AStA Adv Stat Anal
DOI 10.1007/s10182-010-0123-4

ORIGINAL PAPER

A modified signed likelihood ratio test in elliptical structural models

Tatiane F.N. Melo · Silvia L.P. Ferrari

Received: 12 March 2009 / Accepted: 13 February 2010

Abstract In this paper we deal with the issue of performing accurate testing inference on a scalar parameter of interest in structural errors-in-variables models. The error terms are allowed to follow a multivariate distribution in the class of the elliptical distributions, which has the multivariate normal distribution as special case. We derive a modified signed likelihood ratio statistic that follows a standard normal distribution with a high degree of accuracy. Our Monte Carlo results show that the modified test is much less size distorted than its unmodified counterpart. An application is presented.

Keywords Elliptical distribution · Errors-in-variables model · Measurement error · Modified signed likelihood ratio statistic · Structural model

1 Introduction

Regression models are a powerful tool for exploring the dependence of a response on a set of explanatory variables. It is often assumed that the explanatory variables are measured without error. In many practical situations, however, the explanatory variables are not measured exactly. Regression models that account for such measurement errors are often named errors-in-variables models or measurement error

T.F.N. Melo (✉)
Instituto de Matemática e Estatística, Universidade Federal de Goiás, Campus Samambaia, CP 131, Goiânia, GO 74001-970, Brazil
e-mail: tmelo@mat.ufg.br

S.L.P. Ferrari
Departamento de Estatística, Universidade de São Paulo, Rua do Matão, 1010, São Paulo, SP 05508-090, Brazil
e-mail: sferrari@ime.usp.br

Published online: 04 March 2010

models. Structural models assume that the covariates are random variables. In functional models, on the other hand, the explanatory variables are treated as unknown parameters. The structural model formulation poses identifiability problems, while functional models lead to unlimited likelihood function. To circumvent such problems, a number of different restrictions can be imposed, such as that some variances are known (e.g., Chan and Mak 1979, and Wong 1989), or that the intercept is known (Aoki et al. 2001). In the present paper we deal with structural models. For details on errors-in-variables models, see Fuller (1987) and Cheng and Van Ness (1999).

Normality is a common assumption for continuous response data. It is well known, however, that likelihood inference in normal regression models is sensitive to extreme observations. When the dataset contains outlying observations, it is advisable to assume that the data were drawn from a distribution with heavier-than-normal tails, such as the Student-t distribution. The family of the elliptical distributions provides a useful alternative to the normal distribution since it includes the normal distribution, heavy-tailed distributions such as the exponential power, Student-t, Pearson II, Pearson VII, logistic II, and light-tailed distributions, for example, logistic I. Further information on elliptical distributions can be found in Fang et al. (1990) and Fang and Anderson (1990).

Likelihood inference in errors-in-variables models is usually based on first-order asymptotic theory, which can lead to inaccurate inference when the sample is small or of moderate size. This is the case of the signed likelihood ratio test, often used to perform testing inference on a scalar parameter of interest in the presence of nuisance parameters. Asymptotically, its statistic has a standard normal distribution under the null hypothesis, with error of order $n^{-1/2}$, where n is the sample size. Aiming to improve this approximation, Barndorff-Nielsen (1986) proposed an adjustment to the test statistic, in such a way that the modified signed likelihood ratio statistic has, under the null hypothesis, a standard normal distribution with error of order $n^{-3/2}$. In order to obtain such an adjustment, it is necessary to identify a suitable ancillary statistic. It is required that the maximum likelihood estimator coupled with the ancillary statistic constitutes a sufficient statistic for the model. In many situations it is very difficult or even impossible to find an appropriate ancillary, which makes the approach unfeasible. In this paper, we give the ancillary statistic and obtain a modified signed likelihood ratio statistic in elliptical structural models.

The paper unfolds as follows. Section 2 introduces the elliptical structural models. Section 3 contains our main results, namely the ancillary statistic and an explicit formula for the modified signed likelihood ratio statistic. Section 4 presents a simulation study on the finite sample behavior of the standard signed likelihood ratio test and its modified counterpart. Our simulation results show that the signed likelihood ratio test tends to be liberal and its modified version is much less size-distorted. An application that uses real data is presented and discussed in Sect. 5. Finally, Sect. 6 concludes the paper. Technical details are collected in two appendices.

2 Elliptical structural models

A $p \times 1$ random vector $\boldsymbol{Z}$ is said to have a p-variate elliptical distribution with location vector $\boldsymbol{\mu}$ $(p \times 1)$, dispersion matrix Σ $(p \times p)$, and density generating function

p_0, and we write $\boldsymbol{Z} \sim El_p(\boldsymbol{\mu}, \Sigma; p_0)$, if

$$\boldsymbol{Z} \stackrel{d}{=} \boldsymbol{\mu} + A\boldsymbol{Z}^*,$$

where A is a $p \times k$ matrix with $\text{rank}(A) = k$, $AA^\top = \Sigma$, and $\boldsymbol{Z}^*$ is a $p \times 1$ random vector with density function $p_0(z^\top z)$ for $z \in \Re^p$. The notation $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{Y}$ indicates that $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same distribution. It is assumed that $\int_0^\infty y^{p/2-1} p_0(y)\,dy < \infty$. The density function of $\boldsymbol{Z}$ is

$$p(z, \mu, \Sigma) = |\Sigma|^{-1/2} p_0\big((z - \boldsymbol{\mu})^\top \Sigma^{-1}(z - \boldsymbol{\mu})\big). \tag{1}$$

Some special cases of (1) are the following multivariate distributions: normal, exponential power, Pearson II, Pearson VII, Student-t, generalized Student-t, logistic I, logistic II, and Cauchy. The elliptical distributions share many properties with the multivariate normal distribution. In particular, marginal distributions are elliptical. For a full account of the properties of the elliptical distributions, see Fang et al. (1990, Sect. 2.5).

