RT-MAE 2005-11

Estimation and influence diagnostics for structural comparative calibration models under the skew-normal distributions

by

Victor H. Lachos, Heleno Bolfarine
Filidor Vilca
and
Manuel Galea

Palavras-Chave: Maximum likelihood, local influence, EM – algorithm, skewness.
Classificação AMS: 62F05, 62J12.

- Março de 2005 -

Estimation and Influence Diagnostics for Structural Comparative Calibration Models under the Skew-Normal Distributions

Víctor H. Lachos[1]
Bayes Forecast, Brasil
Heleno Bolfarine[2]
Departamento de Estatística, Universidade de São Paulo, Brasil
Filidor Vilca[3]
Departamento de Estatística, Universidade Estadual de Campinas, Brasil
and
Manuel Galea[4]
Departamento de Estadística, Universidad de Valparaíso, Chile

Abstract

In this paper we extent the usual normal (symmetric) structural comparative calibration model by considering that the latent variable follows a skew-normal distribution which includes the normal ones as special case and provides robust estimation in this type of models. The marginal likelihood function is obtained which is expressed in closed form, so that inference may be carried out using existing statistical software and standard optimization techniques. In order to get reliability in the estimation process we also implement an EM-type algorithm by exploring statistical properties of the model considered. Additionally, we derive the appropriate matrices to assessing the local influence on the parameters estimates under different perturbation schemes. Applications of the results and methods developed in the paper are illustrated with examples using two real data sets previously analyzed in the literature.

Key Words: *Skew-normal distribution, Local influence, EM algorithm, Skewness.*

1 Introduction

Recent statistical literature has seen an increasing interest for models that provide flexibility in capturing a broad range of non-normal behavior and thus, represent features of the data as adequately as possible and to reduce unrealistic assumptions. Advantages of using such general structures include easiness of interpretation, as well as estimation efficiency.

Comparing measuring devices which varies in pricing, fastness and other features, such as efficiency, has been of growing interest in several areas like engineering, medicine, psychology and agriculture. Barnett (1969) reports on the comparison of four combinations of two instruments and two operators for measuring vital capacity. Several other examples in the medical area are reported in the literature specially in Kelly (1984, 1985), Chipkevitch et al. (1996) and Lu et al. (1997).

[1]e-mail: vhlachos@bayesforecast.com
[2]e-mail: hbolfar@ime.usp.br
[3]e-mail: fily@ime.unicamp.br
[4]e-mail: manuel.galea@uv.cl

In this paper we consider the comparative calibration models designed to compare the efficiency of several measuring devices (or instruments) when measuring the same unknown quantity x in a common group of individuals or experimental units. It is assumed that the observed measurements follow a multivariate skew-normal distribution. This study is motivated by the fact that many data sets considered in the comparative calibration literature seem to present nonnormal behavior such as asymmetry. This is the case with the data sets studied in Barnett (1969) and in Chipkevitch et al. (1996) (see also Bolfarine et al., 2002) which require data transformation in order to be better approximated by the normal distribution.

Outliers and detection of influential observations is an important step in the analysis of a data set. There are several ways of evaluating the influence of perturbations in the data set and in the model given the parameter estimates. On the other hand, there are just a few works in the literature for diagnostic and influence of observations in models with measurement errors. Kelly (1984) considered a diagnostic procedure in the structural linear model based on the influence function. Tanaka et al. (1991) also consider the influence function introduced by Hampel for evaluating the influence of observations in the analysis of covariance structures. Rather than eliminating cases, the approach proposed by Cook (1986) is a general method for evaluating, under the maximum likelihood estimators, the influence of small perturbations in the model or data set. Recently, Galea et al. (2002) apply the local influence method in functional and structural normal comparative calibration models (NCCM). However, no application of local influence has been considered for comparative calibration under the skew-normal distribution.

Specifically, we first extend the NCCM by considering that the observed responses follow the skew-normal distribution so that the *Skew-Normal Structural Comparative Calibration Model* (SNCCM) is defined. Closed form expressions are obtained for the likelihood function which extends results in Arellano-Valle et al. (2005b). The likelihood function can then be directly maximized by using statistical software such as the *R* program. Alternatively, an EM-type algorithm is developed which seems to be more robust with respect to starting values. A local influence study is also conducted for such general models in order to study the influence of observations on the maximum likelihood estimators. We develop model curvature for two types of perturbation schemes.

