Statistical Papers manuscript No.
(will be inserted by the editor)

On Estimation and Local Influence Analysis for Measurement Errors Models under Heavy-tailed Distributions

V. H. Lachos · T. Angolini · C. A. Abanto-Valle

Received: date / Accepted: date

Abstract Scale mixtures of normal distribution is a class of symmetric thick–tailed distributions that includes the normal one as a special case. In this paper we consider local influence analysis for measurement error model (MEM) when the random error and the unobserved value of the covariates follows jointly a scale mixtures of normal

The first and second authors acknowledges the partial financial support from CNPq and FAPESP-Brazil. The third author acknowledges the partial financial support from FAPERJ-Brazil.

V. H. Lachos
Department of Statistics, Campinas State University, CP 6065, CEP 13083-859, SP, Brazil
E-mail: hlachos@ime.unicamp.br

T. Angolini
Department of Statistics, Campinas State University, CP 6065, CEP 13083-859, SP, Brazil

C. A. Abanto–Valle
Department of Statistics, Federal University of Rio de Janeiro, CP 68530, CEP 21945-970, RJ, Brazil
Tel.: +55-21-2562-7505 ext. 201 Fax: +55-21-2562-7374
E-mail: cabantovalle@im.ufrj.br

distribution, providing an appealing robust alternative to the usual Gaussian process in measurement error models. As the observed data log-likelihood associated with this model is intractable, Cook's well–known approach may be hard to be applied to obtain measures of local influence. Instead we develop local influence measures following the approach of Zhu and Lee (2001), which is based on the use of an EM-type algorithm. Three specific perturbation schemes are discussed. Results obtained from a real data set are reported, illustrating the usefulness of the proposed methodology.

Keywords EM algorithm · Scale mixtures of normal distribution · Mahalanobis distance · Measurement error models.

1 Introduction

Linear regression is one of the most widely used statistical tools and it has been subject to extensive research in the literature for over a century. In many applications the explanatory variable (x) is not directly observed, so it is measured with errors. There is a considerable amount of work on the problem of parameter estimation when the explanatory variable is measured with errors, see for instance, Fuller (1987) and Cheng and Van-Ness (1999) for references. As in Fuller (1987), consider a bivariate random variable (x_i, η_i) satisfying the linear relation $\eta_i = \alpha + \beta x_i$, $i = 1, \cdots, n$, where α represents the (unknown) intercept and β stands for the (unknown) slope. Suppose that x_i and η_i can not be observed directly, but instead we observe

$$X_i = x_i + u_i, \tag{1}$$

$$Y_i = \eta_i + e_i, \quad i = 1, \ldots, n, \tag{2}$$

i.e., the observed quantities X_i and Y_i are measured with error. Let x_i denote the true (unknown) value of the observed value of the covariate, X_i, and the errors u_i

and e_i to have zero means and unknown variances ϕ_u and ϕ_e, respectively and to be independently distributed, $i = 1, \ldots, n$. It is common to assume that all the random variables in the error-in-variables regression model are jointly normal. In this case it is well known that such a model is not identifiable and to bypass this inconvenience, we must make an extra assumption about the parameters (Kendall and Stuart, 1973). Among the available alternatives, we identify i) ϕ_u or (and) ϕ_e known; ii) $\lambda = \phi_u/\phi_e$ known; iii) $k_x = \phi_x/(\phi_x + \phi_e)$ known or iv) α known, where ϕ_x is the variance of x_i, $i = 1, \cdots, n$.

In the normal structural model with no side conditions , $E(x_i) = \mu_x$ is the only parameter that is identifiable (see Lindley, 1972, p. 46, for some comments on this model considering the Bayesian approach). Chan and Mak (1979) studied the MLE of the parameters when the intercept is known and presented the information matrix explicitly. Patefield (1985) considered the case where the intercept and the ratio of the variances are known. Aoki et al (2003) discussed the model with known intercept and repeated measurement data. The MEM are useful in many disciplines, including linear and nonlinear errors–in–variables regression models, factor analysis models, latent structural models and simultaneous equations models. It is also used in the problem of comparing measurement device (see, Barnett, 1969; Theobald and Mallison, 1978; Shyr and Gleser, 1986; Bolfarine and Galea-Rojas, 1996; Chipkevitch et al, 1996), which varies in price, time spent to measure and other features, such as efficiency. Grubbs (1973) analyzed such an experiment for comparing three cronometres; Barnett (1969) used MEM for the comparison of four combinations of two instruments and two operators for measuring vital capacity. Several other examples of MEM model in the medical area are reported in the literature, especially in Kelly (1984), Chipkevitch et al (1996) and Lu et al (1997). Examples in agriculture are considered in Fuller (1987), and examples in psychology and education were considered by Dunn (1992). However, all these works mostly assumed normality as the basis of their inference; there are relatively few works that have considered MEM model under symmetric elliptical distributions. Some results on MEM using t–distribution can be found in Bolfarine and Galea-Rojas (1996), and Galea-Rojas et al (2005). Recently, Arellano-Valle et al (2005) considered a study using the skew–normal distribution. Hence, a study of its properties under nonstandard assumptions, like normality for example, is very pertinent.

In this paper, we consider a classical approach for measurement error models, where the unobserved value of the covariate (x) and the vector of random errors are assumed to follows jointly a scale mixtures of normal distribution (Andrews and Mallows, 1974). This extension result in a flexible class of models for robust measurement error models called of measurement error model with scale mixtures of normal distribution (hereafter, SAN-MEM). Advantages of using such general structures include ease of interpretation, as well as estimation efficiency. Moreover, the class of scale mixtures of normal distribution is a rich class of distributions that contains as proper elements, the normal (N–MEM), the Student-t (T–MEM), the slash (SL–MEM), the power exponential (PE–MEM) and the contaminated normal (CN–MEM) one. All these distributions have heavier tails than the normal, and thus can be used for robust inference in many type of models. This work is motivated by the fact that many data sets considered in the literature seem to present significant non–normal behavior, such as heavy tails.

On the other hand, an investigation of influence is an important step in data analysis following parameter estimation. This can be achieved by conducting local influence analysis, a general statistical technique used to assess the stability of the estimation outputs with respect to the model inputs. Model inputs may include data, parameters or other information regarding the model (e.g., heteroscedasticity). Outputs may include the parameter estimates, final objective function values, estimates of residuals, etc. Following the pioneering work of Cook (1986), this area of research has received considerable attention in the recent statistical literature: see Zhu and Lee (2001), Lee and Xu (2004), Galea-Rojas et al (2005) and Osorio et al (2007), among others. However, as the observed log–likelihood function of the SMN-MEM involves intractable integrals, direct application of Cook's (1986) approach for SMN-LMM is very difficult, because these measures involve the first and second partial derivatives of this function. Recently, Zhu and Lee (2001) developed an approach to performing local influence analysis for general statistical models with missing data by working with a Q-displacement function closely related to the conditional expectation of the complete-data log-likelihood at the E-step of the EM algorithm. Inspired by Lee and Xu (2004) and Lee and Tang (2004), in this paper we apply Zhu and Lee's (2001) local influence approach to the SMN-MEM. As we will show, such an approach leads to closed-form influence measures.

The paper is organized as follows. In Section 2, for the sake of completeness, we give a brief sketch of SMN distribution. In Section 3 we present the elliptical–MEM. In Section 4 we discuss the EM–algorithm for maximum likelihood (ML) estimation in SMN–MEM and we derived analytically the observed information matrix. In Section 5, we give a brief sketch of the local influence approach for models with incomplete-data and develop the methodology required for the SMN-MEM. Three different perturbation schemes are considered. The methodology proposed for SMN–MEM models is illustrated in Section 6 considering a real data set and finally, some conclusive remarks are presented in Section 7.

2 Scale mixtures of normal distribution

A SMN distribution (Andrews and Mallows, 1974) is defined as the p−dimensional random vector

$$\mathbf{Y} = \boldsymbol{\mu} + \kappa^{1/2}(U)\mathbf{Z}, \tag{3}$$

where $\boldsymbol{\mu}$ is a location vector, $\mathbf{Z}$ is a normal random vector with mean vector $\mathbf{0}$, variance–covariance matrix $\boldsymbol{\Sigma}$, $\kappa(.)$ is a weight function and U is a mixing positive random variable with cumulative distribution function (cdf) $H(u;\boldsymbol{\nu})$ and probability density function (pdf) $h(u;\boldsymbol{\nu})$, independent of $\mathbf{Z}$, where $\boldsymbol{\nu}$ is a scalar or parameter vector indexing the distribution of U. Given U, $\mathbf{Y}$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance–covariance matrix $\kappa(u)\boldsymbol{\Sigma}$, i.e., $\mathbf{Y}|U = u \sim N_p(\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})$. Hence, the pdf of $\mathbf{Y}$ is given by

$$f(\mathbf{y}) = \int_0^{\infty} \phi_p(\mathbf{y}|\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})dH(u), \tag{4}$$

where $\phi_p(.|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the pdf of the p–variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariate matrix $\boldsymbol{\Sigma}$. One particular case of this distribution is the normal distribution, for which H is degenerate, with $\kappa(u) = 1$, $u > 0$. In the sequel, we denote $\mathrm{SMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$ as the SMN distribution with the pdf (4). We notice that when $\boldsymbol{\lambda} = \mathbf{0}$ and $k(u) = u^{-1}$ in (3), the distribution of $\mathbf{Y}$ reduces to the normal/independent (NI) distribution family discussed, for instance, in Lange and Sinsheimer (1993).