We consider the simple linear measurement error regression model

$$Y_i = \alpha + \beta x_i + e_i, \tag{2}$$

$$X_i = x_i + u_i, \tag{3}$$

for $i = 1, \dots, n$. Here, $\alpha \in \Re$ and $\beta \in \Re$ are unknown parameters, and e_i is an unobservable random error associated with the response Y_i. Also, the explanatory variables $x_1, \dots, x_n$ are random variables and are not observed directly but, instead, are observed with measurement errors, $u_1, \dots, u_n$ respectively. Equations (2)–(3) can be written as

$$\boldsymbol{Z}_i = \boldsymbol{\delta} + \Delta \boldsymbol{b}_i, \quad i = 1, 2, \dots, n, \tag{4}$$

where

$$\boldsymbol{Z}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}, \qquad \boldsymbol{\delta} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \qquad \Delta = \begin{pmatrix} \beta & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{b}_i = \begin{pmatrix} x_i \\ e_i \\ u_i \end{pmatrix}.$$

We assume that $\boldsymbol{b}_1, \dots, \boldsymbol{b}_n$ are independent and $\boldsymbol{b}_i \sim El_3(\boldsymbol{\eta}, \Omega; p_0)$, with

$$\boldsymbol{\eta} = \begin{pmatrix} \mu_x \\ 0 \\ 0 \end{pmatrix}, \qquad \Omega = \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_e^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{pmatrix},$$

$\mu_x \in \Re$, $\sigma_x^2 > 0$, $\sigma_e^2 > 0$, and $\sigma_u^2 > 0$. Therefore, $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n$ are independent random variables, and $\boldsymbol{Z}_i \sim El_2(\boldsymbol{\mu}, \Sigma; p_0)$ with $\boldsymbol{\mu} = \boldsymbol{\delta} + \Delta\boldsymbol{\eta}$ and $\Sigma = \Delta\Omega\Delta^\top$ (Fang et al. 1990, Sect. 2.5). We assume that Σ is a positive definite matrix. We can write $\boldsymbol{\mu}$ and Σ as

$$\boldsymbol{\mu} = \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \beta^2\sigma_x^2 + \sigma_e^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix}.$$

The structural regression model (4) is not identifiable; see Fuller (1987, Sect. 1.1.3). To circumvent this problem, it is often assumed one of the following three conditions: (i) $\lambda_e = \sigma_e^2/\sigma_u^2$ is known; (ii) $\lambda_x = \sigma_x^2/\sigma_u^2$ is known; (iii) the intercept (α) is known. Then, the vector of unknown parameters, $\boldsymbol{\theta}$ say, is

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)^\top = \begin{cases} (\beta, \alpha, \mu_x, \sigma_x^2, \sigma_u^2)^\top & \text{if } \lambda_e \text{ is known,} \\ (\beta, \alpha, \mu_x, \sigma_u^2, \sigma_e^2)^\top & \text{if } \lambda_x \text{ is known,} \\ (\beta, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_e^2)^\top & \text{if } \alpha \text{ is known.} \end{cases}$$

Under any of these conditions, the log-likelihood function is

$$\ell(\boldsymbol{\theta}) = -\frac{n}{2} \log |\Sigma| + \sum_{i=1}^{n} \log p_0\big(\boldsymbol{d}_i^\top \Sigma^{-1} \boldsymbol{d}_i\big), \tag{5}$$

where $\boldsymbol{d}_i = \boldsymbol{d}_i(\boldsymbol{\theta}) = z_i - \boldsymbol{\mu}$.

When $\boldsymbol{Z}$ has a normal distribution and λ_x or λ_e is known, the maximum likelihood estimator of $\boldsymbol{\theta}$ has closed form; see Wong (1991), Arellano-Valle and Bolfarine (1996), and Fuller (1987, Sect. 1.3.2). For distributions other than the normal or in the case where the intercept α is known, the maximum likelihood estimator must be numerically obtained by maximizing the log-likelihood function (5) through an iterative algorithm, such as the Newton–Raphson, Fisher scoring, EM, or BFGS. In this paper, we use the BFGS method, which is implemented in `Ox` as the `MaxBFGS` function (Doornik 2006).

3 Modified signed likelihood ratio test

Let the parameter vector $\boldsymbol{\theta}$ be partitioned as $\boldsymbol{\theta} = (\psi, \boldsymbol{\omega}^\top)^\top$, where ψ is the parameter of interest (scalar), and $\boldsymbol{\omega}$ is the vector of nuisance parameters. We focus on tests of the null hypotheses $\mathcal{H}_{0a}: \psi \geq \psi^{(0)}$, $\mathcal{H}_{0b}: \psi = \psi^{(0)}$, and $\mathcal{H}_{0c}: \psi \leq \psi^{(0)}$, where $\psi^{(0)}$ is a constant. Here, $\psi = \theta_1 = \beta$ and $\boldsymbol{\omega} = (\theta_2, \theta_3, \theta_4, \theta_5)^\top$. The unrestricted maximum likelihood estimator of the $\boldsymbol{\theta}$ is denoted by $\widehat{\boldsymbol{\theta}} = (\widehat{\psi}, \widehat{\boldsymbol{\omega}}^\top)^\top$, and the corresponding estimator assuming that $\psi = \psi^{(0)}$, by $\widetilde{\boldsymbol{\theta}} = (\psi^{(0)}, \widetilde{\boldsymbol{\omega}}^\top)^\top$. We use hat and tilde to indicate evaluation at $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$, respectively.

Let r_S be the signed likelihood ratio statistic,

$$r_S = \operatorname{sgn}\big(\widehat{\psi} - \psi^{(0)}\big)\sqrt{2\big(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widetilde{\boldsymbol{\theta}})\big)},$$

which can be used to construct a confidence interval for ψ or to test either of the hypotheses mentioned above. In any case, one must compute its exact or approximate distribution. It is well known that, if $\psi = \psi^{(0)}$, r_S has a standard normal distribution with error of order $n^{-1/2}$. If the sample size is large, the standard normal distribution provides a good approximation. On the other hand, if the sample size is not large enough to guarantee a good agreement between the true distribution of r_S and its limiting distribution, inference can be misleading.

Barndorff-Nielsen (1986, 1991) proposed a modified version of the signed likelihood ratio statistic that has a standard normal distribution with a higher order of approximation, $n^{-3/2}$ instead of $n^{-1/2}$. The modified statistic, denoted here by r^*, retains the essential character of r_S but can be difficult to obtain. The difficulty arises from the need of an appropriate ancillary statistic and of derivatives of the log-likelihood function with respect to the data. By "ancillary statistic" we mean a statistic, $\boldsymbol{a}$ say, whose distribution does not depend on the unknown parameter $\boldsymbol{\theta}$ and, together with the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$, is a minimal sufficient statistic for the model. If $(\widehat{\boldsymbol{\theta}}, \boldsymbol{a})$ is sufficient, but not minimal sufficient, Barndorff-Nielsen's results still hold; see Severini (2000, Sect. 6.5). In fact, minimal sufficiency is only required for the ancillary $\boldsymbol{a}$ to be relevant to the statistical analysis. Sufficiency implies that the log-likelihood function depends on the data only through $(\widehat{\boldsymbol{\theta}}, \boldsymbol{a})$, and we then write $\ell(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}, \boldsymbol{a})$. The required derivatives of $\ell(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}, \boldsymbol{a})$ with respect to the data are

$$\ell' = \frac{\partial \ell(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}, a)}{\partial \widehat{\boldsymbol{\theta}}}, \qquad U' = \frac{\partial^2 \ell(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}, \boldsymbol{a})}{\partial \widehat{\boldsymbol{\theta}} \partial \boldsymbol{\theta}^\top}.$$