The paper is organized as follows. In Section 2 the skew-normal distribution is revised in univariate and multivariate contexts. In Section 3 we present the SNCCM model, the marginal likelihood function of the observed data is derived in closed form and an EM-type algorithm for maximum likelihood estimation is developed by exploring statistical properties of the model considered. In Section 4 we discuss the main concepts of local influence and the related concept of diagnostic. Section 5 is dedicated to the derivation of the appropriate matrices for the curvature calculation under two perturbation schemes. Finally, in Section 6 applications the results and methods are illustrated with two examples using data sets previously analyzed in the literature. Global and local influence for the skew-normal comparative calibration model are compared.

2 The skew-normal distribution

In this section we present a review of the literature in skew-normal models. We start by giving the notation that will be used throughout the paper. Let $\phi_k(\mathbf{x}|\boldsymbol{\mu},\Sigma)$ and $\Phi_k(\mathbf{x}|\boldsymbol{\mu},\Sigma)$ be the probability density function (pdf) and the cumulative distribution function (cdf), respectively, of the $N_k(\boldsymbol{\mu},\Sigma)$ distribution evaluated at $\mathbf{x}$. When $\boldsymbol{\mu}=\mathbf{0}$ and $\Sigma=\mathbf{I}_k$ (the $k\times k$ identity matrix), we denote these functions as $\phi_k(\mathbf{x})$ and $\Phi_k(\mathbf{x})$, *respectively.*

The skew-normal distribution was previously considered among others by O'Hagan and Leonard (1976) as a prior distribution in Bayesian analysis. This family of distributions was further studied from a classical point of view by Azzalini (1985, 1986) in a univariate context as a natural extension of the classical normal density to accommodate asymmetry. Arnold et al. (1993) consider some asymmetric models by using truncated normal distributions and using this idea more general skew distributions have been developed. In a recent review, Azzalini (2005) provides a detailed accounting on previous appearances of the skew-normal distribution. Following Azzalini (1985), a random variable Y follows a univariate skew-normal distribution with location parameter μ, scale parameter σ^2 and skewness parameter λ, which we denote by $Y\sim SN_1(\mu,\sigma^2,\lambda)$, if the pdf of Y is given by

$$f_Y(y)=2\phi_1\left(y|\mu,\sigma^2\right)\Phi_1\left(\lambda\frac{y-\mu}{\sigma}\right). \tag{1}$$

Note that if $\lambda=0$ then the density of Y in (1) reduces to the density of the normal distribution. We use the notation $Y\sim SN_1(\lambda)$ when $\mu=0$ and $\sigma^2=1$, that is, the standard skew-normal distribution. Probabilistic properties of this distribution can be found in Azzalini (1985,1986), Henze (1986), Pewsey (2000) among others. Henze (1986) develops a stochastic representation for the above distribution which allows obtaining easily many of its properties such as the asymmetry index, the kurtosis index and all its moments.

Studies on multivariate skew-normal distributions are considered in Azzalini and Dalla-Valle (1996), Azzalini and Capitanio (1999), Branco and Dey (2001), Sahu et al. (2003), Genton (2004), among others. Arellano-Valle and Genton (2005) introduce the class of fundamental skewed distributions, giving an unified approach to obtain multivariate skew distributions starting from symmetric ones. Arellano-Valle et al. (2005a) propose a multivariate skew-normal distribution, which is suitable to make inferential studies in many types of models. Following their approach, we say that a k-dimensional random vector $\mathbf{Y}$ follows a skew-normal distribution with location vector $\boldsymbol{\mu}\in\mathbb{R}^k$, dispersion matrix Σ (a $k\times k$ positive definite matrix) and skewness vector $\boldsymbol{\lambda}\in\mathbb{R}^k$, if its pdf is given by

$$f_{\mathbf{Y}}(\mathbf{y})=2\phi_k(\mathbf{y}|\boldsymbol{\mu},\Sigma)\Phi_1(\boldsymbol{\lambda}^\top\Sigma^{-1/2}(\mathbf{y}-\boldsymbol{\mu})),\quad \mathbf{y}\in\mathbb{R}^k, \tag{2}$$

