to build a Wald-type statistic for testing $\mathrm{H}_0 : \boldsymbol{G\beta} = \boldsymbol{d}$, which is given by

$$\xi_{ML} = n\left(\boldsymbol{G}\widehat{\boldsymbol{\beta}}_{ML} - \boldsymbol{d}\right)^\top \left[\boldsymbol{G}h'(\widehat{\boldsymbol{\theta}^*})\widehat{\boldsymbol{\Gamma}}^{-1}(\widehat{\boldsymbol{\theta}^*})h'(\widehat{\boldsymbol{\theta}^*})^\top\boldsymbol{G}^\top\right]^{-1}\left(\boldsymbol{G}\widehat{\boldsymbol{\beta}}_{ML} - \boldsymbol{d}\right), \quad (3.2)$$

where $\widehat{\boldsymbol{\Gamma}}(\widehat{\boldsymbol{\theta}^*}) = \dfrac{1}{n}\boldsymbol{I}_n(\widehat{\boldsymbol{\theta}^*})$.

The three Wald-type statistics (2.6), (2.7) and (3.2) have the same asymptotic distribution. Hence, confidence regions for $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ can be built from these Wald-type statistics. They are given, respectively, by

$$n\left(\widehat{\boldsymbol{\beta}}_{ME} - \boldsymbol{\beta}\right)^\top \widehat{\boldsymbol{\Psi}}^{-1}(\widehat{\boldsymbol{\theta}}_{ME})\left(\widehat{\boldsymbol{\beta}}_{ME} - \boldsymbol{\beta}\right) \leq \chi^2_{2,\gamma}, \quad (3.3)$$

$$n\left(\widehat{\boldsymbol{\beta}}_{ME}-\boldsymbol{\beta}\right)^\top \boldsymbol{A}_n\boldsymbol{B}_n^{-1}\boldsymbol{A}_n\left(\widehat{\boldsymbol{\beta}}_{ME}-\boldsymbol{\beta}\right)\leq\chi^2_{2,\gamma} \tag{3.4}$$

and

$$n\left(\widehat{\boldsymbol{\beta}}_{ML}-\boldsymbol{\beta}\right)^\top\left[h'(\widehat{\boldsymbol{\theta}^*})\widehat{\boldsymbol{\Gamma}}^{-1}(\widehat{\boldsymbol{\theta}^*})h'(\widehat{\boldsymbol{\theta}^*})^\top\right]^{-1}\left(\widehat{\boldsymbol{\beta}}_{ML}-\boldsymbol{\beta}\right)\leq\chi^2_{2,\gamma}, \tag{3.5}$$

where $\chi^2_{2,\gamma}$ denotes the γ-th quantile of the χ^2_2 distribution, $0<\gamma<1$. Moreover, ellipses can be plotted from (3.3)-(3.5).

4. Simulation study

In this section simulation studies are conducted for examining the behavior of the distributions of the Wald-type statistics (2.6), (2.7) and (3.2) for small and moderate sample sizes. In view of this, we planned Monte Carlo simulations to evaluate the empirical test sizes and the powers of the statistics at the 5% nominal level. The simulation setting is as follows:

$$\sigma^2=10, \qquad \sqrt{\tau_{xi}}\overset{iid}{\sim}U(0.5,1.5) \qquad \text{and} \qquad \sqrt{\tau_{yi}}\overset{iid}{\sim}U(0.5,4), \tag{4.1}$$

where $U(a,b)$ means uniform distribution on $[a,b]$. The setting (4.1) was selected in order to simulate observations similar to the real data sets from the WHO MONICA project. The empirical test size and the power are computed as follows. We consider a finite grid in the neighborhood of $(\beta_0,\beta_1)=(0,1)$; namely, $(-2,0.4)$, $(0,0.4)$, $(2,0.4)$, $(-2,1)$, $(0,1)$, $(2,1)$, $(-2,1.6)$, $(0,1.6)$ and $(2,1.6)$. We take the sample sizes $n=40$, 80 and 160. The moment estimators are used as initial values for starting the EM algorithm. The null hypothesis was $H_0:(\beta_0,\beta_1)=(0,1)$ in Tables 2–4 and $H_0:\beta_1=0$ in Tables 5–7 (under the latter hypothesis we consider the following values for β_1: -0.50, -0.25, 0.00, 0.25 and 0.50). For each triplet (β_0,β_1,n) we generate 10 000 Monte Carlo simulations and utilize the Wald-type statistics (2.6), (2.7) and (3.2) for testing if there exists evidence against the null hypothesis at the 5% (nominal) significance level. Under the null hypothesis, we expect to reject only 5% of the time. The variances τ_{xi} and τ_{yi} are generated for each sample size but kept fixed in all Monte Carlo simulations for each sample size.