The modified signed likelihood ratio statistic is

$$r^* = r_S - \frac{1}{r_S} \log \rho$$

with

$$\rho = \left|\widehat{J}\right|^{1/2} \left|\widetilde{U}'\right|^{-1} \left|\widetilde{J}_{\boldsymbol{\omega}\boldsymbol{\omega}}\right|^{1/2} \frac{r_S}{[(\widehat{\ell}' - \widetilde{\ell}')^\top (\widetilde{U}')^{-1}]_\psi}, \tag{6}$$

where J is the observed information matrix, and $J_{\boldsymbol{\omega}\boldsymbol{\omega}}$ is the lower right submatrix of J corresponding to the nuisance parameter $\boldsymbol{\omega}$. Here, $[\boldsymbol{v}]_\psi$ denotes the element of the vector $\boldsymbol{v}$ that corresponds to the parameter of interest ψ.

In order to find an ancillary statistic, we first note that model (4) is a transformation model. Hence any maximal invariant statistic is an ancillary statistic; see Barndorff-Nielsen et al. (1989, Chap. 8) and Barndorff-Nielsen (1986). A statistic $\boldsymbol{a}(\cdot)$ defined on the sample space $\boldsymbol{\mathcal{Z}}$ is invariant under a group of transformations $\mathcal{G}$ if $\boldsymbol{a}(g(z)) = \boldsymbol{a}(z)$ for all $g \in \mathcal{G}$. Moreover, $\boldsymbol{a}(\cdot)$ is a maximal invariant statistic if it is a function of any invariant statistic, i.e., $\boldsymbol{a}(z) = \boldsymbol{a}(z')$ implies that there exists $g \in \mathcal{G}$ such that $z' = g(z)$. Pace and Salvan (1997, Theorem 7.2) show that, in transformation models, all invariant statistics are distribution constant, i.e., their distributions do not depend on $\boldsymbol{\theta}$.

Consider the log-likelihood function (5) and assume that the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, based on $z = (z_1^\top, z_2^\top, \ldots, z_n^\top)^\top$, exists and is finite. Let $\mathcal{G}^n$ be the transformation group with elements $g^n(z) = (g(z_1)^\top, g(z_2)^\top, \ldots, g(z_n)^\top)^\top$, where $g(z_i) = \boldsymbol{\xi} + \Psi z_i$, $\boldsymbol{\xi} \in \Re^2$, and Ψ is a (positive definite) lower triangular matrix. Using $g^n(z)$, the induced transformation in the parameter space is $\bar{g}(\boldsymbol{\mu}, P) = (\boldsymbol{\xi} + \Psi \boldsymbol{\mu}, \Psi P)$, where P is a lower triangular matrix such that $P P^\top = \Sigma$. Let $(\widehat{\boldsymbol{\mu}}(z), \widehat{P}(z))$ be the maximum likelihood estimate of (μ, P) based on z. The corresponding estimate based on $g^n(z)$ is $(\boldsymbol{\xi} + \Psi \widehat{\boldsymbol{\mu}}(z), \Psi \widehat{P}(z))$ because of the equivariance of the maximum likelihood estimate, see Pace and Salvan (1997, Sect. 7.6). We now argue that $\boldsymbol{a} = \boldsymbol{a}(z) = (\boldsymbol{a}_1^\top(z), \boldsymbol{a}_2^\top(z), \ldots, \boldsymbol{a}_n^\top(z))^\top$, with

$$\boldsymbol{a}_i(z) = \widehat{P}(z)^{-1}\big(z_i - \widehat{\boldsymbol{\mu}}(z)\big), \tag{7}$$

is maximal invariant with respect to $\mathcal{G}^n$. First, we show that $\boldsymbol{a}$ is invariant with respect to the group $\mathcal{G}^n$, i.e., $\boldsymbol{a}(g^n(z)) = \boldsymbol{a}(z)$. Indeed,

$$\begin{aligned}
\boldsymbol{a}_i\big(g^n(z)\big) &= \widehat{P}\big(g^n(z)\big)^{-1}\big[g(z_i) - \widehat{\boldsymbol{\mu}}\big(g^n(z)\big)\big] \\
&= \big[\Psi \widehat{P}(z)\big]^{-1}\big[\boldsymbol{\xi} + \Psi z_i - \big(\boldsymbol{\xi} + \Psi\widehat{\boldsymbol{\mu}}(z)\big)\big] \\
&= \widehat{P}(z)^{-1}\big(z_i - \widehat{\boldsymbol{\mu}}(z)\big) = \boldsymbol{a}_i(z).
\end{aligned}$$

Now, assume that $\boldsymbol{a}(z) = \boldsymbol{a}(z')$. We have

$$\widehat{P}(z)^{-1}\big(z_i - \widehat{\boldsymbol{\mu}}(z)\big) = \widehat{P}\big(z'\big)^{-1}\big(z_i' - \widehat{\boldsymbol{\mu}}\big(z'\big)\big)$$

or, equivalently,

$$z_i' = \widehat{\boldsymbol{\mu}}\big(z'\big) - \widehat{P}\big(z'\big)\widehat{P}(z)^{-1}\widehat{\boldsymbol{\mu}}(z) + \widehat{P}\big(z'\big)\widehat{P}(z)^{-1}z_i = \boldsymbol{\xi} + \Psi z_i$$

with $\boldsymbol{\xi} = \widehat{\boldsymbol{\mu}}(z') - \widehat{P}(z')\widehat{P}(z)^{-1}\widehat{\boldsymbol{\mu}}(z)$ and $\Psi = \widehat{P}(z')\widehat{P}(z)^{-1}$. Therefore, $\boldsymbol{a}$ is maximal invariant with respect to the group $\mathcal{G}^n$.