The symmetrical class of SMN distribution include distributions such as the Student–t, the slash, the power exponential and the contaminated–normal one; all these distributions have heavier tails than the normal ones and can be used for robust inference. In the next section, we present some special cases of SMN distributions, and for each element of this class, we also compute the conditional moments defined by

$u = E[\kappa^{-1}(U)|\mathbf{y}]$. Derivation of these moments will be useful in the implementation of the EM–algorithm. Other members of SMN distributions can be found in Andrews and Mallows (1974). However, for many of the cases, the scale distribution $H(u;\boldsymbol{\nu})$ does not have a computationally attractive form and thus it will not be dealt with in this work.

2.1 Multivariate Student–t distribution

The multivariate Student–t distribution with ν degrees of freedom, $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$, can be derived from the mixture model (4), by taking $\kappa(u) = 1/u$, with U distributed as $Gamma(\nu/2, \nu/2)$, $u > 0$, $\nu > 0$. The pdf of $\mathbf{Y}$ takes the following form:

$$f(\mathbf{y}) = \frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}}\nu^{-p/2}|\Sigma|^{-1/2}\left(1 + \frac{d}{\nu}\right)^{-(p+\nu)/2}, \quad \mathbf{y} \in \mathbb{R}^p.$$

A particular case of the Student–t distribution is the Cauchy one, when $\nu = 1$. Also, when $\nu \uparrow \infty$, we get the normal distribution as the limiting case. It follows that

$$q(d) = E[\kappa^{-1}(U)|\mathbf{y}] = \frac{\nu + p}{\nu + d}, \;\; d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \sim pF(p, \nu).$$

Applications of the Student–t distribution to robust estimation in MEM can be found in Galea-Rojas et al (2005).

2.2 Multivariate slash distribution

Another SMN distribution, termed multivariate slash distribution and denoted by $SL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$, arise when $\kappa(u) = 1/u$ and the distribution of U is $Beta(\nu, 1)$, $0 < u < 1$ and $\nu > 0$. Its pdf is given by

$$f(\mathbf{y}) = \nu \int_0^1 u^{\nu-1}\phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})du, \quad \mathbf{y} \in \mathbb{R}^p.$$

The slash distribution reduces to the normal distribution when $\nu \uparrow \infty$. It follows that

$$q(d) = E[\kappa^{-1}(U)|\mathbf{y}] = \left(\frac{p+2\nu}{d}\right)\frac{P_1(p/2+\nu+1, d/2)}{P_1(p/2+\nu, d/2)},$$

where $P_x(a,b)$ denotes the cdf of the $Gamma(a,b)$ distribution evaluated at x. From Lange and Sincheimer (1993), the Mahalanobis distance has cdf

$$Pr(d \leq r) = Pr(\chi^2_p \leq r) - \frac{2^\nu \Gamma(p/2+\nu)}{r^\nu \Gamma(p/2)} Pr(\chi^2_{p+2\nu} \leq r).$$

2.3 Multivariate contaminated normal distribution

The multivariate contaminated normal distribution is denoted by $CN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu, \gamma)$, where $0 \leq \nu \leq 1$, $0 < \gamma \leq 1$. Here, $\kappa(u) = 1/u$ and U is a discrete random variable taking one of two states. The probability density function of U, given the parameter vector $\boldsymbol{\nu} = (\nu, \gamma)^\top$, is denoted by

$$h(u; \boldsymbol{\nu}) = \nu \mathbb{I}_{(u=\gamma)} + (1-\nu)\mathbb{I}_{(u=1)}, \;\; 0 \leq \nu \leq 1, 0 < \gamma \leq 1,$$

It follows then that

$$f(\mathbf{y}) = \nu \phi_p(\mathbf{y}|\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Sigma}) + (1-\nu)\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Parameter ν can be interpreted as the proportion of outliers while γ may be interpreted as a scale factor. The skew–contaminated normal distribution reduces to the normal one when $\gamma = 1$. In this case, we have

$$q(d) = E[\kappa^{-1}(U)|\mathbf{y}] = \frac{1-\nu+\nu\gamma^{p/2+1}e^{(1-\gamma)d/2}}{1-\nu+\nu\gamma^{p/2}e^{(1-\gamma)d/2}},$$

and

$$Pr(d \leq r) = \nu Pr(\chi^2_p \leq \gamma r) + (1-\nu) Pr(\chi^2_p \leq r).$$

2.4 Multivariate power exponential distribution

The multivariate power exponential distribution is denoted by $PE_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$, where $\nu > 0$ is the shape parameter. The pdf of $\mathbf{Y}$ is given by

$$f(\mathbf{y}) = \frac{\nu \Gamma(\frac{p}{2})}{\pi^{p/2}\Gamma(\frac{p}{2\nu})2^{p/2\nu}} |\boldsymbol{\Sigma}|^{-1/2} \exp(-d^{\nu}/2).$$

It is possible to express the power exponential distribution in the class of scale mixtures of normal distribution (West, 1987; Lange and Sinsheimer, 1993), however the evaluation of $E[\kappa^{-1}(U)|\mathbf{y}]$ may be hard. In this work we apply the approach proposed by Lange and Sinsheimer (1993, Sec. 3), for which we obtain

$$q(d) = \nu d^{\nu-1},$$

where $d \neq 0$ and $\nu \neq \frac{1}{2}$, and

$$P(d \leq r) = r^{p/2} G(p/\nu, (r^{\nu/2})/2)/(\Gamma(p/\nu)2^{p/\nu}),$$

where $G(\alpha, \beta)$ is the incomplete Gamma function.

3 The model and likelihood based inference

Let $\boldsymbol{\epsilon}_i = (u_i, e_i)^{\top}$ and $\mathbf{Z}_i = (X_i, Y_i)^{\top}$, then the model defined by equations (1)-(2) can be written, in matrix notation, as

$$\mathbf{Z}_i = \mathbf{a} + \mathbf{b}x_i + \boldsymbol{\epsilon}_i = \mathbf{a} + \mathbf{B}\mathbf{r}_i, \tag{5}$$

where $\mathbf{a} = (0, \alpha)^{\top}$ and $\mathbf{b} = (1, \beta)^{\top}$ are 2×1 vectors, $\mathbf{B} = [\mathbf{b}; \mathbf{I}_2]$ is a 2×3 matrix and $\mathbf{r}_i = (x_i, \boldsymbol{\epsilon}_i^{\top})^{\top}$ is a 3×1 vector, $i = 1, \ldots, n$. Thus, using (5), the distribution of $\mathbf{Z}_i$

become specified once the distribution of $\mathbf{r}_i$ is specified, $i = 1, \ldots, n$. In this paper for robust estimation of the parameter, we assume that

$$\mathbf{r}_i = \begin{pmatrix} x_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \overset{\text{iid}}{\sim} \text{SMN}_3 \left(\begin{pmatrix} \mu_x \\ \mathbf{0} \end{pmatrix}, D(\phi_x, \boldsymbol{\phi}); H \right), \tag{6}$$

$i = 1, \ldots, n$, where $D(\phi_x, \boldsymbol{\phi}) = diag(\phi_x, \phi_u, \phi_e)^\top$, with $\boldsymbol{\phi} = (\phi_u, \phi_e)$, further we assume that $\lambda = \phi_u/\phi_e$ is known in order to make the model identifiable. Model (5) together with Model (6) will be called structural SMN-MEM. From (3), this formulation implies that

$$\begin{pmatrix} x_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} | U_i = u_i \sim N_3 \left(\begin{pmatrix} \mu_x \\ \mathbf{0} \end{pmatrix}, \kappa(u_i) D(\phi_x, \boldsymbol{\phi}) \right), \tag{7}$$

$$U_i \sim H(u_i; \boldsymbol{\nu}), \tag{8}$$

$i = 1, \ldots, n$. As the scale matrix of the normal distribution defined in (7) is diagonal it follows that conditional on U_i, $\boldsymbol{\epsilon}_i$ and x_i are independent, with

$$\boldsymbol{\epsilon}_i | U_i = u_i \overset{\text{ind}}{\sim} N_2(\mathbf{0}, \kappa(u_i)\text{D}(\boldsymbol{\phi})) \ \text{ and } \ x_i | U_i = u_i \overset{\text{ind}}{\sim} N_1(\mu_x, \kappa(u_i)\phi_x),$$

and further marginally $\boldsymbol{\epsilon}_i \sim \text{SMN}_2(\mathbf{0}, D(\boldsymbol{\phi}); H)$ and $x_i \sim \text{SMN}_1(\mu_x, \phi_x; H)$. Since for each $i = 1, \ldots, n$, $\boldsymbol{\epsilon}_i$ and x_i are indexed by the same scale mixing factor U_i, they are not independent in general. The independence corresponds to the case where H is degenerate, with $\kappa(U_i) = 1$, $i = 1, \ldots, n$, so that the SMN–MEM reduces to the N–MEM defined in Fuller (1987). However, as stated in (7)-(8), conditional on U_i, $\boldsymbol{\epsilon}_i$ and x_i are independent for each $i = 1, \ldots, n$, which implies that $\boldsymbol{\epsilon}_i$ and x_i are not correlated, once $Cov(\boldsymbol{\epsilon}_i x_i) = E[\boldsymbol{\epsilon}_i x_i | U_i] = 0$. As in Lange et al (1989), Osorio (2006) and Osorio et al (2007), we assumed that $\boldsymbol{\nu}$ is known. Classical inference on the parameter vector