We also study the rejection rates by using erroneously a naive model which does not consider the measurement errors in the covariate. We use the Wald statistic *MM 1* taking $\tau_{x*} = \tau_{x**} = \tau_{xy*} = \tau_{xi} = 0$ for $i = 1, 2, \ldots, n$ and denote this procedure as the naive approach and the Wald statistic using this procedure as *NA* (weighted least squares method). Table 1.a presents the test sizes of the hypothesis $H_0 : (\beta_0, \beta_1) = (0, 1)$ for the normal, half normal and Student t cases when the sample sizes are 40, 80 and 160 using the *NA*. Note that, as expected, the empirical test sizes depicted in Table 1.a are far away from the expected 5% nominal level. Other parameter settings were taken, but they have the same behavior (if we maintain the variances of the measurement errors with the same magnitude as previously defined). However, when we are testing hypothesis specifying that $\beta_1 = 0$ (see Table 1.b and Tables 5–7), the *NA* produces coherent results (test sizes close to the adopted nominal level). This happens because, under this hypothesis, there is no covariate effect and, consequently, there is no measurement error effect associated with the covariate.

Tables 2–4 depict the empirical test sizes (in the middle cells) and powers (around the middle cells) considering $x \overset{iid}{\sim} \mathcal{N}(-2, 4)$, $x \overset{iid}{\sim} \mathcal{HN}(-2, 4)$ and $x \overset{iid}{\sim} t(-2, 4, 5)$, respectively, where " $\mathcal{HN}(\mu, \sigma^2)$" means the half normal distribution with location μ and scale σ^2; " $t(\mu, \sigma^2, v)$" means the Student t distribution with location μ, scale σ^2 and v degrees of freedom. The same distributions set up was considered in Tables 5–7, but in these tables we maintained fixed $\beta_0 = -2$ (other simulations were developed but they had similar results and, for this reason, we omit them). The perturbation on the distribution of x may be severe. Firstly it is not perturbed, i.e, a normal distribution is considered. Next, we consider asymmetric and heavy tailed distributions for it in order to verify whether the Wald-type statistics are much affected. We denote the Wald-type statistic (2.6) as *MM 1*, the Wald-type statistic (2.7) as *MM 2* (it uses the asymptotic covariance matrix derived in Cheng and Riu (2006)) and the Wald-type statistic (3.2) as *ML*. Both asymptotic covariance matrices used in *MM 1* and *ML* have been derived in this paper.

Tables 5–7 show the rejection rates for $H_0 : \beta_1 = 0$ considering two sorts of heteroscedasticity; namely: (a) when $\sqrt{\tau_{xi}}$ and $\sqrt{\tau_{yi}}$ have uniform distributions as defined in (4.1), i.e., the variances do not depend on the covariate x_i and (b) when $\sqrt{\tau_{xi}} = 0.1|x_i|$ and $\sqrt{\tau_{yi}} = 0.1|\beta_0 + \beta_1 x_i|$. In general, in setting (b) the testing becomes more sensible and rejects more often than when (a) is considered, i.e., this sort of heteroscedasticity can interfere in the inferences.

As can be seen in all Tables 2–7, the *MM 1* and *ML*'s performances seem not to be affected by the distribution of x. Additionally, Cheng and Riu's approach is the most affected by the perturbations in the distribution of x when the sample size is small. These results are still valid for other parameter settings. Moreover, in the majority of cases, *ML* is the most powerful test, as expected. The low power presented when $x_i \stackrel{iid}{\sim} \mathcal{HN}(\mu, \sigma^2)$ in both Tables 3 and 6 might be explained by the fact that the x-values are generated with measurement error in a short range, making difficult the identification of the line's intercept and slope.

5. Applications

5.1. Epidemiology

Trends in cardiovascular diseases have been monitored by the WHO MONICA (World Health Organization Multinational MONItoring of trends and determinants in CArdiovascular disease) Project which was established in the early 1980s. The main objective of this project is related to changes in known risk factors (x) with the trends in cardiovascular mortality and coronary heart disease (y). In this paper, we analyze the same data set analyzed by Kulathinal et al. (2002) which are trends of the annual change in event rate (cardiovascular mortality) and trends of the risk scores for women (n=36) and for men (n=38) in each population. The risk score was defined as a linear combination of smoking status, systolic blood pressure, body mass index and total cholesterol level. Its coefficients were derived from a follow up study using proportional hazards models which can provide the observed risk score and its variance. For additional information, see Kulathinal et al. (2002). The observed response variable,

Y, is the average annual change in event rate (%) and the observed covariate, X, is the average annual change in the observed risk score (%).