which we denote this by $\mathbf{Y}\sim SN_k(\boldsymbol{\mu},\Sigma,\boldsymbol{\lambda})$ and by $\mathbf{Y}\sim SN_k(\boldsymbol{\lambda})$ when $\boldsymbol{\mu}=\mathbf{0}$ and $\Sigma=\mathbf{I}_k$, the k-dimensional identity matrix. A stochastic representation for the distribution in (2) follows by considering that if $\mathbf{Y}\sim SN_k(\boldsymbol{\mu},\Sigma,\boldsymbol{\lambda})$, then

$$\mathbf{Y}\stackrel{d}{=}\boldsymbol{\mu}+\Sigma^{1/2}(\boldsymbol{\delta}|X_0|+(\mathbf{I}_k-\boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{X}_1),\quad\text{with}\quad \boldsymbol{\delta}=\frac{\boldsymbol{\lambda}}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}, \tag{3}$$

where $X_0\sim N_1(0,1)$ independent of $\mathbf{X}_1\sim N_k(\mathbf{0},\mathbf{I}_k)$ and " $\stackrel{d}{=}$ " meaning "distributed as". This representation is a generalization of the univariate representation in Henze (1986). As in the multivariate case, many properties of the multivariate distribution in (2) can be derived from this stochastic representation. See Arellano-Valle and Genton (2005) and Arellano-Valle et al. (2005a) for details.

3 The SNCCM and Maximum Likelihood Estimation

Suppose that we have at our disposal $p \geq 3$ instruments for measuring a characteristic of interest x in a group of n experimental units. Let x_i the true (unknown) value in unit i and y_{ij} the measured value obtained with the instrument j in unit i, $i = 1, \ldots, n$ and $j = 1, \ldots, p$ which, we consider to be related through the following linear model (see, Barnett, 1969 and Shyr and Gleser, 1986),

$$\mathbf{Y}_i = \mathbf{a} + \mathbf{b}x_i + \boldsymbol{\epsilon}_i, \tag{4}$$

where $\mathbf{a} = (0, \boldsymbol{\alpha}^\top)^\top = (0, \alpha_2, \ldots, \alpha_p)^\top$ and $\mathbf{b} = (1, \boldsymbol{\beta}^\top)^\top = (1, \beta_2, \ldots, \beta_p)^\top$ are $p \times 1$ vectors; $\mathbf{Y}_i = (y_{i1}, \ldots, y_{ip})^\top$ and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})^\top$ are $p \times 1$ random vectors, which will be called structural SNCCM if

$$\boldsymbol{\epsilon}_i \overset{iid}{\sim} N_p(0, D(\boldsymbol{\phi})) \text{ and } x_i \overset{iid}{\sim} SN_1(\mu_x, \phi_x, \lambda_x), \tag{5}$$

$i = 1, \ldots, n$, with $D(\boldsymbol{\phi}) = diag(\phi_1, \ldots, \phi_p)^\top$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$. The above model is considering that in the case of Barnett's (1969) data set, vital capacity is not symmetrically distributed in the population. The same seems to be the case with the testicular density data set in Chepkevitch et al. (1996). On the other hand, the errors $\boldsymbol{\epsilon}_i$, are related to measurement errors so that it reasonable to expect them to be normally distributed. The asymmetry parameter λ_x incorporates asymmetry in the latent variable x_i and consequently in the observed quantities $\mathbf{Y}_i$, $i = 1, \ldots, n$, which will be shown to have marginally a multivariate skew-normal distribution. If $\lambda_x = 0$, then the asymmetric model reduces to the normal comparative calibration model (NCCM) considered in Barnett (1969) in which inferences are extensively treated in the literature. Note from (3) that, the regression set up defined in (4)-(5) can be written hierarchically as

$$\mathbf{Y}_i \mid x_i \overset{ind}{\sim} N_p(\mathbf{a} + \mathbf{b}x_i, D(\boldsymbol{\phi})), \tag{6}$$

$$x_i \mid T_i = t_i \overset{ind}{\sim} N_1(\mu_x + \phi_x^{1/2}\delta_x t_i, \phi_x(1 - \delta_x^2)), \tag{7}$$

$$T_i \overset{iid}{\sim} HN_1(0, 1), \tag{8}$$

$i = 1, \ldots, n$, all independent, where $HN_1(0, 1)$ denote the standardized univariate half-normal distribution and $\delta_x = \lambda_x/(1 + \lambda_x^2)^{1/2}$. Classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \mu_x, \phi_x, \lambda_x)^\top$ in this type of model is based on the marginal distribution for the response $\mathbf{Y}_i$ (see, Bolfarine and Galea-Rojas, 1995), given in the following proposition. The proof is given in the appendix.