Using the maximal invariant statistic $\boldsymbol{a}$, we can find the quantities needed for the Barndorff-Nielsen adjustment in the elliptical structural model (4). Let T be a 5×5 matrix, whose (j,k)th element is $t_{jk} = \operatorname{tr}(\Sigma^{\theta_k}\Sigma_{\theta_j}) + \operatorname{tr}(\Sigma^{-1}\Sigma_{\theta_j\theta_k})$, where $\Sigma_{\theta_j} = \partial\Sigma/\partial\theta_j$, $\Sigma_{\theta_j\theta_k} = \partial\Sigma_{\theta_j}/\partial\theta_k$, and $\Sigma^{\theta_j} = \partial\Sigma^{-1}/\partial\theta_j = -\Sigma^{-1}\Sigma_{\theta_j}\Sigma^{-1}$ for $j,k = 1,2,3,4,5$. We define the $5n \times 5$ block diagonal matrices $R = \operatorname{diag}(\boldsymbol{r},\boldsymbol{r},\boldsymbol{r},\boldsymbol{r},\boldsymbol{r})$ and $V = \operatorname{diag}(\boldsymbol{v},\boldsymbol{v},\boldsymbol{v},\boldsymbol{v},\boldsymbol{v})$, where the ith elements of the vectors $\boldsymbol{r}$ and $\boldsymbol{v}$ are $r_i = W_{p_0}(\boldsymbol{d}_i^\top\Sigma^{-1}\boldsymbol{d}_i)$ and $v_i = W_{p_0}'(\boldsymbol{d}_i^\top\Sigma^{-1}\boldsymbol{d}_i)$, respectively, and $W_{p_0}(u) = \partial\log p_0(u)/\partial u$. Let $\boldsymbol{w} = (\boldsymbol{w}^{(1)\top}, \boldsymbol{w}^{(2)\top}, \boldsymbol{w}^{(3)\top}, \boldsymbol{w}^{(4)\top}, \boldsymbol{w}^{(5)\top})$ be a column vector of dimension $5n$, where the ith element of $\boldsymbol{w}^{(j)}$ is $w_i^{(j)} = (\widehat{P}_{\theta_j}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}}_{\theta_j})^\top\Sigma^{-1}(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})$, where $P_{\theta_j} = \partial P/\partial\theta_j$ and $\boldsymbol{\mu}_{\theta_j} = \partial\boldsymbol{\mu}/\partial\theta_j$. Additionally, let B, C, M, and Q be the $5n \times 5$ block matrices whose (j,k)th blocks are the vectors $\boldsymbol{b}^{(jk)}$, $\boldsymbol{c}^{(jk)}$, $\boldsymbol{m}^{(jk)}$, and $\boldsymbol{q}^{(jk)}$, respectively, with the corresponding ith elements

$$\begin{aligned}
b_i^{(jk)} &= \big(\widehat{P}_{\theta_k}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}}_{\theta_k}\big)^\top\Sigma^{\theta_j}\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big) - \boldsymbol{\mu}_{\theta_j}^\top\Sigma^{-1}\big(\widehat{P}_{\theta_k}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}}_{\theta_k}\big), \\
c_i^{(jk)} &= \big(\widehat{P}_{\theta_k}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}}_{\theta_k}\big)^\top\Sigma^{-1}\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)\big(\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)^\top\Sigma^{\theta_j}\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big) \\
&\quad - 2\boldsymbol{\mu}_{\theta_j}^\top\Sigma^{-1}\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)\big), \\
m_i^{(jk)} &= \boldsymbol{d}_i^\top\Sigma^{\theta_j\theta_k}\boldsymbol{d}_i - 2\boldsymbol{\mu}_{\theta_k}^\top\Sigma^{\theta_j}\boldsymbol{d}_i - 2\boldsymbol{\mu}_{\theta_j}^\top\Sigma^{\theta_k}\boldsymbol{d}_i - 2\boldsymbol{\mu}_{\theta_j\theta_k}^\top\Sigma^{-1}\boldsymbol{d}_i + 2\boldsymbol{\mu}_{\theta_j}^\top\Sigma^{-1}\boldsymbol{\mu}_{\theta_k}, \\
q_i^{(jk)} &= \big(\boldsymbol{d}_i^\top\Sigma^{\theta_k}\boldsymbol{d}_i - 2\boldsymbol{\mu}_{\theta_k}^\top\Sigma^{-1}\boldsymbol{d}_i\big)\big(\boldsymbol{d}_i^\top\Sigma^{\theta_j}\boldsymbol{d}_i - 2\boldsymbol{\mu}_{\theta_j}^\top\Sigma^{-1}\boldsymbol{d}_i\big),
\end{aligned}$$

where $\boldsymbol{\mu}_{\theta_j\theta_k} = \partial\boldsymbol{\mu}_{\theta_j}/\partial\theta_k$ and $\Sigma^{\theta_j\theta_k} = \partial\Sigma^{\theta_j}/\partial\theta_k = -2\Sigma^{\theta_k}\Sigma_{\theta_j}\Sigma^{-1} - \Sigma^{-1}\Sigma_{\theta_j\theta_k}\Sigma^{-1}$. We are now able to write the observed information matrix and the derivatives with respect to the data as

$$J = \frac{n}{2}T - R^\top M - V^\top Q, \qquad \ell' = 2R^\top \boldsymbol{w}, \qquad U' = 2\big(R^\top B + V^\top C\big); \tag{8}$$

see Appendix A. The adjustment term of the signed likelihood ratio statistic, ρ, is obtained by replacing $\widehat{J}$, $\widetilde{J}_{\omega\omega}$, $\widehat{\ell}'$, $\widetilde{\ell}'$, and $\widetilde{U}'$ in (6).

Computer packages that perform simple operations on matrices and vectors can be used to calculate ρ. Note that ρ depends on the model through $\boldsymbol{\mu}$ and P with the corresponding first derivative, Σ, with its first two derivatives, and Σ^{-1}. The dependence on the specific distribution of $\boldsymbol{Z}_i$ in the class of elliptical distributions occurs through W_{p_0}. Appendix B gives the required derivatives of $\boldsymbol{\mu}$, Σ, and P for each identifiability condition mentioned in Sect. 2.