The model without equation error ($\sigma^2 = 0$) is not adequate for this data set as shown by de Castro et al. (2008). Figure 2 displays 95% confidence regions for the three distinct methods (3.3)-(3.5) applied to men and women data sets. Notice from Table 8 that the standard errors of the estimators (in parenthesis) for β_0 and β_1 are always smaller if one uses *MM 2* than the other two approaches (*MM 1* and *ML*). The estimates seem to be close to each other (including the naive approach) except for σ^2, for which the ML estimate (4.89 for men data and 11.08 for women data) is very different from the *MM 1* (3.06 for men data and 6.43 for women data). Moreover, from Figure 2, it is clear that the hypothesis $H_0 : (\beta_0, \beta_1) = (0, 1)$ should be rejected for men but not for women. Data reveal that in the women's population, annual changes in event rate and risk score have the same numerical value.

Figure 1 presents the ellipses using (2.6), (2.7), (3.2) and it also presents ellipses using the naive approach. Figure 2 shows the fitted lines using the *MM 1*, *MM 2*, *ML* and *NA* approaches. Notice that the naive method produces attenuated estimates for the model slope.

5.2. Astrophysics

Active Galaxies and quasars emit a considerable fraction of their energy in X-rays. It is well accepted that the source of the X-ray emission involves accretion of hot plasma onto a supermassive black hole; however, there is considerable uncertainty regarding the structure of the accretion flow, and significant effort has gone into understanding it. In particular, we consider two applications related to X-ray emissions for using the proposed model and methods derived in this paper. There are many problems regarding the data collection such as sample selection and censoring, as discussed in Akritas and Bershady (1996) and Kelly (2007). The data set analyzed in this paper has no censoring, however, it is subject to sample selection as related in Kelly et al. (2008). We modeled this data set disregarding the bias produced by the data collection just to show the applicability of our approach. We are engaged in future researches to take into account these sample peculiarities.

In both data sets, the covariate is the base-10 logarithm of the ratio of luminosity (intrinsic brightness) at 2500 angstroms (250 nanometers) to the Eddington luminosity. The Eddington luminosity is a function of black hole mass. However, the black hole masses are unknown and must be estimated from the optical emission for each object. Because the estimated black hole masses are subject to measurement error, the estimated Eddington ratio is as well. In addition, it is possible to assess the precision related to this measure in each experimental unit (defining heteroscedastic errors).

The response variable for the first application is the X-ray photon index (also known as Γ_X), where larger values of it mean that more of the X-ray emission is being emitted at lower energies. The value of Γ_X and its uncertainty are obtained by fitting a model to an empirical spectrum. The fit is done by maximum-likelihood, and the standard error is essentially obtained by inverting the information matrix. In astronomy, standard errors are almost never estimated from replications, but instead are derived from an analytical treatment of the data collection process. The *Chandra* X-ray observatory collects light particles (photons) in the X-ray region of the electromagnetic spectrum. When it detects X-ray photons, it also records the energy of these photons. The result is a table of the number of X-rays detected as a function of energy; this is called a spectrum. The data are Poisson distributed, and a theoretical function (e.g., a power-law) is fitted to this data by maximum-likelihood. The estimate of Γ_X is the best fitting value of the exponent of this power-law, and the standard error in Γ_X is the estimated asymptotic variance, calculated by inverting the information matrix. By studying how the X-ray emission depends on this covariate, one can help to shed light on the nature of the X-ray emitting region.

The response variable for the second application is proportional to the base-10 logarithm of the ratio of optical/UV flux to X-ray flux (also known as α_{ox}). The variable α_{ox} is defined to be the ratio of the luminosity in the optical/UV band to that in the X-ray band, which is calculated from two separate observations – one in the optical/UV and one in the X-ray. A model spectrum is fitted separately to each band to calculate the luminosities, and measurement errors are derived from the best fit parameters, similar to the case for Γ_X. The optical/UV and X-ray observations are not simultaneous, and can be separated by several years. Because these objects (active galaxies, i.e., 'quasars') are known to be variable in their brightness, this contributes to the measurement error on the response (i.e., α_{ox}).

Tables 9–10 show the estimates (using the *MM 1*, *MM 2*, *ML* and *NA* approaches) and their standard deviations (in parenthesis) for the first and second applications, respectively. In the first application, all methods agree that the coefficient of the inclination is not significant and, for both applications, the estimation methods are very close to each other except for the *NA* approach which produces standard errors much lower than the other approaches. Figure 3 shows the fitted lines. We can see a very significant difference between the naive approach and the others. Figure 4 presents the ellipses from these approaches, the *NA* produces the smallest confidence region. It can be explained by the simulation results on Table 1.a, the test sizes are greater than the expected nominal level, which means underestimated standard errors.