Proposition 1. *Under the structural SNCCM defined in (4)-(5), the marginal distribution of* $\mathbf{Y}_i$ *is given by*

$$f_{\mathbf{Y}_i}(\mathbf{y}_i|\boldsymbol{\theta}) = 2\phi_p(\mathbf{y}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\bar{\boldsymbol{\lambda}}_x^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu})), \tag{9}$$

i.e.,

$$\mathbf{Y}_i \overset{iid}{\sim} SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \bar{\boldsymbol{\lambda}}_x), \text{ with } \boldsymbol{\mu} = \mathbf{a} + \mathbf{b}\mu_x, \ \boldsymbol{\Sigma} = D(\boldsymbol{\phi}) + \phi_x \mathbf{b}\mathbf{b}^\top, \ \bar{\boldsymbol{\lambda}}_x = \frac{\lambda_x \phi_x \boldsymbol{\Sigma}^{-1/2}\mathbf{b}}{\sqrt{\phi_x + \lambda_x^2 \Lambda_x}},$$

$i = 1, \ldots, n$, *where* $\Lambda_x = (\phi_x{}^{-1} + \mathbf{b}^\top D^{-1}(\phi)\mathbf{b})^{-1}$.

Notice that the marginal distribution in (9) is also a multivariate skew-normal density of the form defined in Azzalini and Dalla-Valle (1996) since the skewing function is of dimension one. The log-likelihood function for $\boldsymbol{\theta}$ given the observed sample $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \tag{10}$$

where $\ell_i(\boldsymbol{\theta}) = \log(2) - (p/2)\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}d_i + \log(K_i)$, with $d_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$, $K_i = \Phi_1(\bar{\boldsymbol{\lambda}}_x^\top \Sigma^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}))$, $i = 1, \ldots, n$, and $\boldsymbol{\mu}$, Σ, $\bar{\boldsymbol{\lambda}}_x$ as in Proposition 1.

The result presented in Proposition 1 is important because it avoids, for example, using more complex numerical techniques such as Monte Carlo integration to carry out inferences in this type of models, given that it allows a closed form for the marginal distribution of $\mathbf{Y}_i$, $i = 1, \ldots, n$, facilitating straightforward implementation of inferences with standard optimization routines and existing statistical softwares. The asymptotic covariance matrix of the maximum likelihood estimators can be estimated by using the Hessian matrix, which can also be computed numerically using, for instance, the program R with the *optim* routine. If λ_x is suspected to be close to zero (symmetrical model) then it is more adequate to consider a normal model, given that under $\lambda_x = 0$ the information matrix can be singular although this is only proved for simpler models (see, Diciccio and Monti, 2004). Here, to obtain the maximum likelihood estimator of $\boldsymbol{\theta}$ we use the EM algorithm and derived algebraically its asymptotic covariance matrix.

The EM algorithm (Dempster, Laird, and Rubin, 1977) is a popular iterative algorithm for ML estimation in models with incomplete data. More specifically, let $\mathbf{y}$ denotes the observed data and $\mathbf{s}$ denotes the missing data. The complete data $\mathbf{y}_{comp} = (\mathbf{y}, \mathbf{s})$ is $\mathbf{y}$ augmented with $\mathbf{s}$. We denote by $\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{s})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the complete-data log-likelihood function and by $Q(\boldsymbol{\theta}|\boldsymbol{\theta}')$ the expected complete-data log- likelihood, that is,

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}') = E[\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{s})|\mathbf{y}, \boldsymbol{\theta}'].$$

Each iteration of the EM algorithm involves two steps, the expectation step and the maximization step:

E-step: Compute $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$ as a function of $\boldsymbol{\theta}$;

M-step: Find $\boldsymbol{\theta}^{(r+1)}$ such that $Q(\boldsymbol{\theta}^{(r+1)}|\boldsymbol{\theta}^{(r)}) = \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$.