$\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \phi_x, \phi_u)$ in this type of model is based on the marginal distribution for $\mathbf{Z}_i$ (see, Bolfarine and Galea-Rojas, 1996), which is given in the following proposition:

Proposition 1 *Under the structural SMN–MEM defined in (5)-(6), the marginal distribution of $\mathbf{Z}_i$ is given by*

$$f(\mathbf{z}_i|\boldsymbol{\theta}) = \int_0^\infty \phi_p(\mathbf{z}_i|\boldsymbol{\mu}, \kappa(u_i)\boldsymbol{\Sigma})dH(u_i), \tag{9}$$

i.e., $\mathbf{Z}_i \overset{\text{iid}}{\sim} SMN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$, $i = 1, \ldots, n$, *where*

$$\boldsymbol{\mu} = \mathbf{a} + \mathbf{b}\mu_x, \;\; \boldsymbol{\Sigma} = \phi_x \mathbf{b}\mathbf{b}^\top + D(\boldsymbol{\phi}).$$

Proof The proof is direct from (5) by using the stochastic representation given in (3)

The result presented in Proposition 1 facilitates straightforward implementation of inferences with standard optimization routines and existing statistical softwares. The asymptotic covariance matrix of the ML estimators can be estimated by using the Hessian matrix, which can also be computed numerically using, for instance, the *optim* routine in platform R. A disadvantage of direct maximization of the log-likelihood function is that it may not converge unless good starting values are used. Thus, consider using the EM algorithm (Dempster et al, 1977) for parameter estimation, which is quite insensitive to the stating values and considered as a powerful computational tool. One of the major reason for the popularity of the EM algorithm is that the M-step involves only complete data ML estimation, which is often computationally simple. Moreover, the EM algorithm is stable and straightforward to implement since the iterations converge monotonically and no second derivatives are required. When the M-step of EM turns out to be analytically intractable, it can be replaced with a sequence of conditional maximization (CM) steps. Such modification is referred to as the ECM algorithm (Meng and Rubin, 1993). In this article, to obtain the ML estimator of $\boldsymbol{\theta}$ we use the ECM algorithm and derived algebraically its asymptotic covariance matrix.

3.1 The EM–algorithm

In this section, we demonstrate how to use the EM–type algorithm for ML estimation of the SMN–MEM. A key feature of this model is that it can be formulated in a flexible hierarchical representation that is useful for theoretical derivations. From (3) it follows that

$$\mathbf{Z}_i \mid x_i, U_i = u_i \overset{\text{ind}}{\sim} N_p(\mathbf{a} + \mathbf{b}x_i, \kappa(u_i)D(\boldsymbol{\phi})), \tag{10}$$

$$x_i \mid U_i = u_i \overset{\text{ind}}{\sim} N_1(\mu_x, \kappa(u_i)\phi_x), \tag{11}$$

$$U_i \overset{\text{iid}}{\sim} H(u_i; \boldsymbol{\nu}), \tag{12}$$

$i = 1, \ldots, n$, all independent. Let $\mathbf{z} = (\mathbf{z}_1^\top, \ldots, \mathbf{z}_n^\top)^\top$, $\mathbf{x} = (x_1, \ldots, x_n)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$ and $\boldsymbol{\theta}^{(k)} = (\alpha^{(k)}, \beta^{(k)}, \mu_x^{(k)}, \phi_x^{(k)}, \phi_u^{(k)})^\top$, denotes the estimates of $\boldsymbol{\theta}$ at the k–th iteration. It follows from (10)-(12) that the complete log–likelihood function associated with $\mathbf{z}_c = (\mathbf{z}, \mathbf{x}, \mathbf{u})$ is of the form

$$\begin{aligned} \ell_c(\boldsymbol{\theta}|\mathbf{z}_c) = & -n\log(\phi_u) - \frac{1}{2}\sum_{i=1}^{n}\kappa^{-1}(u_i)(\mathbf{z}_i - \mathbf{a} - \mathbf{b}x_i)^\top D^{-1}(\boldsymbol{\phi})(\mathbf{z}_i - \mathbf{a} - \mathbf{b}x_i) \\ & -\frac{n}{2}\log(\phi_x) - \frac{1}{2\phi_x}\sum_{i=1}^{n}\kappa^{-1}(u_i)(x_i - \mu_x)^2 + C, \end{aligned}$$

where C is a constant that is independent of the parameter vector $\boldsymbol{\theta}$. Given the current estimate $\widehat{\boldsymbol{\theta}}^{(k)}$, the E–step calculates $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = E[\ell_c(\boldsymbol{\theta}|\mathbf{z}_c)|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}] = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$, where $Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = Q_{1i}(\alpha, \beta, \phi_u|\widehat{\boldsymbol{\theta}}^{(k)}) + Q_{2i}(\mu_x, \phi_x|\widehat{\boldsymbol{\theta}}^{(k)})$, with

$$\begin{aligned}
Q_{1i}(\alpha, \beta, \phi_u|\widehat{\boldsymbol{\theta}}^{(k)}) &= -\log(\phi_u) - \frac{1}{2}\widehat{u}_i^{(k)}(\mathbf{z}_i - \mathbf{a} - \mathbf{b}\widehat{x}_i^{(k)})^\top D^{-1}(\boldsymbol{\phi})(\mathbf{z}_i - \mathbf{a} - \mathbf{b}\widehat{x}_i^{(k)}) \\
&\quad -\frac{1}{2}\widehat{\Lambda}_x^{(k)}\mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}, \\
Q_{2i}(\mu_x, \phi_x|\widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{1}{2}\log(\phi_x) - \frac{1}{2\phi_x}(\widehat{\Lambda}_x^{(k)} + \widehat{u}_i^{(k)}(\widehat{x}_i^{(k)} - \mu_x)^2),
\end{aligned}$$

and its calculation requires of the expressions $\widehat{u}_i^{(k)} = E[\kappa^{-1}(U_i)|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}_i]$,

$$\begin{aligned}
\widehat{\Lambda}_x^{(k)} &= \widehat{u}_i^{(k)} Var[x_i|u_i, \mathbf{z}, \widehat{\boldsymbol{\theta}}^{(k)}] = \phi_x^{(k)}/(1 + \phi_x^{(k)}\mathbf{b}^{\top(k)}D^{-1}(\boldsymbol{\phi}^{(k)})\mathbf{b}^{(k)}) \\
\widehat{x}_i^{(k)} &= E[x_i|u_i, \mathbf{z}, \widehat{\boldsymbol{\theta}}^{(k)}] = \mu_x^{(k)} + \widehat{\Lambda}_x^{(k)}\mathbf{b}^{\top(k)}D^{-1}(\boldsymbol{\phi}^{(k)})(\mathbf{z}_i - \mathbf{a}^{(k)} - \mathbf{b}^{(k)}\mu_x^{(k)}).
\end{aligned}$$

In each step, the conditional expectation $\widehat{u}_i$ can be easily derived from the result given in Section 2, where we present computationally attractive expression for each distribution considered in this paper.

The CM steps then conditionally maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$, obtaining a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as described below:

$$\begin{aligned}
\widehat{\alpha}^{(k+1)} &= \frac{\sum_{i=1}^n \widehat{u}_i^{(k)}(Y_i - \widehat{\beta}^{(k)}\widehat{x}_i^{(k)})}{\sum_{i=1}^n \widehat{u}_i^{(k)}}; \ \widehat{\beta}^{(k+1)} = \frac{\sum_{i=1}^n (Y_i - \widehat{\alpha}^{(k+1)})\widehat{u_i}^{(k)}\widehat{x_i}^{(k)}}{\sum_{i=1}^n (\widehat{\Lambda}_x^{(k)} + \widehat{u}_i\widehat{x}_i^{2(k)})} \\
\widehat{\mu}_x^{(k+1)} &= \frac{\sum_{i=1}^n \widehat{u_i}^{(k)}\widehat{x_i}^{(k)}}{\sum_{i=1}^n \widehat{u_i}^{(k)}}; \ \widehat{\phi}_x^{(k+1)} = \frac{1}{n}\sum_{i=1}^n \left(\widehat{\Lambda}_x^{(k)} + \widehat{u}_i^{(k)}(\widehat{x}_i^{(k)} - \widehat{\mu}_x^{(k+1)})^2\right); \\
\widehat{\phi}_u^{(k+1)} &= \frac{1}{2n\lambda}\sum_{i=1}^n \left(\widehat{u}_i^{(k)}[\lambda(X_i - \widehat{x}_i^{(k)})^2 + (Y_i - \widehat{\alpha}^{(k+1)} - \widehat{\beta}^{(k+1)}\widehat{x}_i^{(k)})^2] + \widehat{\Lambda}_x^{(k)}(\lambda + \widehat{\beta}^{2(k+1)})\right).
\end{aligned}$$

The iterations of the above algorithm are repeated until a suitable convergence rule is satisfied, e.g., $||\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}||$ is sufficiently small. Note that this algorithm is computationally no expensive and guarantees no negative scale parameter estimates.