4 Simulation study

In this section we shall present the results of a Monte Carlo simulation in which we evaluate the finite sample performance of the signed likelihood ratio test (r_S) and the modified signed likelihood ratio test (r^*). The simulations were based on model (4) when the random vector $\boldsymbol{Z}_i$ has a normal or a Student-t bivariate distribution. All simulations were performed using the `Ox` matrix programming language (Doornik 2006). The number of Monte Carlo replications was 10,000, and the sample sizes considered were $n = 10, 20, 30$, and 40. The tests were carried out at the following nominal levels: $\gamma = 1\%, 5\%, 10\%$. The rejection rates were obtained assuming that the ratio λ_x or λ_e is known or that the intercept α is known and equal to zero. We test $\mathcal{H}_{0a} : \beta \geq 1$ against $\mathcal{H}_{1a} : \beta < 1$ and $\mathcal{H}_{0b} : \beta = 1$ against $\mathcal{H}_{1b} : \beta \neq 1$. The parameter values are $\alpha = 0.5$, except for the case where α is assumed to be known; in this case $\alpha = 0$, $\mu_x = 5.0$, $\sigma_x^2 = 1.5$, $\sigma_u^2 = 0.5$, and $\sigma_e^2 = 2.0$. When the distribution considered is the Student-t, we set the number of degrees of freedom at 3. The null rejection rates of the tests are displayed in Tables 1, 2, and 3 for different sample sizes; entries are percentages.

For λ_x or λ_e known, the signed likelihood ratio test is markedly liberal when the sample is small. For instance, for the case where the distribution is Student-t, the ratio λ_x is known and the sample size is $n = 10$, the two-sided unmodified test displays rejection rates equal to 2.2% ($\gamma = 1\%$), 8.5% ($\gamma = 5\%$), and 15.2% ($\gamma = 10\%$). When the intercept is known, the test based on the signed likelihood ratio statistic is less liberal relatively to the cases where one of the ratios, λ_x or λ_e, is known. In other words, the adjustment is less needed when the intercept is known.

The test based on the modified signed likelihood ratio statistic, r^*, shows better performance than the test based on the signed likelihood ratio statistic in all cases, with rejection rates close to the nominal levels. For example, when the test is two-sided, the distribution is Student-t, λ_e is known, $n = 10$, and $\gamma = 10\%$, the rejection rates are 10.0% (r^*) and 14.5% (r_S). Overall, the best performing test is the one that employs r^* as the test statistic.

A comment on confidence intervals is now in order. Tests may be inverted to give confidence sets. For instance, a confidence interval for a scalar parameter may consist

Table 1 Null rejection rates of the tests of $\mathcal{H}_{0a}$ and $\mathcal{H}_{0b}$—λ_x known

One-sided test ($\mathcal{H}_{0a}$)

	Normal distribution						Student-t distribution ($\nu = 3$)					
	$\gamma = 1\%$		$\gamma = 5\%$		$\gamma = 10\%$		$\gamma = 1\%$		$\gamma = 5\%$		$\gamma = 10\%$	
n	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*
10	2.2	1.2	7.5	4.9	13.3	10.1	2.1	1.1	7.8	5.5	13.2	10.5
20	1.5	1.1	6.4	5.3	11.7	10.4	1.4	1.0	6.0	5.0	11.5	10.0
30	1.3	1.0	6.1	5.5	10.9	9.9	1.4	1.1	6.4	5.7	11.5	10.6
40	1.2	1.0	5.8	5.1	11.2	10.6	1.2	1.0	5.5	4.9	10.8	10.0

Two-sided test ($\mathcal{H}_{0b}$)

	Normal distribution						Student-t distribution ($\nu = 3$)					
	$\gamma = 1\%$		$\gamma = 5\%$		$\gamma = 10\%$		$\gamma = 1\%$		$\gamma = 5\%$		$\gamma = 10\%$	
n	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*
10	2.6	1.0	8.6	5.2	14.9	9.9	2.2	1.2	8.5	5.3	15.2	10.4
20	1.6	1.0	6.4	4.8	12.2	10.0	1.4	1.0	6.3	5.1	11.8	9.8
30	1.3	0.9	5.9	5.0	11.6	10.3	1.3	1.0	5.9	5.1	12.1	10.8
40	1.2	1.0	5.6	4.8	11.2	10.0	1.2	1.0	5.7	5.0	10.8	9.7

Table 2 Null rejection rates of the tests of $\mathcal{H}_{0a}$ and $\mathcal{H}_{0b}$—λ_e known

One-sided test ($\mathcal{H}_{0a}$)

	Normal distribution						Student-t distribution ($\nu = 3$)					
	$\gamma = 1\%$		$\gamma = 5\%$		$\gamma = 10\%$		$\gamma = 1\%$		$\gamma = 5\%$		$\gamma = 10\%$	
n	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*
10	2.2	0.9	7.3	5.0	12.9	9.8	1.8	1.0	7.3	5.2	12.7	10.1
20	1.3	0.8	5.7	4.6	11.2	9.6	1.4	1.0	5.8	4.8	11.1	9.6
30	1.0	0.7	5.7	4.8	10.9	9.9	1.3	1.1	5.7	5.0	10.5	9.8
40	1.1	0.8	5.2	4.7	10.3	9.5	1.3	1.1	5.5	5.0	10.2	9.7

Two-sided test ($\mathcal{H}_{0b}$)

	Normal distribution						Student-t distribution ($\nu = 3$)					
	$\gamma = 1\%$		$\gamma = 5\%$		$\gamma = 10\%$		$\gamma = 1\%$		$\gamma = 5\%$		$\gamma = 10\%$	
n	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*
10	2.4	1.0	8.5	5.1	14.6	9.8	1.8	0.9	7.8	4.8	14.5	10.0
20	1.4	1.0	6.1	4.5	11.7	9.5	1.5	1.1	6.5	5.1	11.9	10.0
30	1.3	0.9	5.6	4.7	11.6	9.9	1.4	1.1	6.1	5.3	11.6	10.2
40	1.2	0.9	5.9	5.0	10.9	10.1	1.2	0.9	5.4	4.8	11.1	10.2

of the parameter values that are not rejected by a given test at the corresponding level. If the level of test is γ, the confidence coefficient of the corresponding confidence interval is $1 - \gamma$. Our simulations then suggest that confidence intervals should be constructed from the inversion of the modified signed likelihood ratio statistic.