6. Conclusions and final remarks

We have presented the asymptotic covariance matrix for the line (under the maximum likelihood and method of moments approaches) estimators in a heteroscedastic structural errors-in-variables model (this model is largely applied in the astrophysics field) which leads to more accurate confidence regions and the hypotheses testing is more trustworthy. Furthermore, all methods are robust against the distribution of the unobservable covariate (although the maximum likelihood approach depends on that distribution, the simulation studies indicate that the tests regarding the line regression parameters seem to be not affected by the distribution of x). The simulation study in Section 4 can be used as guidance to the practitioner having to select a statistical test. Moreover, it was shown that the naive approach (that does not consider errors in the covariate) may produce results much different from the expected.

7. Acknowledgements

This work was partially supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil). The authors thank Dr. Kari Kuulasmaa (National Public Health Institute, Finland) and Dr. Brandon C. Kelly for kindly supplying the data of our epidemiological and astrophysical applications, respectively. The authors also acknowledge helpful comments and suggestions from three anonymous referees and an associated editor which led to an improved presentation.

References

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

Cheng CL, Van Ness JW. (1999). *Statistical Regression with Measurement Error.* Arnold Publishers, London.

Cheng CL, Riu J. (2006). On estimating linear relationships when both variables are subject to heteroscedastic measurement errors. *Technometrics.* **48**:511–519.

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

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

Huber, PJ (1964). Robust estimation of a location parameter. *Annals of Mathematics Statistics.* **35**:73–101.

Kelly BC. (2007). Some aspects of measurement error in linear regression of astronomical data. *The Astrophysical Journal.* **665**:1489–1506.

Kelly BC, Bechtold J, Trump JR, 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**(2): 355–373.

Kulathinal SB, 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.

Lehmann, EL, Casella, G. (1998). *Theory of Point Estimation*, 2nd ed. Springer-Verlag: New York.

Table 1: Test sizes (%) for the hypotheses (a) $H_0 : (\beta_0, \beta_1) = (0, 1)$ and (b) $H_0 : (\beta_0, \beta_1) = (0, 0)$ using a naive procedure, i.e., the Wald statistic (2.6) taking $\tau_{x*} = \tau_{x**} = \tau_{xy*} = \tau_x = 0$. The expected behavior for all cells is to converge to 5% when the sample size n increases.

		Distribution of x		
		Normal	Half normal	Student t
(a)	$n = 40$	12.34	18.06	10.07
	$n = 80$	17.06	25.92	10.99
	$n = 160$	24.38	44.31	18.62
(b)	$n = 40$	7.00	6.88	7.10
	$n = 80$	6.03	5.85	5.85
	$n = 160$	5.30	5.19	5.47

Table 2: Rejection rates (%) for the hypothesis $H_0 : (\beta_0, \beta_1) = (0, 1)$ (at a 5% nominal level) using the Wald statistics (2.6), (2.7) and (3.2) for $n = 40$, $n = 80$, $n = 160$ and $x \overset{iid}{\sim} \mathcal{N}(-2, 4)$. It is expected to be 5% in the middle cells.

	MM 1			*MM 2*			*ML*		
β_0	β_1			β_1			β_1		
$n = 40$	0.6	1	1.4	0.6	1	1.4	0.6	1	1.4
-2	51.02	79.58	98.00	54.75	81.62	98.09	55.11	83.26	98.73
0	34.54	6.77	29.11	38.20	9.39	33.69	36.83	7.93	33.84
2	98.46	80.16	45.84	98.69	82.57	50.32	98.92	84.10	51.20
$n = 80$	0.6	1	1.4	0.6	1	1.4	0.6	1	1.4
-2	78.22	97.75	99.99	79.10	97.82	100	81.98	98.55	100
0	55.86	5.73	50.08	57.41	6.93	52.89	59.30	6.22	55.38
2	99.99	98.00	74.76	99.99	98.17	76.21	100	98.80	78.59
$n = 160$	0.6	1	1.4	0.6	1	1.4	0.6	1	1.4
-2	97.22	99.98	100	97.27	99.99	100	98.33	100	100
0	83.83	5.09	80.19	84.03	5.57	81.06	87.00	5.46	84.57
2	100	99.97	96.20	100	99.98	96.28	100	99.97	97.41

Table 3: Rejection rates (%) for the hypothesis $H_0 : (\beta_0, \beta_1) = (0, 1)$ (at a 5% nominal level) using the Wald statistics (2.6), (2.7) and (3.2) for $n = 40$, $n = 80$, $n = 160$ and $x \overset{iid}{\sim} \mathcal{HN}(-2, 4)$. It is expected to be 5% in the middle cells.