3.2 The observed information matrix

The log-likelihood function for $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \phi_x, \phi_u)^\top$, given the observed sample $\mathbf{z} = (\mathbf{z}_1^\top, \ldots, \mathbf{z}_n^\top)^\top$, is of the form

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \tag{13}$$

where $\ell_i(\boldsymbol{\theta}) = -\log 2\pi - \frac{1}{2}\log(|\boldsymbol{\Sigma}|) + \log K_i$, with

$$K_i = K_i(\boldsymbol{\theta}) = \int_0^\infty \kappa^{-1}(u_i)\exp\{-\frac{1}{2}\kappa^{-1}(u_i)d_i\}dH(u_i),$$

and $d_i = (\mathbf{z}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z}_i - \boldsymbol{\mu})$ is the Mahalanobis distance. Using the notation

$$\mathbb{I}_i^\phi(w) = \int_0^\infty \kappa^{-w}(u_i)\exp\{-\frac{1}{2}\kappa^{-1}(u_i)d_i\}dH(u_i),$$

so that, $K_i(\boldsymbol{\theta})$ can be expressed as $K_i(\boldsymbol{\theta}) = \mathbb{I}_i^\Phi(1)$, $i = 1, \ldots, n$, it follows that the matrix of second derivatives with respect to $\boldsymbol{\theta}$ is given by

$$\mathbf{L} = -\sum_{i=1}^{n} \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \frac{1}{2}\sum_{i=1}^{n}\frac{\partial^2 \log(|\boldsymbol{\Sigma}|)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} + \sum_{i=1}^{n}\frac{1}{K_i^2}\frac{\partial K_i}{\partial\boldsymbol{\theta}}\frac{\partial K_i}{\partial\boldsymbol{\theta}^\top} - \sum_{i=1}^{n}\frac{1}{K_i}\frac{\partial^2 K_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}, \tag{14}$$

where

$$\frac{\partial K_i}{\partial\boldsymbol{\theta}} = -\frac{1}{2}\mathbb{I}_i^\phi(2)\frac{\partial d_i}{\partial\boldsymbol{\theta}}, \quad \frac{\partial^2 K_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \frac{1}{4}\mathbb{I}_i^\phi(3)\frac{\partial d_i}{\partial\boldsymbol{\theta}}\frac{\partial d_i}{\partial\boldsymbol{\theta}^\top} - \frac{1}{2}I_i^\phi(2)\frac{\partial^2 d_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$$

For each distribution considered in this work, the integrates above can be written as

- *Skew–t.*

$$\mathbb{I}_i^\phi(w) = \frac{2^w \nu^{\nu/2}\Gamma(w+\nu/2)}{\Gamma(\nu/2)(\nu+d_i)^{\nu/2+w}}$$

- *Skew–slash.*

$$\mathbb{I}_i^\phi(w) = \frac{\nu 2^{w+\nu}\Gamma(w+\nu)}{d_i^{w+\nu}}P_1(w+\nu, \frac{d_i}{2})$$

- *Skew–contaminated normal.*

$$\mathbb{I}_i^\phi(w) = \sqrt{2\pi}\{\nu\gamma^{w-1/2}\phi_1(\sqrt{d_i}|0, \frac{1}{\gamma}) + (1-\nu)\phi_1(\sqrt{d_i}|0,1)\}.$$

3.3 The expected information matrix

In this section we find the expected information matrix of the parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \phi_x, \phi_u)^\top$. Adopting the parameterization $\boldsymbol{\theta}^* = (\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})^\top$, where $\mu_1 = \mu_x$, $\mu_2 = \alpha + \beta\mu_x$, $\sigma_{11} = \phi_x$, $\sigma_{12} = \beta\phi_x$ and $\sigma_{22} = \beta^2\phi_x + \phi_u$, it follows from Osorio et al (2007) that the expected information matrix of $\boldsymbol{\theta}^* = (\boldsymbol{\mu}, \boldsymbol{\sigma})$, where $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and $\boldsymbol{\sigma} = (\sigma_{11}, \sigma_{12}, \sigma_{22})$, assumes the following block diagonal form:

$$I(\boldsymbol{\theta}^*) = \begin{pmatrix} I(\boldsymbol{\mu}) & \mathbf{0} \\ \mathbf{0} & I(\boldsymbol{\sigma}) \end{pmatrix},$$

where $I(\boldsymbol{\mu}) = \boldsymbol{\Sigma}^{-1} \sum_{i=1}^{n} 2d_{Hi}$ and $I(\boldsymbol{\sigma}) = \sum_{i=1}^{n} I_i(\boldsymbol{\sigma})$.

The (r, s)th element of $I_i(\boldsymbol{\sigma})$ is given by

$$I_{i,rs}(\boldsymbol{\sigma}) = \frac{b_{i,rs}}{4}\left(\frac{f_{Hi}}{2} - 1\right) + \frac{f_{Hi}}{4} tr\left\{\boldsymbol{\Sigma}^{-1}\,\dot{\boldsymbol{\Sigma}}\,(r)\boldsymbol{\Sigma}^{-1}\,\dot{\boldsymbol{\Sigma}}\,(s)\right\},$$

where $d_{Hi} = E\left[W^2(D_i)D_i\right]$, $f_{Hi} = E\left[W^2(D_i)D_i^2\right]$ with $D_i = ||\mathbf{X}_i||, \mathbf{X}_i \sim SMN_2(\mathbf{0}, \mathbf{I}_2; H)$, $W(D_i) = -\frac{1}{2}E[\kappa^{-1}(U)|\mathbf{y}] = -\frac{q(D_i)}{2}$ and $b_{i,rs} = tr\left\{\boldsymbol{\Sigma}^{-1}\,\dot{\boldsymbol{\Sigma}}\,(r)\right\} tr\left\{\boldsymbol{\Sigma}^{-1}\,\dot{\boldsymbol{\Sigma}}\,(s)\right\}$ with $\dot{\boldsymbol{\Sigma}}\,(j) = \dfrac{\partial\boldsymbol{\Sigma}}{\partial\boldsymbol{\sigma}_j}$, $j = 1, 2, 3$. For some cases of SMN distributions it is possible to get the functions d_{Hi} and f_{Hi} explicitly, for other distributions we can get they using Monte Carlo approximations. It is possible to obtain closed-form expressions for d_{Hi} and f_{Hi} for some SMN distributions. In particular, for the Student-t distribution with ν degrees of freedom one has $d_{Hi} = \frac{\nu+2}{2(\nu+4)}$ and $f_{Hi} = \frac{2(\nu+2)}{\nu+4}$, and for the power exponential with shape parameter ν we find $d_{Hi} = \frac{\nu^2}{2^{1/\nu}\Gamma(1/\nu)}$, where $\Gamma(.)$ denotes the gamma function, and $f_{Hi} = 1 + \nu$.

Turning back to our parameterizations, we find the Fisher information matrix for $\boldsymbol{\theta}$,

$$I(\boldsymbol{\theta}) = J(\boldsymbol{\theta}^*)I(\boldsymbol{\theta}^*)J(\boldsymbol{\theta}^*)^\top,$$

with the Jacobian given by

$$J(\boldsymbol{\theta}^*) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & \mu_1 & 0 & \gamma & 2\sigma_{12} \\ 1 & \frac{\sigma_{12}}{\gamma} & 0 & 0 & 0 \\ 0 & 0 & 1 & \frac{\sigma_{11}}{\gamma} & \frac{(\sigma_{12})^2}{\gamma^2} \\ 0 & 0 & 1 & 0 & \lambda^{-1} \end{pmatrix},$$

where $\gamma = -\frac{1}{2}(\lambda\sigma_{22} - \sigma_{11}) + \sqrt{(\frac{1}{4}(\lambda\sigma_{22} - \sigma_{11})^2 - \lambda(\sigma_{12})^2)}$. Confidence regions for the parameters can be constructed from asymptotic results. Thus, if $\mathbf{J} = -\mathbf{L}$ denotes the expected information matrix for the log-likelihood $\ell(\boldsymbol{\theta})$, then under some regularity conditions, asymptotic confidence intervals and hypotheses tests for the parameter $\boldsymbol{\theta} \in \mathbb{R}^q$ are obtained assuming that the MLE $\widehat{\boldsymbol{\theta}}$ has approximately a $N_q(\boldsymbol{\theta}, \mathbf{J}^{-1})$ distribution. In practice, $\mathbf{J}$ is usually unknown and has to be replaced by the MLE $\widehat{\mathbf{J}}$, that is, the matrix $\widehat{\mathbf{J}}$ evaluated at the ML estimate $\hat{\boldsymbol{\theta}}$.