Each iteration of the EM algorithm increases the likelihood function $\ell(\boldsymbol{\theta})$ and the EM algorithm typically converges to a local or global maximum of the likelihood function.

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{x} = (x_1, \ldots, x_n)^\top$ and $\mathbf{t} = (t_1, \ldots, t_n)^\top$. In the following we implement the EM algorithm for the structural SNCCM by considering that $(\mathbf{x}, \mathbf{t})$ are missing data, i.e, using double augmentation. Thus, under the hierarchical representation (6)-(8), with $\nu^2 = \phi_x(1 - \delta_x^2)$ and $\varsigma = \phi_x^{1/2}\delta_x$, it follows that the complete log-likelihood function associated with $(\mathbf{y},\mathbf{x},\mathbf{t})$ is

$$\ell_c(\theta|\mathbf{y},\mathbf{x},\mathbf{t}) \propto -\frac{n}{2}\log(|D(\phi)|) - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{a} - \mathbf{b}x_i)^\top D^{-1}(\phi)(\mathbf{y}_i - \mathbf{a} - \mathbf{b}x_i)$$
$$-\frac{n}{2}\log(\nu^2) - \frac{1}{2\nu^2}\sum_{i=1}^{n}(x_i - \mu_x - \varsigma t_i)^2. \tag{11}$$

Letting $\widehat{x}_i = E[x_i|\theta = \widehat{\theta}, \mathbf{y}_i]$, $\widehat{x_i^2} = E[x_i^2|\theta = \widehat{\theta}, \mathbf{y}_i]$, $\widehat{t}_i = E[t_i|\theta = \widehat{\theta}, \mathbf{y}_i]$, $\widehat{t_i^2} = E[t_i^2|\theta = \widehat{\theta}, \mathbf{y}_i]$ and $\widehat{t_i x_i} = E[t_i x_i|\theta = \widehat{\theta}, \mathbf{y}_i]$, we obtain using double conditional expectations and the moments of the truncated normal distributions (see Johnson et al., 1994, Section 10.1) that

$$\begin{aligned}
\widehat{t}_i &= \widehat{\mu}_{Ti} + W_{\Phi_1}\left(\frac{\widehat{\mu}_{T_i}}{\widehat{M}_T}\right)\widehat{M}_T, \\
\widehat{t_i^2} &= \widehat{\mu}_{T_i}^2 + \widehat{M_T^2} + W_{\Phi_1}\left(\frac{\widehat{\mu}_{T_i}}{\widehat{M}_T}\right)\widehat{M}_T\widehat{\mu}_{T_i}, \\
\widehat{x}_i &= \widehat{r}_i + \widehat{s}\,\widehat{t}_i, \\
\widehat{x_i^2} &= \widehat{T^2}_x + \widehat{r}_i^2 + 2\widehat{r}_i\,\widehat{s}\,\widehat{t}_i + \widehat{s}^2\,\widehat{t_i^2}, \quad \text{and} \\
\widehat{t_i x_i} &= \widehat{r}_i\,\widehat{t}_i + \widehat{s}\,\widehat{t_i^2},
\end{aligned} \tag{12}$$

where $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$, $M_T^2 = [1 + \varsigma^2\mathbf{b}^\top(D(\phi) + \nu^2\mathbf{b}\mathbf{b}^\top)^{-1}\mathbf{b}]^{-1}$, $\mu_{T_i} = \varsigma M_T^2\mathbf{b}^\top(D(\phi) + \nu^2\mathbf{b}\mathbf{b}^\top)^{-1}(\mathbf{y}_i - \mathbf{a} - \mathbf{b}\mu_x)$, $T_x^2 = \nu[1 + \nu\mathbf{b}^T D^{-1}(\phi)\mathbf{b}]^{-1}$, $r_i = \mu_x + T_x^2\mathbf{b}^T D^{-1}(\phi)(\mathbf{y}_i - \mathbf{a} - \mathbf{b}\mu_x)$ and $s = \varsigma(1 - T_x^2\mathbf{b}^T D^{-1}(\phi)\mathbf{b})$.