Table 3 Null rejection rates of the tests of $\mathcal{H}_{0a}$ and $\mathcal{H}_{0b}$—null intercept

One-sided test ($\mathcal{H}_{0a}$)

	Normal distribution						Student-t distribution ($\nu = 3$)					
	$\gamma = 1\%$		$\gamma = 5\%$		$\gamma = 10\%$		$\gamma = 1\%$		$\gamma = 5\%$		$\gamma = 10\%$	
n	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*
10	1.7	1.1	6.8	5.3	12.7	10.7	1.6	1.2	6.5	5.5	11.8	10.6
20	1.2	0.9	5.7	5.1	11.1	10.0	1.2	1.1	5.5	5.1	10.8	10.3
30	1.2	1.0	5.4	4.9	11.0	10.4	1.3	1.1	5.3	5.1	10.7	10.3
40	1.1	1.0	5.5	5.2	10.8	10.4	1.3	1.2	5.6	5.3	10.7	10.5

Two-sided test ($\mathcal{H}_{0b}$)

	Normal distribution						Student-t distribution ($\nu = 3$)					
	$\gamma = 1\%$		$\gamma = 5\%$		$\gamma = 10\%$		$\gamma = 1\%$		$\gamma = 5\%$		$\gamma = 10\%$	
n	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*	r_S	r^*
10	1.7	1.0	7.0	5.2	13.0	10.2	1.6	1.1	6.4	5.3	12.0	10.3
20	1.0	0.8	5.6	4.6	11.4	10.1	1.1	0.9	5.3	4.7	10.6	9.7
30	1.2	1.1	5.5	4.8	10.4	9.7	1.3	1.2	5.4	5.0	10.7	10.1
40	1.2	1.1	5.3	4.9	10.6	9.9	1.3	1.2	5.9	5.5	11.2	10.8

5 Babies data

In this section we illustrate an application of the usual signed likelihood ratio test and its modified version in a real dataset. The observations are pairs of measurements of serum kanamycin levels in blood samples drawn from twenty premature babies (Kelly 1984). One of the measurements was obtained by a heelstick method (x), the other using an umbilical catheter (y). Since both methods are subject to measurement errors, model (2)–(3) seems to be adequate to fit the data. Following Kelly (1984), we assume the normality for the error terms and that $\lambda_e = 1$, that is, $\sigma_e^2 = \sigma_u^2$.

The maximum likelihood estimates of the parameters (standard errors between parentheses) are $\widehat{\beta} = 1.070$ (0.159), $\widehat{\alpha} = -1.160$ (3.390), $\widehat{\mu}_x = 20.855$ (1.112), $\widehat{\sigma}_x^2 = 20.352$ (7.811), and $\widehat{\sigma}_u^2 = 4.374$ (1.383). Here, the interest lies in testing $\mathcal{H}_0 : \beta = 1$ against $\mathcal{H}_1 : \beta \neq 1$. We have $r_S = 0.452$ (p-value $= 0.651$) and $r^* = 0.424$ (p-value $= 0.671$). Both tests lead to the same conclusion, namely that $\mathcal{H}_0$ should not be rejected at the usual significance levels. Here, the adjustment had negligible impact on the value of the test statistic.

If the sample size were smaller, the adjustment could be much more pronounced. To illustrate this, a randomly chosen subset of the data set with 17 premature babies was drawn. The maximum likelihood estimates are now $\widehat{\beta} = 1.285$ (0.186), $\widehat{\alpha} = -4.935$ (3.737), $\widehat{\mu}_x = 19.800$ (0.953), $\widehat{\sigma}_x^2 = 12.803$ (5.276), and $\widehat{\sigma}_u^2 = 2.634$ (0.904). The observed values of the test statistics are $r_S = 1.677$ (p-value $= 0.093$) and $r^* = 1.554$ (p-value $= 0.120$). Clearly, the unmodified test rejects the null hypothesis at the 10% significance level unlike the adjusted test. Also, approximately 90% confidence intervals for β constructed from the inversion of the tests that use r

and r^* are, respectively, (1.007, 1.667) and (0.984, 1.593). Therefore, the confidence interval obtained from r^* does not contradict the hypothesis that $\beta = 1$ unlike the one obtained from r^*. The tests lead to different conclusions, and the conclusion obtained using the adjusted test is compatible with that achieved using the complete dataset.

6 Concluding remarks

We considered a class of measurement error regression models that allow the error terms to follow an elliptical distribution. Inference in this class of models rely on asymptotic approximations, which can be inaccurate if the sample is not large. The modified signed likelihood ratio statistic proposed by Barndorff-Nielsen (1986) has an approximate standard normal distribution with high degree of accuracy and hence is very attractive for inference purposes. The main difficulty in deriving the modified statistic is the identification, if it is possible, of an appropriate ancillary statistic. In this paper, we found an adequate ancillary statistic and derived Barndorff-Nielsen's modified signed likelihood ratio statistic. We gave closed-form expressions for the adjustment term for different identifiability conditions. Our simulation results indicated that the modified test is much more reliable than the unmodified test when the sample size is small.

Acknowledgement We gratefully acknowledge financial support from FAPESP and CNPq. We thank an anonymous referee for constructive comments that improved the paper.

Appendix A: The observed information matrix and derivatives with respect to the data

The first- and second-order derivatives of the log-likelihood function (5) with respect to the parameters are

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_j} = -\frac{n}{2}\mathrm{tr}\big(\Sigma^{-1}\Sigma_{\theta_j}\big) + \sum_{i=1}^{n} W_{p0}\big(\boldsymbol{d}_i^\top \Sigma^{-1}\boldsymbol{d}_i\big)\big(\boldsymbol{d}_i^\top \Sigma^{\theta_j}\boldsymbol{d}_i - 2\boldsymbol{\mu}_{\theta_j}\Sigma^{-1}\boldsymbol{d}_i\big),$$

$$\begin{aligned}
J_{\theta_j\theta_k} = -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k} &= \frac{n}{2}\mathrm{tr}\big(\Sigma^{\theta_k}\Sigma_{\theta_j}\big) + \frac{n}{2}\mathrm{tr}\big(\Sigma^{-1}\Sigma_{\theta_j\theta_k}\big) \\
&\quad - \sum_{i=1}^{n}\big\{W'_{p0}\big(\boldsymbol{d}_i^\top\Sigma^{-1}\boldsymbol{d}_i\big)\big[\boldsymbol{d}_i^\top\Sigma^{\theta_k}\boldsymbol{d}_i - 2\boldsymbol{\mu}_{\theta_k}^\top\Sigma^{-1}\boldsymbol{d}_i\big] \\
&\quad \times \big[\boldsymbol{d}_i^\top\Sigma^{\theta_j}\boldsymbol{d}_i - 2\boldsymbol{\mu}_{\theta_j}^\top\Sigma^{-1}\boldsymbol{d}_i\big] + W_{p0}\big(\boldsymbol{d}_i^\top\Sigma^{-1}\boldsymbol{d}_i\big)\big[\boldsymbol{d}_i^\top\Sigma^{\theta_j\theta_k}\boldsymbol{d}_i - 2\boldsymbol{\mu}_{\theta_k}^\top\Sigma^{\theta_j}\boldsymbol{d}_i \\
&\quad - 2\boldsymbol{\mu}_{\theta_j}^\top\Sigma^{\theta_k}\boldsymbol{d}_i - 2\boldsymbol{\mu}_{\theta_j\theta_k}^\top\Sigma^{-1}\boldsymbol{d}_i + 2\boldsymbol{\mu}_{\theta_j}^\top\Sigma^{-1}\boldsymbol{\mu}_{\theta_k}\big]\big\}.
\end{aligned}$$

The (j,k)th element of the observed information matrix $J = J(\boldsymbol{\theta})$ is $J_{\theta_j\theta_k}$. In matrix notation, the observed information matrix can be written as in (8).