	MM 1			*MM 2*			*ML*		
β_0	β_1			β_1			β_1		
$n = 40$	0.6	1	1.4	0.6	1	1.4	0.6	1	1.4
-2	74.44	78.62	83.61	77.09	80.88	85.57	78.72	82.94	88.74
0	9.94	4.69	4.78	13.12	6.93	7.80	11.28	5.97	7.92
2	86.68	77.81	70.02	88.19	80.28	73.06	90.04	82.61	76.29
$n = 80$	0.6	1	1.4	0.6	1	1.4	0.6	1	1.4
-2	97.11	9.801	99.22	97.29	98.20	99.28	98.00	98.75	99.48
0	14.24	4.83	7.90	15.58	6.17	11.03	15.02	5.64	12.35
2	99.54	98.17	96.09	99.60	98.31	96.44	99.81	98.85	97.55
$n = 160$	0.6	1	1.4	0.6	1	1.4	0.6	1	1.4
-2	99.92	99.99	100	99.92	99.99	100	99.97	99.99	100
0	19.89	4.89	13.21	20.63	5.43	15.52	21.49	5.44	17.81
2	100	99.97	99.95	100	99.97	99.95	100	99.99	99.98

Table 4: Rejection rates (%) for the hypothesis $H_0 : (\beta_0, \beta_1) = (0, 1)$ (at a 5% nominal level) using the Wald statistics (2.6), (2.7) and (3.2) for $n = 40$, $n = 80$, $n = 160$ and $x \overset{iid}{\sim} t(-2, 4, 5)$. It is expected to be 5% in the middle cells.

	MM 1			*MM 2*			*ML*		
β_0	β_1			β_1			β_1		
$n = 40$	0.6	1	1.4	0.6	1	1.4	0.6	1	1.4
-2	56.60	78.70	97.79	60.87	81.08	98.28	60.81	82.75	98.77
0	40.85	6.95	36.26	45.99	10.24	43.03	44.17	7.76	41.56
2	98.71	77.86	51.65	98.79	80.71	57.02	99.17	81.98	57.38
$n = 80$	0.6	1	1.4	0.6	1	1.4	0.6	1	1.4
-2	86.29	97.94	100	87.00	98.24	100	88.37	98.55	100
0	68.98	5.70	64.48	70.73	7.32	67.29	72.36	6.06	69.16
2	99.99	98.15	83.61	99.99	98.31	84.73	100	98.61	86.25
$n = 160$	0.6	1	1.4	0.6	1	1.4	0.6	1	1.4
-2	99.01	100	100	99.02	99.99	100	99.31	99.99	100
0	92.92	5.31	91.62	93.02	6.17	91.85	94.47	5.06	93.97
2	100	99.98	98.49	100	99.98	98.51	100	99.99	99.07

Table 5: Rejection rates (%) of the hypothesis $H_0 : \beta_1 = 0$ (at a 5% nominal level) for $n = 40$, $n = 80$, $n = 160$ and $x \overset{iid}{\sim} \mathcal{N}(-2, 4)$: (a) $\sqrt{\tau_x}$ and $\sqrt{\tau_y}$ uniform and (b) $\sqrt{\tau_x} = 0.1|x|$, $\sqrt{\tau_y} = 0.1|\beta_0 + \beta_1 x|$.

		MM 1		*MM 2*		*ML*		*NA*	
	β_1	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
$n = 40$									
	-0.5	46.70	51.55	48.48	53.90	49.57	51.62	46.83	52.13
	-0.25	18.03	18.23	19.63	20.94	19.03	18.29	18.10	18.47
	0.0	6.75	6.35	7.87	8.46	7.03	6.63	6.80	6.58
	0.25	17.54	18.75	18.73	21.40	18.81	18.96	17.61	19.14
	0.5	46.57	50.28	48.09	53.13	49.67	50.75	46.79	50.89
$n = 80$									
	-0.5	64.43	78.21	65.09	78.40	68.07	77.37	64.64	78.43
	-0.25	22.48	28.57	23.60	30.71	24.43	28.28	22.63	28.77
	0.0	5.54	5.65	6.33	6.86	5.78	5.60	5.61	5.78
	0.25	22.53	28.80	23.59	30.55	24.30	28.27	22.68	29.02
	0.5	64.93	78.01	65.98	78.20	68.97	77.37	65.18	78.25
$n = 160$									
	-0.5	80.01	95.17	80.35	94.89	82.56	94.45	80.01	95.21
	-0.25	29.18	44.73	30.22	45.59	31.22	43.42	29.19	44.95
	0.0	5.34	5.26	5.83	6.00	5.65	5.20	5.35	5.32
	0.25	29.29	45.14	30.06	46.08	31.39	43.99	29.30	45.31
	0.5	79.23	94.65	79.67	94.43	82.13	94.07	79.25	94.71

Table 6: Rejection rates (%) of the hypothesis $H_0 : \beta_1 = 0$ (at a 5% nominal level) for $n = 40$, $n = 80$, $n = 160$ and $x \overset{iid}{\sim} \mathcal{HN}(-2, 4)$: (a) $\sqrt{\tau_x}$ and $\sqrt{\tau_y}$ uniform and (b) $\sqrt{\tau_x} = 0.1|x|$, $\sqrt{\tau_y} = 0.1|\beta_0 + \beta_1 x|$.