Using a simple algebra we get

$$E[\ell_c(\theta|\mathbf{y},\mathbf{x},\mathbf{t})|\mathbf{y},\widehat{\theta}] \propto -\frac{n}{2}\log(|D(\phi)|)$$
$$-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{a} - \mathbf{b}\widehat{x}_i)^\top D^{-1}(\phi)(\mathbf{y}_i - \mathbf{a} - \mathbf{b}\widehat{x}_i) - \frac{1}{2}({\Lambda_x}^{-1} - {\phi_x}^{-1})\sum_{i=1}^{n}(\widehat{x^2}_i - \widehat{x}_i^2)$$
$$-\frac{n}{2}\log(\nu^2) - \frac{1}{2}\sum_{i=1}^{n}(\widehat{x^2}_i + \mu_x^2 + \varsigma^2\widehat{t_i^2} - 2\widehat{x}_i\mu_x - 2\widehat{x_i t_i}\mu_x + 2\varsigma\mu_x\widehat{t}_i). \tag{13}$$

We then have the following EM algorithm:

E-step: Given $\theta = \widehat{\theta}$, compute $\widehat{t}_i$, $\widehat{t_i^2}$, $\widehat{x}_i$, $\widehat{x_i^2}$ and $\widehat{x_i t_i}$ for $i = 1, \ldots, n$, using (12).

M-step: Update $\widehat{\theta}$ by maximizing $E[\ell_c(\theta|\mathbf{y},\mathbf{x},\mathbf{t})|\mathbf{y},\widehat{\theta}]$ over θ, which leads to

$$\widehat{\alpha}_j = \bar{y}_j - \bar{\widehat{x}}\beta_j, \tag{14}$$

$$\widehat{\beta}_j = \frac{\sum_{i=1}^{n}\widehat{x}_i(y_{ij} - \bar{y}_j)}{\sum_{i=1}^{n}\widehat{x_i^2} - n(\bar{\widehat{x}})^2}, \tag{15}$$

$$\widehat{\phi}_1 = \frac{1}{n}\sum_{i=1}^{n}(y_{i1}^2 - 2\widehat{x}_i y_{i1} + \widehat{x_i^2}), \tag{16}$$

$$\widehat{\phi}_j = \frac{1}{n}\sum_{i=1}^{n}(y_{ij}^2 + \alpha_j^2 + \beta_j^2\widehat{x_i^2} - 2\alpha_j y_{ij} - 2y_{ij}\beta_j\widehat{x}_i + 2\alpha_j\beta_j\widehat{x}_i), \tag{17}$$

$$\widehat{\mu}_x = \frac{1}{n}\sum_{i=1}^{n}(\widehat{x}_i - \varsigma\widehat{t}_i), \tag{18}$$

$$\widehat{\nu}^2 = \frac{1}{n}\sum_{i=1}^{n}(\widehat{x^2}_i + \mu_x^2 + \varsigma^2\widehat{t_i^2} - 2\mu_x\widehat{x}_i - 2\varsigma\widehat{t_i x_i} + 2\varsigma\mu_x\widehat{t}_i), \text{ and} \tag{19}$$

$$\widehat{\varsigma} = \frac{\sum_{i=1}^{n}(\widehat{t_i x_i} - \mu_x\widehat{t}_i)}{\sum_{i=1}^{n}\widehat{t_i^2}}, \tag{20}$$

where $\bar{y}_j = \frac{1}{n}\sum_{i=1}^{n} y_{ij}$, $\bar{\widehat{x}} = \frac{1}{n}\sum_{i=1}^{n}\widehat{x}_i$ and $j = 2, \ldots, p$.

The shape and scale parameters of the latent variable x, can be estimated by noting that $\varsigma/\nu = \lambda_x$, and $\phi_x = \varsigma^2 + \nu^2$. Starting values are often chosen to be the corresponding estimates under a normal assumption, where the starting values for the asymmetric parameters are set to be 0 and as recommended in the literature, it is useful to run the EM-algorithm several times with different starting values. Following Arellano-Valle et al. (2005a) we also propose selecting the best fit by inspection of information criteria such as Akaike's Information Criterion (AIC, $-\ell(\widehat{\theta})/N + P/N$), Schawarz's Bayesian Information Criterion (BIC, $-\ell(\widehat{\theta})/N + 0.5\log(N)P/N$), and the Hannan-Quinn Criterion (HQ, $-\ell(\widehat{\theta})/N + \log(\log(N))P/N$), where P is the number of free parameters in the model and $N = p \times n$. This approach can be used in practice to select between NCCM and SNCCM fits. Note that under $\lambda_x = 0$ (or $\varsigma = 0$) the M-step equations reduce to the equations obtained in Bolfarine and Galea-Rojas (1995).