4 The local influence approach

The aim of local influence Cook (1986) is to investigate the behavior of some influence measure $T(\boldsymbol{\omega})$ when small perturbations are made into the model/data, where $\boldsymbol{\omega}$ is a g-dimensional vector of perturbations restricted to some open subset $\boldsymbol{\Omega} \in \mathbb{R}^g$. In this work we assess local influence by using an appropriate measure based on the log-likelihood function and particularly recommended for incomplete data.

Let $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c)$, $\boldsymbol{\theta} \in \mathbb{R}^h$ be the complete-data log-likelihood of the perturbed model. We assume that there is a $\boldsymbol{\omega}_0$ such that $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{Y}_c) = \ell_c(\boldsymbol{\theta}|\mathbf{Y}_c)$ for all $\boldsymbol{\theta}$. Let $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ denote the maximum of the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}]$. The influence graph is defined as $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$, where $f_Q(\boldsymbol{\omega})$ is the Q-displacement function defined as

$$f_Q(\boldsymbol{\omega}) = 2\left[Q\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) - Q\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\widehat{\boldsymbol{\theta}}\right)\right].$$

Following the approach developed in Cook (1986) and Zhu and Lee (2001), the normal curvature $C_{f_Q,\mathbf{d}}$ of $\boldsymbol{\alpha}(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ in the direction of some unit vector $\mathbf{d}$ can be used to summarize the local behavior of the Q-displacement function. It can be shown (see, Zhu and Lee 2001) that

$$C_{f_Q,\mathbf{d}} = -2\mathbf{d}^\top \ddot{Q}_{\boldsymbol{\omega}_o}\mathbf{d} \;\text{ and }\; -\ddot{Q}_{\boldsymbol{\omega}_0} = \Delta^\top_{\boldsymbol{\omega}_o}\left\{-\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}})\right\}^{-1}\Delta_{\boldsymbol{\omega}_0},$$

where $\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) = \dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ and $\Delta_{\boldsymbol{\omega}} = \dfrac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}$.

Under mild regularity conditions (Zhu and Lee, 2001), $-\ddot{Q}_{\boldsymbol{\omega}_0}$ is semi-positive definite. Thus, from the spectral decomposition of a symmetric matrix,

$$-2\ddot{Q}_{\boldsymbol{\omega}_o} = \sum_{k=1}^{g}\lambda_k\mathbf{e}_k\mathbf{e}_k',$$

where $(\lambda_1,\mathbf{e}_1),\ldots,(\lambda_g,\mathbf{e}_g)$ are the eigenvalue-eigenvector pairs for the matrix $-2\ddot{Q}_{\boldsymbol{\omega}_o}$ with $\lambda_1 \geq \ldots \geq \lambda_h, \lambda_{h+1} = \ldots = \lambda_g = 0$ and $\mathbf{e}_1,\ldots,\mathbf{e}_g$ are the elements of the associated orthonormal basis. On the basis of results given in Cook (1986), $C_{f_Q,\mathbf{d}} = \lambda_1$, the largest eigenvalue. Thus, the eigenvector $\mathbf{e}_1 = \mathbf{d}_{max}$ corresponding to λ_1 gives important information for assessing local influence of a minor perturbation. When g is larger, it is computationally expensive to obtain the eigenvalues-eigenvectors $\{(\lambda_k,\mathbf{e}_k), k = 1,\ldots,g\}$. Moreover, as λ_1 can take any positive value, sometimes it is not easy to judge its size. In order to reduce the computational burden and to aid interpretation, Zhu and Lee (2001) consider an aggregated contribution vector of all those eigenvectors with nonzero eigenvalues. Introducing some further notation, let $\widetilde{\lambda}_k = \lambda_k/(\lambda_1+\ldots+\lambda_h)$, $\mathbf{e}_k^2 = (e_{k1}^2,\ldots,e_{kg}^2)$ and

$$M(0) = \sum_{k=1}^{h}\widetilde{\lambda}_k\mathbf{e}_k^2,$$

with the lth component of $M(0)$, denoted by $M(0)_l$, equal to $\sum_{k=1}^{h} \widetilde{\lambda}_k \mathbf{e}_{kl}^2$. Assessment of influential cases is based on the visual inspection of the $\{M(0)_l, l = 1, \ldots, g\}$ plotted against the index l. Inspired by the work of Poon and Poon (1999) in using a conformal normal curvature to modify Cook's (1986) normal curvature, one can obtain $M(0)_l$ via $B_{f_Q,\mathbf{u}_l} = -2\mathbf{u}_l^\top \ddot{Q}_{\boldsymbol{\omega}_0} \mathbf{u}_l / \mathrm{tr}[-2\ddot{Q}_{\boldsymbol{\omega}_0}]$, where $\mathbf{u}_l$ is a column vector in $\mathbb{R}^g$ with the lth entry equal to one and all other entries zero. Note that the evaluation of $B_{f_Q,\mathbf{u}_l}$ is very simple, and it is not necessary to compute the eigenvalues-eigenvectors of $\ddot{Q}_{\boldsymbol{\omega}_0}$. We refer the reader to Zhu and Lee (2001) for other theoretical properties of $B_{f_Q,\mathbf{u}_l}$, such as its invariance under reparametrizations of $\boldsymbol{\theta}$.

Let $\bar{M}(0)$ and $SM(0)$ be the mean and standard error of $\{M(0)_l, l = 1, \ldots, g\}$. Clearly, $\bar{M}(0) = 1/m$. It is natural to regard observations with $M(0)_l$-values that are significantly larger than $\bar{M}(0)$ as influential. This motivates the use of $\bar{M}(0)$ in defining a benchmark. However, one can clearly use different functions of $\bar{M}(0)$. For example, in the context of the approach Cook (1986), Poon and Poon (1999) proposed the use of $2\bar{M}(0)$, similar to that which is often employed for identifying leverage in regression analysis. In order to take into account the variation in $M(0)_l$, Zhu and Lee (2001) proposed the use of $\bar{M}(0) + 2SM(0)$, and Lee and Xu (2004) the use of $\bar{M}(0)+c^*SM(0)$, where c^* is a selected constant taken to be 2 or larger. The lth case is then regarded as influential if $M(0)_l$ is larger than the benchmark. So far in influence analysis, there are no general rules for selecting and benchmark. Thus, we cannot regard $2\bar{M}(0)$ as necessarily being a better benchmark than $\bar{M}(0) + c^*SM(0)$, or vice versa. Also, there are no clear guidelines for selecting c^*. Nevertheless, we recommend starting any analysis using $c^* = 2$ and, if overly many cases are identified as being influential, increase the value to $c^* = 4$, say. This strategy appeared to work well in the empirical studies that we have conducted.

4.1 The Hessian matrix, $\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})$

In order to obtain the diagnostic measures for local influence of a particular perturbation scheme, it is necessary to compute

$$\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) = -\frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \begin{pmatrix} \ddot{Q}_1(\alpha,\beta,\phi_u) & \mathbf{0} \\ \mathbf{0} & \ddot{Q}_2(\mu_x,\phi_x) \end{pmatrix},$$

where the matrices $\ddot{Q}_1(\alpha,\beta,\phi_u) = \left[-\ddot{Q}_{1,\tau\pi}\right]$, $\ddot{Q}_{1,\tau\pi} = \dfrac{\partial^2 Q_1(\alpha,\beta,\phi_u|\widehat{\boldsymbol{\theta}})}{\partial\tau\partial\pi}$, with $\tau,\pi = \alpha,\beta$ or ϕ_u, $\ddot{Q}_2(\mu_x,\phi_x) = \left[-\ddot{Q}_{2,\tau\pi}\right]$, $\ddot{Q}_{2,\tau\pi} = \dfrac{\partial^2 Q_2(\mu_x,\phi_x|\widehat{\boldsymbol{\theta}})}{\partial\tau\partial\pi}$, with τ, $\pi = \mu_x$ or ϕ_x. Hence, the Hessian matrix have elements given by

$$\ddot{Q}_{1,\alpha\alpha} = -\frac{1}{\lambda\phi_u}\sum_{i=1}^n \widehat{u}_i,\;\; \ddot{Q}_{1,\alpha\beta} = -\frac{1}{\lambda\phi_u}\sum_{i=1}^n \widehat{u}_i\widehat{x}_i,\;\; \ddot{Q}_{1,\alpha\phi_u} = -\frac{1}{\lambda\phi_u^2}\sum_{i=1}^n \widehat{u}_i(Y_i-\alpha-\beta\widehat{x}_i),$$

$$\ddot{Q}_{1,\beta\beta} = -\frac{1}{\lambda\phi_u}\sum_{i=1}^n(\widehat{\Lambda}_x + \widehat{x}_i^2\widehat{u}_i),\;\; \ddot{Q}_{1,\beta\phi_u} = -\frac{1}{\lambda\phi_u^2}\sum_{i=1}^n((Y_i-\alpha-\beta\widehat{x}_i)\widehat{x}_i\widehat{u}_i - \beta\widehat{\Lambda}_x),$$

$$\ddot{Q}_{1,\phi_u\phi_u} = \frac{n}{\phi_u^2}(1-\frac{\widehat{\Lambda}_x}{\lambda\phi_u}(\lambda+\beta^2)) - \frac{1}{\phi_u^3\lambda}\sum_{i=1}^n \widehat{u}_i[\lambda(X_i-\widehat{x}_i)^2 + (Y_i-\alpha-\beta\widehat{x}_i)^2],$$

$$\ddot{Q}_{2,\mu_x\mu_x} = -\frac{1}{\phi_x}\sum_{i=1}^n \widehat{u}_i,\;\; \ddot{Q}_{2,\mu_x\phi_x} = -\frac{1}{\phi_x^2}\sum_{i=1}^n \widehat{u}_i(\widehat{x}_i-\mu_x),$$

$$\ddot{Q}_{2,\phi_x\phi_x} = \frac{n}{2\phi_x^2} - \frac{1}{2\phi_u^3}\sum_{i=1}^n(\widehat{u}_i(x_i-\mu_x)^2 + \widehat{\Lambda}_x).$$

4.2 Interesting perturbation schemes

In this section we consider three different perturbation schemes for the baseline model defined in (5) and (6).