		MM 1		*MM 2*		*ML*		*NA*	
	β_1	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
$n = 40$									
	-0.5	8.47	31.47	11.00	39.31	9.61	28.32	11.49	31.73
	-0.25	5.21	12.85	7.19	19.44	6.07	12.36	7.29	13.01
	0.0	4.33	6.23	6.53	11.46	5.25	6.35	6.41	6.35
	0.25	5.32	12.52	7.47	19.13	6.41	11.99	7.58	12.63
	0.5	8.96	31.49	11.81	39.27	9.46	28.37	11.88	31.79
$n = 80$									
	-0.5	14.31	44.03	17.11	45.75	18.53	42.68	17.10	44.13
	-0.25	6.28	15.51	8.22	17.43	8.32	15.38	7.85	15.57
	0.0	4.17	5.47	5.28	6.91	5.30	5.73	5.15	5.49
	0.25	6.71	15.47	8.38	17.11	8.54	15.31	8.32	15.51
	0.5	13.55	43.39	16.39	45.19	17.85	42.29	16.03	43.45
$n = 160$									
	-0.5	22.41	64.93	23.74	65.42	26.37	63.59	24.21	64.99
	-0.25	8.63	21.35	9.69	22.89	10.23	20.97	9.77	21.37
	0.0	4.25	5.37	5.15	6.04	4.77	5.49	4.89	5.39
	0.25	8.87	22.84	9.81	23.76	10.11	22.03	9.91	22.87
	0.5	22.46	64.89	23.96	65.71	26.39	63.69	24.46	64.91

Table 7: Rejection rates (%) of the hypothesis $H_0 : \beta_1 = 0$ (at a 5% nominal level) for $n = 40$, $n = 80$, $n = 160$ and $x \overset{iid}{\sim} t(-2, 4, 5)$: (a) $\sqrt{\tau_x}$ and $\sqrt{\tau_y}$ uniform and (b) $\sqrt{\tau_x} = 0.1|x|$, $\sqrt{\tau_y} = 0.1|\beta_0 + \beta_1 x|$.

		MM 1		*MM 2*		*ML*		*NA*	
	β_1	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
$n = 40$									
	-0.5	41.30	87.87	45.49	89.09	45.14	81.58	41.69	88.11
	-0.25	15.00	36.93	18.52	45.35	16.53	32.15	15.27	37.46
	0.0	5.85	6.52	8.43	12.33	6.62	6.55	6.01	6.68
	0.25	14.81	36.87	18.68	45.67	16.79	32.17	15.03	37.28
	0.5	41.91	88.00	46.78	89.15	46.07	81.62	42.28	88.27
$n = 80$									
	-0.5	65.98	90.33	68.90	90.16	70.88	88.47	66.39	90.43
	-0.25	22.21	37.72	25.33	40.46	24.43	35.87	22.55	37.95
	0.0	4.97	5.32	6.76	7.01	5.12	5.31	5.12	5.37
	0.25	22.05	37.91	25.19	40.77	24.49	36.25	22.33	38.11
	0.5	66.14	90.01	68.79	89.61	71.01	88.47	66.51	90.18
$n = 160$									
	-0.5	97.70	98.32	97.69	98.15	98.63	97.19	97.73	98.34
	-0.25	49.59	55.67	52.83	56.65	54.62	50.23	49.71	55.90
	0.0	4.59	5.69	5.65	6.75	4.54	5.55	4.41	5.75
	0.25	50.47	55.64	53.45	57.03	55.44	51.19	50.53	55.82
	0.5	97.59	98.36	97.91	98.14	98.68	97.19	97.61	98.41

Table 8: Estimates of regression parameters (standard errors in parentheses) for the WHO MONICA Project data sets.

		β_0	β_1	μ_x	σ_x^2	σ^2
Men	*MM 1*	-1.84 (0.50)	0.35 (0.22)	-1.08	4.50	3.06
	MM 2	-1.84 (0.44)	0.35 (0.22)	-1.08	4.37	2.87
	ML	-2.08 (0.53)	0.47 (0.23)	-1.09	4.32	4.89
	NA	-1.88 (0.49)	0.31 (0.20)	-1.08	5.00	3.12
Women	*MM 1*	-0.33 (1.04)	0.58 (0.38)	-2.02	3.93	6.43
	MM 2	-0.33 (0.90)	0.58 (0.33)	-2.02	3.80	5.94
	ML	0.03 (1.11)	0.68 (0.41)	-2.07	3.62	11.08
	NA	-0.47 (0.97)	0.51 (0.33)	-2.02	4.47	6.59

Table 9: Estimates of regression parameters (standard errors in parentheses) for the first application of the *Chandra* observatory.