Let $\boldsymbol{a} = (\boldsymbol{a}_1^\top, \boldsymbol{a}_2^\top, \ldots, \boldsymbol{a}_n^\top)^\top$, where $\boldsymbol{a}_i = \boldsymbol{a}_i(z)$ is given in (7). For simplicity, we write $\widehat{\mu} = \widehat{\mu}(z)$ and $\widehat{P} = \widehat{P}(z)$. Replacing z_i by $\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}}$ in the log-likelihood function (5), we have

$$\ell(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}, \boldsymbol{a}) = -\frac{n}{2} \log|\Sigma| + \sum_{i=1}^{n} \log p_0\big[\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)^\top \Sigma^{-1}\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)\big].$$

Therefore, the jth element of the vector ℓ' is

$$\begin{aligned}\ell'_j = 2\sum_{i=1}^{n} W_{p0}\big[\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)^\top \Sigma^{-1}\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)\big] \\ \times \big[\big(\widehat{P}_{\theta_j}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}}_{\theta_j}\big)^\top \Sigma^{-1}\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)\big],\end{aligned}$$

where $\widehat{P}_{\theta_j}$ is given in Sect. 3. Now, the (j,k)th element of the matrix U' is

$$\begin{aligned}U'_{jk} = \frac{\partial}{\partial\widehat{\theta}_k}\left[\frac{\partial\ell(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}, \boldsymbol{a})}{\partial\theta_j}\right] &= 2\sum_{i=1}^{n}\big\{W_{p0}\big[\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)^\top \Sigma^{-1}\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)\big] \\ &\times \big[\big(\widehat{P}_{\theta_k}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}}_{\theta_k}\big)^\top \Sigma^{\theta_j}\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big) - \boldsymbol{\mu}_{\theta_j}^\top \Sigma^{-1}\big(\widehat{P}_{\theta_k}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}}_{\theta_k}\big)\big] \\ &+ W'_{p0}\big[\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)^\top \Sigma^{-1}\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)\big] \\ &\times \big[\big(\widehat{P}_{\theta_k}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}}_{\theta_k}\big)^\top \Sigma^{-1}\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)\big] \\ &\times \big[\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)^\top \Sigma^{\theta_j}\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big) - 2\boldsymbol{\mu}_{\theta_k}^\top \Sigma^{-1}\big(\widehat{P}\boldsymbol{a}_i + \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big)\big]\big\}.\end{aligned}$$

In matrix notation, ℓ' and U' can be written as in (8).

Appendix B: Derivatives of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{P}$

In the following we give the first and second derivatives of $\boldsymbol{\mu}$, Σ, and P with respect to the unknown parameters for each identifiability condition. Only nonnull derivatives are presented. The matrix

$$P = \begin{pmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{pmatrix}$$

comes from the Cholesky decomposition $PP^\top = \Sigma$, and p_{11}, p_{21}, and p_{22} are given below.

Condition 1: $\lambda_e = \sigma_e^2/\sigma_u^2$ known.

When $\lambda_e = \sigma_e^2/\sigma_u^2$ is known, we have $\boldsymbol{\mu}_{\theta_1} = (\mu_x, 0)^\top$, $\boldsymbol{\mu}_{\theta_2} = (1, 0)^\top$, $\boldsymbol{\mu}_{\theta_3} = (\beta, 1)^\top$,

$$\Sigma_{\theta_1} = \begin{pmatrix} 2\beta\sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & 0 \end{pmatrix}, \qquad \Sigma_{\theta_4} = \begin{pmatrix} \beta^2 & \beta \\ \beta & 1 \end{pmatrix},$$

$$\boldsymbol{\mu}_{\theta_1\theta_3} = \boldsymbol{\mu}_{\theta_3\theta_1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \Sigma_{\theta_1\theta_1} = \begin{pmatrix} 2\sigma_x^2 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \Sigma_{\theta_1\theta_4} = \Sigma_{\theta_4\theta_1} = \begin{pmatrix} 2\beta & 1 \\ 1 & 0 \end{pmatrix}.$$

Here, $p_{11} = \sqrt{\beta^2\sigma_x^2 + \lambda_e\sigma_u^2}$, $p_{21} = \beta\sigma_x^2/p_{11}$, and $p_{22} = \sqrt{p_{11}^2(\sigma_x^2 + \sigma_u^2) - \beta^2(\sigma_x^2)^2}/p_{11}$. The derivatives of P with respect to the parameters $\beta, \sigma_x^2, \sigma_u^2$ are, respectively,

$$P_{\theta_1} = \begin{pmatrix} p_{21} & 0 \\ \lambda_e\sigma_x^2\sigma_u^2/p_{11}^3 & (-\lambda_e\sigma_x^2\sigma_u^2 p_{21})/(p_{11}^3 p_{22}) \end{pmatrix},$$

$$P_{\theta_4} = \begin{pmatrix} \beta^2/(2p_{11}) & 0 \\ (\beta^3\sigma_x^2 + 2\beta\lambda_e\sigma_u^2)/(2p_{11}^3) & \lambda_e^2(\sigma_u^2)^2/(2p_{11}^4 p_{22}) \end{pmatrix},$$

and

$$P_{\theta_5} = \begin{pmatrix} \lambda_e/(2p_{11}) & 0 \\ -\beta\lambda_e\sigma_x^2/(2p_{11}^3) & (\lambda_e p_{21}^2 + p_{11}^2)/(2p_{11}^2 p_{22}) \end{pmatrix}.$$

Condition 2: $\lambda_x = \sigma_x^2/\sigma_u^2$ known.