Case-weight perturbation

First, we consider an arbitrary attribution of weights for the expected value of the complete–data log-likelihood function (perturbed Q-function), which may capture observations with outstanding contribution on the log-likelihood function and that may exercise high influence on the maximum likelihood estimates, represented by writing

$$Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_c(\boldsymbol{\theta},\boldsymbol{\omega}|\mathbf{y}_c)] = \sum_{i=1}^{n}\omega_i \mathrm{E}[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)] = \sum_{i=1}^{n}\omega_i Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$$

where $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_n)^\top$ is an $n\times 1$ vector with $\boldsymbol{\omega}_0 = (1,\ldots,1)^\top$. In this case the matrix $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0} = \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}$ has elements given by

$$\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\alpha\partial\omega_i} = \frac{1}{\lambda\phi_u}\widehat{u}_i(Y_i-\alpha-\beta\widehat{x}_i),\ \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\beta\partial\omega_i} = \frac{1}{\lambda\phi_u}(\widehat{u}_i(Y_i-\alpha-\beta\widehat{x}_i)\widehat{x}_i-\beta\widehat{\Lambda}_x)$$
$$\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\mu_x\partial\omega_i} = \frac{1}{\phi_x}\widehat{u}_i(\widehat{x}_i-\mu_x),\ \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\phi_x\partial\omega_i} = -\frac{1}{2\phi_x}+\frac{1}{2\phi_x^2}(\Lambda_x+\widehat{u}_i(\widehat{x}_i-\mu_x)^2)$$
$$\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\phi_u\partial\omega_i} = -\frac{1}{\phi_u}+\frac{1}{2\phi_u^2\lambda}[\widehat{u}_i(\lambda(X_i-\widehat{x}_i)^2+(Y_i-\alpha-\beta\widehat{x}_i)^2)+\widehat{\Lambda}_x(\lambda+\beta^2)].$$

Response variable perturbation

A additive perturbation of the response variables $(Y_1,\ldots,Y_n)^\top$, which may indicate observations with large influence on their own predicted values, is introduced by replacing Y_i by $Y_{i\omega}^* = Y_i+\omega_i$. In this case, the matrix $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$, with $\boldsymbol{\omega}_o = (0,...,0)^\top$, has elements given by

$$\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\alpha\partial\omega_i} = \frac{\widehat{u}_i}{\lambda\phi_u},\ \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\beta\partial\omega_i} = \frac{\widehat{u}_i\widehat{x}_i}{\lambda\phi_u},\ \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\mu_x\partial\omega_i} = \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\phi_x\partial\omega_i} = 0,$$
$$\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\phi_u\partial\omega_i} = \frac{1}{\lambda\phi_u^2}(Y_i-\alpha)\widehat{u}_i.$$

Variances ratio perturbation

A perturbation that may indicate if the ratio between the variances $\lambda = \phi_u/\phi_e$ is adequate is represented by writing $\lambda^* = \lambda\omega_i$. Under this perturbation scheme, the non-perturbed model is obtained when $\boldsymbol{\omega}_0 = (1,\ldots,1)^\top$. The matrix $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ has elements given by

$$\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\alpha\partial\omega_i} = -\frac{\widehat{u}_i}{\lambda\phi_u}(Y_i - \alpha - \beta\widehat{x}_i), \ \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\beta\partial\omega_i} = \frac{1}{\lambda\phi_u}(\beta\widehat{\Lambda}_x - \widehat{u}_i\widehat{x}_i(Y_i - \alpha - \beta\widehat{x}_i))$$
$$\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\mu_x\partial\omega_i} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\phi_x\partial\omega_i} = 0,$$
$$\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\phi_u\partial\omega_i} = -\frac{1}{2\lambda\phi_u^2}(\widehat{u}_i(Y_i - \alpha - \beta x_i)^2 + \beta^2\Lambda_x).$$

5 Application

In this section we consider a likelihood analysis of a part of the AIS data set (available for download at ($http://stat.umidp.it/SN/index.html$) considering a linear measurement error model relating SSF and $bfat$. We define the model

$$SSF_i = \alpha + \beta bfat_i + e_i,$$

$i = 1, \ldots, 202$, where $bfat_i$ is the body fat percentage of the i-th individual in the sample, SSF_i is the sum of skin folds. We assume that $bfat$ is measured with error according to the equation

$$Bfat_i = bfta_i + u_i,$$

with $Bfat_i$ being an unbiased estimate of the true (unobserved) $bfat_i$, $i = 1, \ldots, 202$. Thus, with the aim of providing robust estimation and inferences, we consider the following marginal formulation:

$$\mathbf{Z}_i = (Bfat_i, SSF_i)^\top \sim SMN_2(\mathbf{a} + \mathbf{b}\mu_x, \boldsymbol{\Sigma}; H), \tag{15}$$

where H denote the distribution function for the mixture variable U_i, $i = 1, \ldots, 202$. In our analysis we will assume Normal (N), Student-t (T), slash (SL) and contaminated normal (CN) distributions from the SMN class for comparative purposes. We set $\nu = 8$ for T–MEM, $\nu = 3$ for SL–MEM and $\boldsymbol{\nu} = (0.1, 0.2)^\top$ for CN–MEM, such values were chosen in order to try accommodating the outlying observations that appear under normal errors. All our results have been obtained using the R software, available upon request from the third author.

Table 1 contains the ML estimates, for the parameters of the four models, viz, N–MEM, T–MEM, SL-MEM and CN-MEM, together with their corresponding standard errors calculated via the observed information matrix. The log-likelihood values (see row $\ell(\widehat{\boldsymbol{\theta}})$) indicate that the SMN distributions with heavy tails presents the best fit than the N-MEM model with the CN-MEM one significantly better. We see that the greatest variation is in the parameter α and that the estimations for the scale parameters are not comparable because there are in different scales. Figure 1 indicate clear attenuation in β due to the use of MSN distribution with heavy tails. It is possible to observe in this figure some atypical individuals that could have an influence on the ML estimators. In this graphics, the individuals 11, 18, 37, 53, 56 and 178 were marked since they were detected as potentially influential by the local influence approach.

Next we conduct a local influence study based on $M(0)$ with interest focussing

	N–MEM		T–MEM		SL–MEM		CN–MEM	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
α	-4.556	1.581	-1.903	1.533	-2.582	1.587	-1.495	1.7812
β	5.447	0.107	5.238	0.113	5.283	0.114	5.186	0.138
μ_x	13.507	0.434	12.702	0.432	13.086	0.422	12.965	0.492
ϕ_x	35.469	3.786	28.566	3.399	23.127	2.625	27.440	5.519
ϕ_u	2.677	0.266	2.141	0.244	1.792	0.191	2.131	0.540
$\ell(\widehat{\boldsymbol{\theta}})$	-1379.188		-1375.205		-1373.732		-1371.022	

Table 1 Results from fitting the four SMN distributions to th AIS data set. The SE values are estimated asymptotic standard errors.

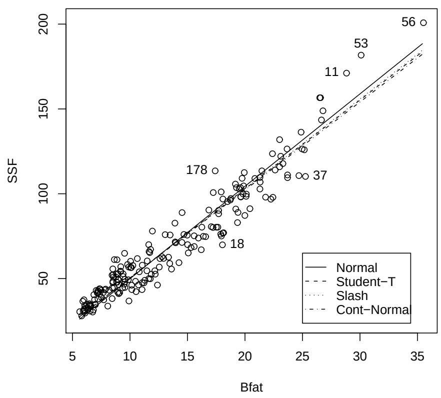

Fig. 1 Fitted regression lines using "AIS" data for the four SMN distributions.

on $\boldsymbol{\theta}$ for the AIS data set. In a preliminary analysis we used $c^* = 3$ when calculating the benchmark. However, far too many observations were identified as being influential for this choice of c^*. Thus, here we present a local influence analysis using $c^* = 4$ to compute the benchmark. Figures 2, 3 and 4 present the index graphs of $M(0)$ for the four selected models.

Case-weight perturbation. This perturbation can be used for detecting observations with outstanding contribution on the log-likelihood function and that may exercise high influence on the ML estimates. As can be observed in Figure 2 cases 18, 37, 53, 56, and 178 seem to be the most influential in the ML estimators under N-MEM, while only cases 18 and 178 seem to be the most influential in the ML estimators under T-MEM and CN–MEM, as depicted in Figures 2b and 2d. In addition, note that the ML estimators are quite stable under SL–MEM, as displayed in Figure 2c. As expected, the influence of atypical points is reduced when we consider distributions with heavier tails than the normal ones. For this data set the slash model with small ν accommodates slightly better the influential observations.