	β_0	β_1	μ_x	σ_x^2	σ^2
MM 1	1.46 (0.41)	−0.50 (0.32)	−1.27	0.07	0.13
MM 2	1.46 (0.42)	−0.51 (0.34)	−1.27	0.07	0.13
ML	1.73 (0.37)	−0.23 (0.28)	−1.30	0.06	0.09
NA	1.86 (0.15)	−0.19 (0.11)	−1.27	0.19	0.14

Table 10: Estimates of regression parameters (standard errors in parentheses) for second application of the *Chandra* observatory.

	β_0	β_1	μ_x	σ_x^2	σ^2
MM 1	2.22 (0.20)	0.56 (0.16)	−1.27	0.07	0.002
MM 2	2.22 (0.22)	0.57 (0.17)	−1.27	0.07	0.002
ML	2.27 (0.22)	0.59 (0.17)	−1.29	0.07	0.002
NA	1.78 (0.04)	0.21 (0.03)	−1.27	0.19	0.020

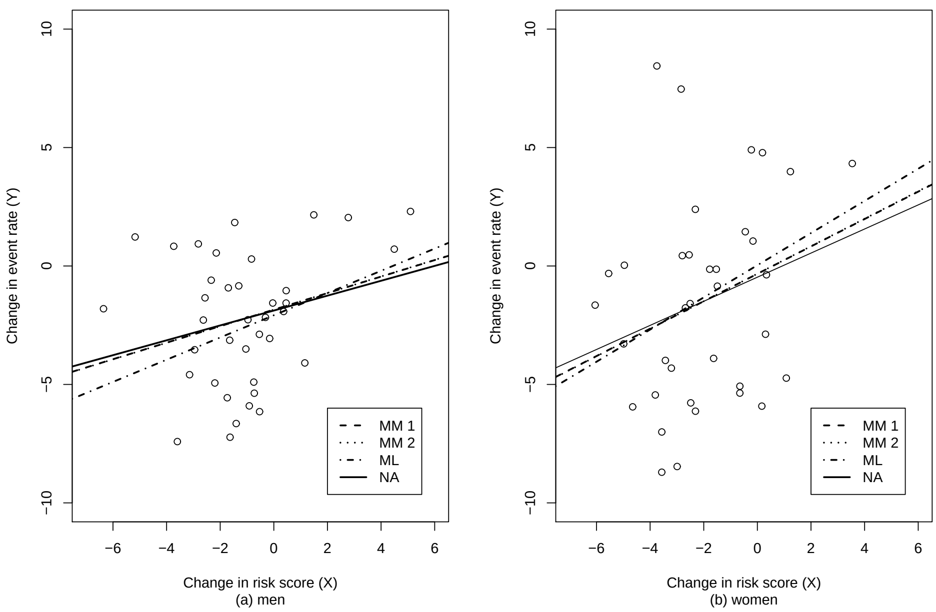

Figure 1: Change in event rate *versus* change in risk score and regression lines for the WHO MONICA Project data sets: (a) men and (b) women.