When λ_x is known, $\boldsymbol{\mu}_{\theta_j}$ and $\boldsymbol{\mu}_{\theta_j\theta_k}$ for $j, k = 1, 2, 3, 4, 5$ coincide with those of the previous case. The matrices Σ_{θ_j} and $\Sigma_{\theta_j\theta_k}$ are

$$\Sigma_{\theta_1} = \begin{pmatrix} 2\lambda_x\beta\sigma_u^2 & \lambda_x\sigma_u^2 \\ \lambda_x\sigma_u^2 & 0 \end{pmatrix}, \qquad \Sigma_{\theta_2} = \Sigma_{\theta_3} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \Sigma_{\theta_4} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\Sigma_{\theta_5} = \begin{pmatrix} \lambda_x\beta^2 & \lambda_x\beta \\ \lambda_x\beta & \lambda_x + 1 \end{pmatrix}, \qquad \Sigma_{\theta_1\theta_1} = \begin{pmatrix} 2\lambda_x\sigma_u^2 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\Sigma_{\theta_1\theta_5} = \Sigma_{\theta_5\theta_1} = \begin{pmatrix} 2\lambda_x\sigma_u^2\beta & \lambda_x \\ \lambda_x & 0 \end{pmatrix}.$$

Here, the nonnull elements of P are $p_{11} = \sqrt{\lambda_x\beta^2\sigma_u^2 + \sigma_e^2}$, $p_{21} = \lambda_x\beta\sigma_u^2/p_{11}$, and $p_{22} = \sqrt{\sigma_u^2(\lambda_x\sigma_e^2 + p_{11}^2)}/p_{11}$. The derivatives of P with respect to $\beta, \sigma_u^2, \sigma_e^2$ are, respectively,

$$P_{\theta_1} = \begin{pmatrix} p_{21} & 0 \\ \lambda_x\sigma_u^2\sigma_e^2/p_{11}^3 & -\lambda_x\sigma_u^2\sigma_e^2 p_{21}/(p_{11}^3 p_{22}) \end{pmatrix},$$

$$P_{\theta_4} = \begin{pmatrix} \lambda_x\beta^2/(2p_{11}) & 0 \\ (\lambda_x\beta p_{11}^2 + \lambda_x\beta\sigma_e^2)/(2p_{11}^3) & (\lambda_x(\sigma_e^2)^2 + p_{11}^4)/(2p_{11}^4 p_{22}) \end{pmatrix},$$

$$P_{\theta_5} = \begin{pmatrix} 1/(2p_{11}) & 0 \\ -p_{21}/(2p_{11}^2) & p_{21}^2/(2p_{11}^2 p_{22}) \end{pmatrix}.$$

Condition 3: Known intercept.

When α is known, $\boldsymbol{\mu}_{\theta_1}$ and Σ_{θ_1} are equal to those given under Condition 1, $\boldsymbol{\mu}_{\theta_2} = (\beta, 1)^\top$,

$$\Sigma_{\theta_3} = \begin{pmatrix} \beta^2 & \beta \\ \beta & 1 \end{pmatrix}, \qquad \Sigma_{\theta_4} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \Sigma_{\theta_5} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\boldsymbol{\mu}_{\theta_1\theta_2} = \boldsymbol{\mu}_{\theta_2\theta_1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \Sigma_{\theta_1\theta_1} = \begin{pmatrix} 2\sigma_x^2 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \Sigma_{\theta_1\theta_3} = \Sigma_{\theta_3\theta_1} = \begin{pmatrix} 2\beta & 1 \\ 1 & 0 \end{pmatrix}.$$

In this case, $p_{11} = \sqrt{\beta^2\sigma_x^2 + \sigma_e^2}$, $p_{21} = \beta\sigma_x^2/p_{11}$, and $p_{22} = \sqrt{\sigma_u^2 p_{11}^2 + \sigma_x^2\sigma_e^2}/p_{11}$. The derivatives of P with respect to the parameters $\beta, \sigma_x^2, \sigma_u^2, \sigma_e^2$ are, respectively, given by

$$P_{\theta_1} = \begin{pmatrix} p_{21} & 0 \\ \sigma_x^2\sigma_e^2/p_{11}^3 & -\beta(\sigma_x^2)^2\sigma_e^2/(p_{11}^4 p_{22}) \end{pmatrix},$$

$$P_{\theta_3} = \begin{pmatrix} \beta^2/(2p_{11}) & 0 \\ (\beta p_{11}^2 + \beta\sigma_e^2)/(2p_{11}^3) & (\sigma_e^2)^2/(2p_{11}^4 p_{22}) \end{pmatrix},$$

$$P_{\theta_4} = \begin{pmatrix} 0 & 0 \\ 0 & 1/(2p_{22}) \end{pmatrix}, \qquad P_{\theta_5} = \begin{pmatrix} 1/(2p_{11}) & 0 \\ -p_{21}/(2p_{11}^2) & p_{21}^2/(2p_{11}^2 p_{22}) \end{pmatrix}.$$

References

Aoki, R., Bolfarine, H., Singer, J.M.: Null intercept measurement error regression models. Test **10**, 441–457 (2001)

Arellano-Valle, R.B., Bolfarine, H.: Elliptical structural models. Commun. Stat. Theory Methods **25**, 2319–2341 (1996)

Barndorff-Nielsen, O.E.: Inference on full or partial parameters, based on the standardized signed log likelihood ratio. Biometrika **73**, 307–322 (1986)

Barndorff-Nielsen, O.E.: Modified signed log likelihood ratio. Biometrika **78**, 557–563 (1991)

Barndorff-Nielsen, O.E., Blaesild, P., Eriksen, P.S.: Decomposition and Invariance of Measures, and Statistical Transformation Models. Springer, Heidelberg (1989)

Chan, L.K., Mak, T.K.: On the maximum likelihood estimation of a linear structural relationship when the intercept is known. J. Multivar. Anal. **9**, 304–313 (1979)

Cheng, C.L., Van Ness, J.W.: Statistical Regression with Measurement Error. Oxford University Press, London (1999)

Doornik, J.A.: Ox: An Object-Oriented Matrix Language. Timberlake Consultants Press, London (2006)

Fang, K.T., Anderson, T.W.: Statistical Inference in Elliptically Contoured and Related Distributions. Allerton Press Inc, New York (1990)

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

Fuller, S.: Measurement Error Models. Wiley, New York (1987)

Kelly, G.: The influence function in the errors in variables problem. Ann. Stat. **12**, 87–100 (1984)

Pace, L., Salvan, A.: Principles of Statistical Inference from a Neo-Fisherian Perspective. World Scientific, Singapore (1997)

Severini, T.A.: Likelihood Methods in Statistics. Oxford University Press, London (2000)

Wong, M.Y.: Likelihood estimation of a simple linear regression model when both variables have error. Biometrika **76**, 141–148 (1989)

Wong, M.Y.: Bartlett adjustment to the likelihood ratio statistic for testing several slopes. Biometrika **78**, 221–224 (1991)