Responde variable perturbation. This perturbation can be used for detecting observations with large influence on their own predicted values. In Figure 2 is appreciated some influence when the measurements of cases 53 and 56 are perturbed under N-MEM. This influence is clearly reduced when we use distributions with heavier tails than the normal ones.

Ratio variances perturbation. This perturbation can be used for detecting observations with large influence on the homoscedastic assumption. Figure 4 presents the conformal curvatures for the four fitted models. We can notice under N-MEM that cases 11, 53 and 56 are the most influential on $\boldsymbol{\theta}$, while under SL-MEM one observation (178) is pointed out with some (slightly) influence in the ML estimators, as depicted in Figure 2c. In the other models, however, influential cases are not observed. Once again, the influence is reduced when we use distributions with heavier tails than the normal ones.

The local influence analysis has detected as the most influential the following six cases: 11, 18, 37, 53, 56 and 178. In order to reveal the impact of these six observations on the parameters, we refitted the models with the totality of these potentially influential observations removed to obtain the ML estimate $\widehat{\boldsymbol{\theta}}^0$. As in Lu and Song (2006), the following two quantities are used to measure the difference between the original ML estimate, $\widehat{\boldsymbol{\theta}}$, and $\widehat{\boldsymbol{\theta}}^0$.

$$TRC = \sum_{j=1}^{n_p} |\frac{\widehat{\boldsymbol{\theta}}_j - \widehat{\boldsymbol{\theta}}_j^0}{\widehat{\boldsymbol{\theta}}_j}|, \quad MRC = \max_{j=1,\ldots,n_p} |\frac{\widehat{\boldsymbol{\theta}}_j - \widehat{\boldsymbol{\theta}}_j^0}{\widehat{\boldsymbol{\theta}}_j}|, \tag{16}$$

where n_p is the number of parameters. The results are summarized in Table 2. From this table we observe that the greater changes take place under the normal distribution. As expected, the results indicated that the ML estimators are less sensitives in the presence of atypical data when we used distributions with heavier tails.

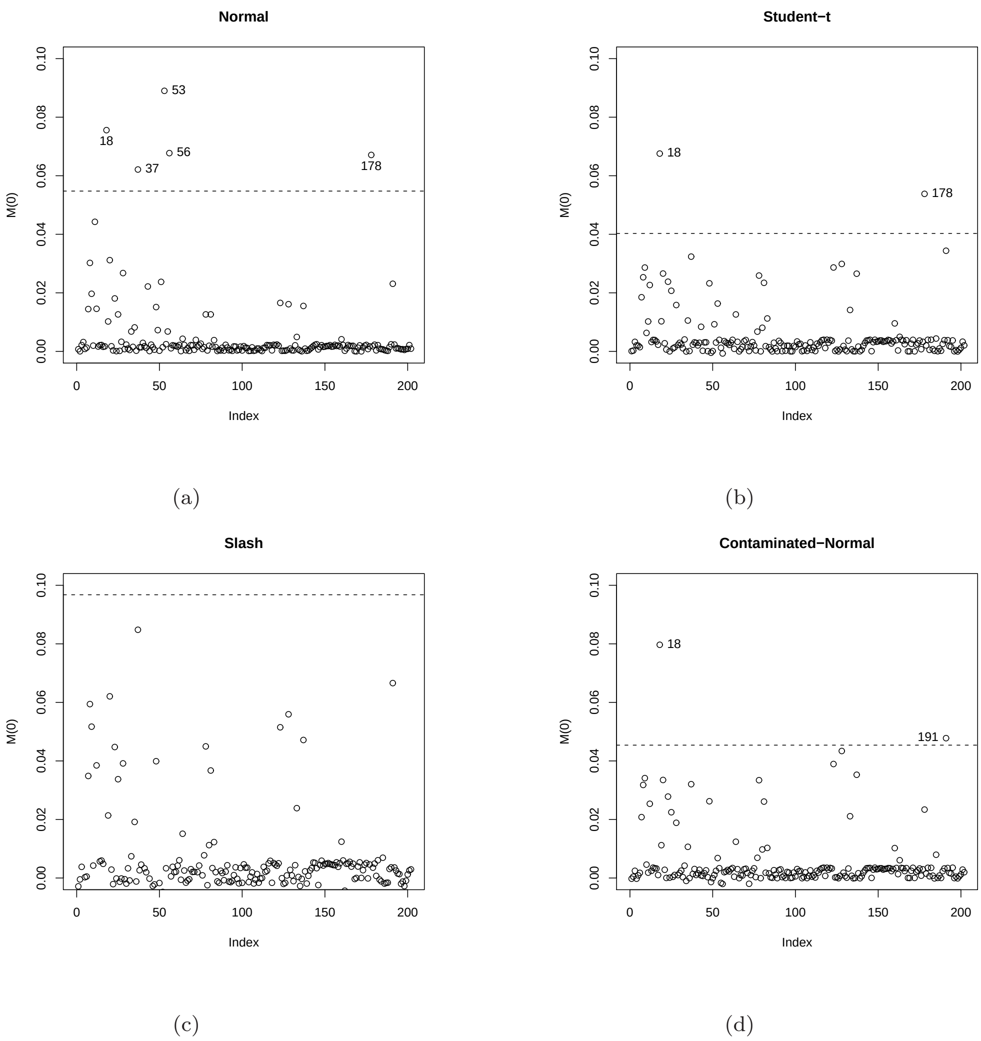

Fig. 2 "AIS" data set. Index plots of $M(0)$ under case-weight perturbation for the four fitted models. The horizontal lines delimit the Lee and Xu (2004) benchmark for $M(0)$ with $c^* = 4$.

6 Final Conclusion

In this work we have developed a new class of SMN–MEM, with the N–MEM as a special case. This class of distributions have additional parameters that can be used for adjusting the distributions, kurtosis and provide more robust procedures than the

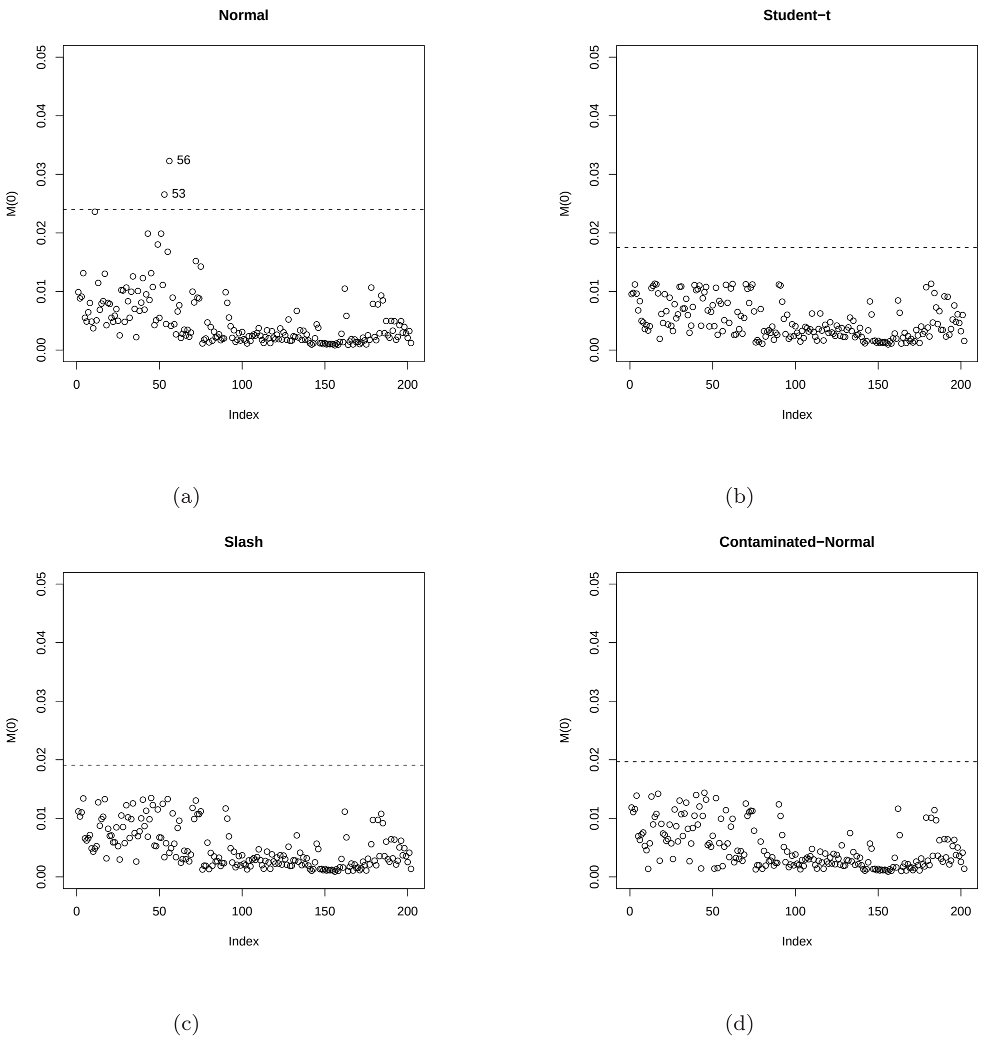

Fig. 3 "AIS" data set. Index plots of $M(0)$ under response perturbation for the four fitted models. The horizontal lines delimit the Lee and Xu (2004) benchmark for $M(0)$ with $c^* = 4$.