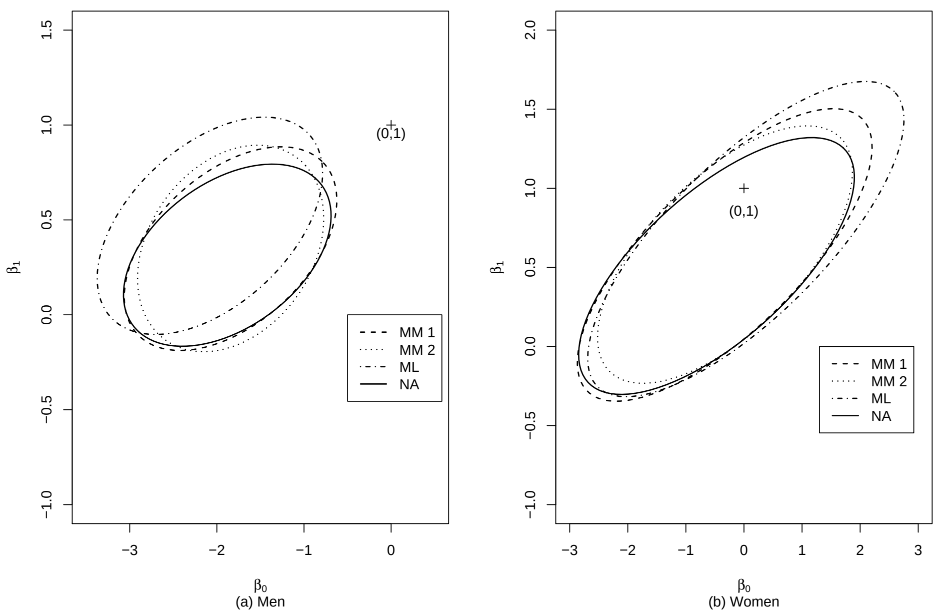

Figure 2: 95% confidence regions for (β_0, β_1) using the WHO MONICA Project data sets.

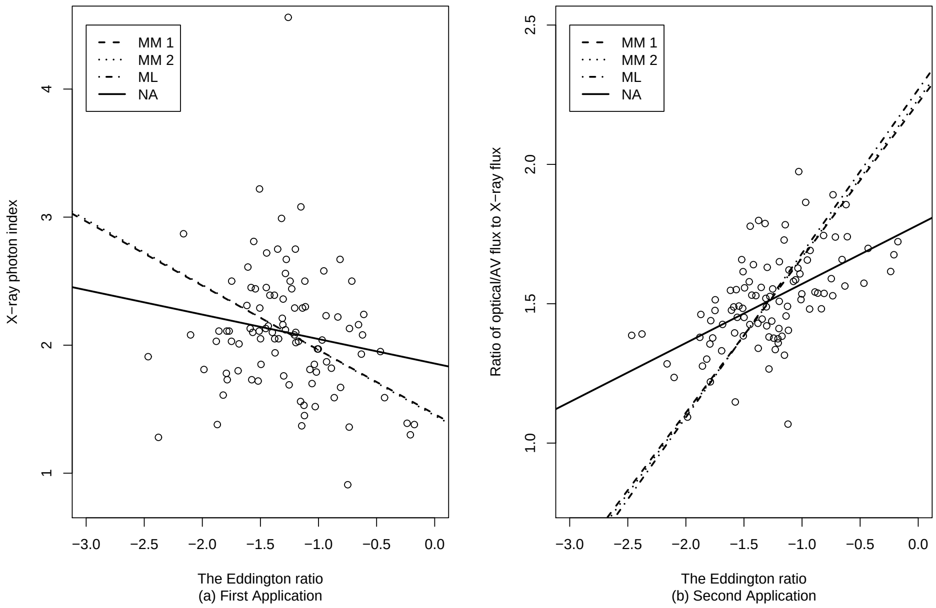

Figure 3: Scatter plots for the *Chandra* data: (a) first application and (b) second application

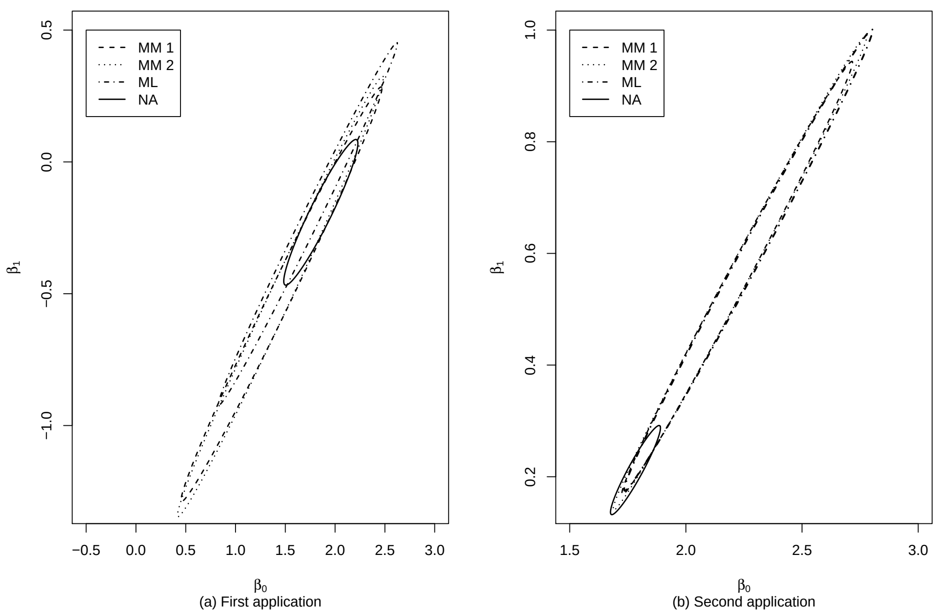

Figure 4: 95% confidence regions for (β_0, β_1) using the *Chandra* data sets.