ones that use the normal distribution, with moderate additional computational effort. Closed form expressions are obtained for the observed information matrix, for the hessian matrix $\ddot{\mathbf{Q}}$ and for the $\boldsymbol{\Delta}$ matrix for the three perturbation schemes. Through a local influence study some aspects of robustness of the maximum likelihood estimators under the scale mixture of normal distributions with heavy-tails were note. The results

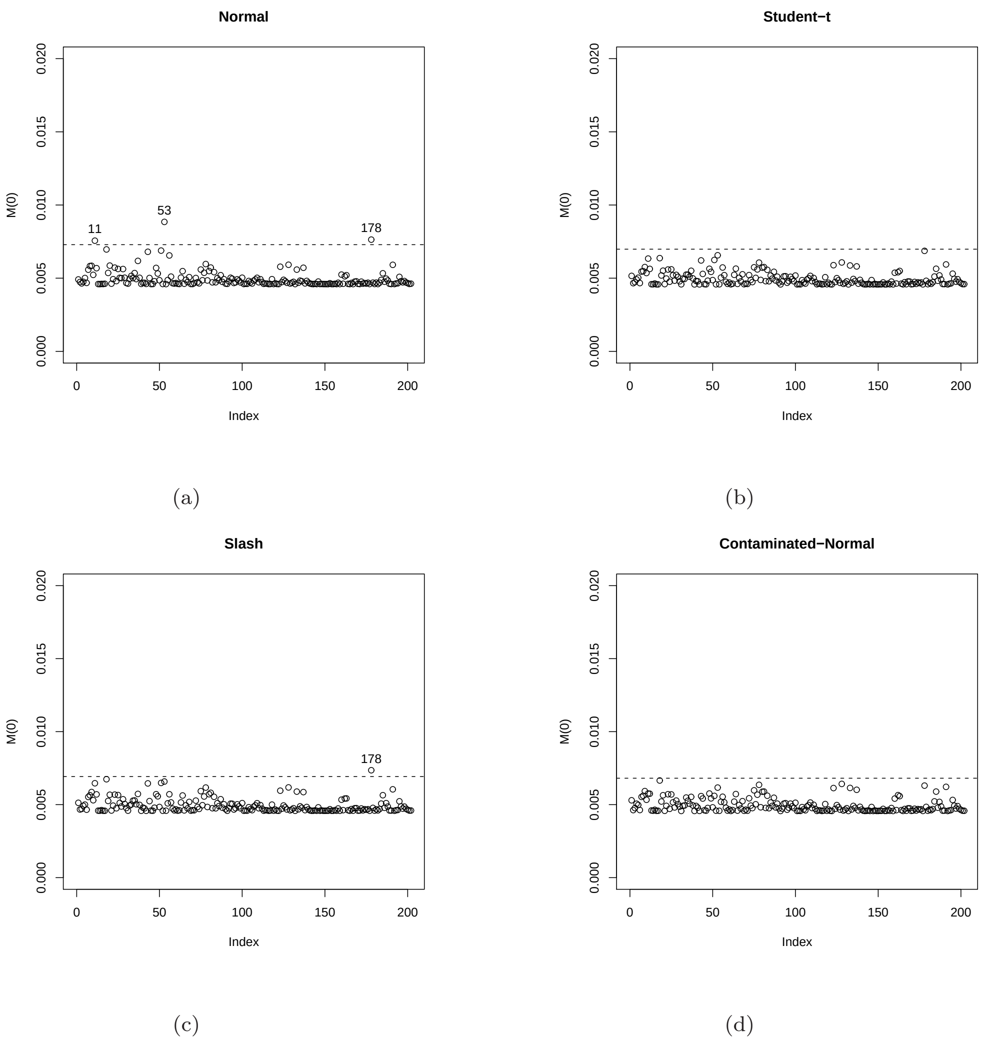

Fig. 4 "AIS" data set. Index plots of $M(0)$ under ratio variances perturbation for the four fitted models. The horizontal lines delimit the Lee and Xu (2004) benchmark for $M(0)$ with $c^* = 4$.

presented in this work represent an extension of the work by Galea-Rojas et al (2002). The results derived in this work agree with the considerations that in this respect are presented in Osorio et al (2007) as well as with the comments of Pinheiro et al (2001).

Table 2 "AIS" data. Comparison of the relative changes in the ML estimators in term of TRC and MRC for the four selected SMN models.

	TRC	MRC
N–MEM	0.312	0.132
T–MEM	0.215	0.108
SL–MEM	0.212	0.111
CN–MEM	0.150	0.081

References

Andrews DF, Mallows SL (1974) Scale mixtures of normal distributions. Journal of the Royal Statistical Society, Series B 36:99–102

Aoki R, Achcar JA, Bolfarine H, Singer JM (2003) Bayesian analysis of null intercept errors-in-variables regression for pretest/post-test data. Journal of Applied Statistics 30:5–14

Arellano-Valle RB, Bolfarine H, Lachos VH (2005) Skew-normal linear mixed models. Journal of Data Science 3:415–438

Barnett VD (1969) Simultaneous pairwise linear structural relationships. Biometrics 25:129–142

Bolfarine H, Galea-Rojas M (1996) On structural comparative calibration under a t-model. Computational Statistics 11:63–85

Chan LK, Mak TK (1979) On the maximun likelihood estimation of a linear structural relationship when the intercept is known. Journal of Multivariate Analyis 9:304–313

Cheng CL, Van-Ness JW (1999) Statistical regression with measurement error. Arnold, London

Chipkevitch E, Nishimura R, Tu D, Galea-Rojas M (1996) Clinical measurement of testicular volume in adolescents: Comparison of the reliability of 5 methods. Journal of Urology 156:2050–2053

Cook RD (1986) Assessment of local influence (with discussion). Journal of the Royal Statistical Society, Series B 48:133–169

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society, Series B 39:1–22

Dunn G (1992) Design and Analysis of Reliability:The statistical evaluation of measurement errors. Edward Arnold, New York

Fuller WA (1987) Measurement error models. Wiley, New York

Galea-Rojas M, Bolfarine H, Vilca LF (2002) Influence diagnostics for the structural errors-in-variables model under the student-t distribution. Journal of Applied Statistics 29:1191–1204

Galea-Rojas M, Bolfarine H, Vilca LF (2005) Local influence in comparative calibration models under elliptical t-distribution. Biometrical Journal 47:691–706

Grubbs FE (1973) Errors of measurement, precision, accuracy, and the statistical comparison of measuring instruments. Technometrics 15:53–66

Kelly G (1984) The influence function in the errors in variables problem. Annals of Statistics 12:87–100

Kendall MG, Stuart A (1973) The Advanced Theory of Statistics, vol 1 and 2, 4th edn. Griffin, London

Lange KL, Sinsheimer JS (1993) Normal/independent distributions and their applications in robust regression. J Comput Graph Stat 2:175–198

Lange KL, Little R, Taylor J (1989) Robust statistical modeling using t distribution. Journal of the American Statistical Association 84:881–896

Lee SY, Tang NS (2004) Local influence analysis of nonlinear structural equation models. Psycometrika 69:573–592

Lee SY, Xu L (2004) Influence analysis of nonlinear mixed-effects models. Computational Statistics and Data Analysis 45:321–341

Lindley DV (1972) Bayesian Statistics. SIAM, Philadelphia, PA

Lu Y, Ye K, Mathur A, Hui S, Fuerst T, Genant H (1997) Comparative calibration without a gold standard. Statistics in Medicine 16:1889–1905

Meng XL, Rubin DB (1993) Maximum likelihood estimation via the ecm algorithm: A general framework. Biometrika 80:267–278

Osorio F, Paula GA, Galea-Rojas M (2007) Assessment of local influence in elliptical linear models with longitudinal structure. Computational Statistics and Data Analysis 51:4354–4368

Osorio FS (2006) Diagnstico de influncia em modelos elpticos com efeitos mistos. Unpublished Ph. D. thesis (in portuguese), Department od Statistics, IME–USP, Brazil

Patefield WM (1985) Information from the maximized likelihood function. Biometrics 72:664–668

Pinheiro JC, Liu CH, Wu YN (2001) Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. Journal of Computational and Graphical Statistics 10:249–276

Poon W, Poon YS (1999) Conformal normal curvature and assessment of local influence. Journal of the Royal Statistical Society, Series B 61:51–61

Shyr I, Gleser L (1986) Inference about comparative precision in linear structural relationships. Journal of Statistical Planning and Inference 14:339–358

Theobald CM, Mallison JR (1978) Comparative calibration, linear structural relationship and congeneric measurements. Biometrics 34:35–45

West M (1987) On scale mixtures of normal distributions. Biometrika 74:646–648

Zhu H, Lee S (2001) Local inffluence for incomplete-data models. Journal of the Royal Statistical Society, Series B 63:121–126