4 Influence diagnostics

Outliers and detection of influential observations is an important step in the analysis of a data set. There are several alternatives to evaluate the influence of perturbation in the data and/or in the model on the parameter estimators. For example, see Cook and Weisberg (1982), Chatterjee and Hadi (1988).

Case deletion is a common way to assess the effect of an observation on the estimation process. This is a global influence analysis, since the effect of observation is evaluated by eliminating it from the data set. Alternatively, local influence is based more on geometric differentiation rather than on the elimination of observations. A differential comparison of estimators is used before and after perturbing the data and/or model assumptions. In order to evaluate the robustness of the maximum likelihood estimator, to possible atypical observations, in the data set, we use the local influence concept introduced by Cook (1986).

Let $l(\boldsymbol{\theta})$ denote the log-likelihood function from the postulated model and let $\boldsymbol{\omega}$ be a $q \times 1$ vector of perturbation restricted to some open subset of $\mathbb{R}^q$. The perturbations are made in the likelihood function such that it takes the form $l(\boldsymbol{\theta}|\boldsymbol{\omega})$, with $\boldsymbol{\omega}$ as the perturbation vector. Denoting the vector of no perturbation by $\boldsymbol{\omega}_0$, we assume $l(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = l(\boldsymbol{\theta})$. To asses the influence of the perturbations on the maximum likelihood estimate of $\boldsymbol{\theta}$, one may consider the likelihood displacement

$$LD(\omega) = 2[l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_\omega)],$$

where $\widehat{\boldsymbol{\theta}}_\omega$ ($\widehat{\boldsymbol{\theta}}$) denotes the maximum likelihood estimator under the model $l(\boldsymbol{\theta}|\omega)$ ($l(\boldsymbol{\theta})$). The idea of local influence (Cook, 1986) is concerned in characterizing the behavior of $LD(\omega)$ at ω_0. The procedure consists in selecting a unit direction $\boldsymbol{d}$, $||\boldsymbol{d}|| = 1$, and then to consider the plot of $LD(\omega_0 + a\boldsymbol{d})$ against $a \in \mathbb{R}$. This plot is called *lifted line.* Notice that since $LD(\omega_0) = 0$, $LD(\omega_0 + a\boldsymbol{d})$ has a local minimum at $a = 0$. Each lifted line can be characterized by considering the normal curvature $C_d(\boldsymbol{\theta})$ around $a = 0$. The suggestion is to consider the direction $\boldsymbol{d}_{\max}$ corresponding to the largest curvature $C_{\mathbf{d}\max}(\boldsymbol{\theta})$. The index plot of $\boldsymbol{d}_{\max}$ may reveal those observations that under small perturbations exert notable influence on $LD(\omega)$. Cook (1986) showed that the normal curvature at the direction $\boldsymbol{d}$ takes the form

$$C_{\mathbf{d}}(\boldsymbol{\theta}) = 2|\boldsymbol{d}^\top \boldsymbol{\Delta}^\top \mathbf{L}^{-1} \boldsymbol{\Delta} \boldsymbol{d}|, \tag{21}$$

where $-\mathbf{L}$ is the observed information matrix for the postulated model ($\omega = \omega_0$) and $\boldsymbol{\Delta}$ is the $p \times q$ matrix with elements

$$\Delta_{ij} = \frac{\partial^2 l(\boldsymbol{\theta}|\omega)}{\partial \theta_i \partial \omega_j},$$

evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\omega = \omega_0, i = 1, ..., p$ and $j = 1, ..., q$. Therefore, the maximization of (21) is equivalent to finding the largest absolute eigenvalue $C_{\mathbf{d}\max}$ of the matrix $\mathbf{B} = \boldsymbol{\Delta}^\top \mathbf{L}^{-1} \boldsymbol{\Delta}$ and $\boldsymbol{d}_{\max}$ is the corresponding eigenvector. In some situations, it may be of interest to assess the influence on a subset $\boldsymbol{\theta}_1$ of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$. For example, one may have interest on $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$ or $\boldsymbol{\theta}_1 = \phi$. In such situations, the curvature at the direction $\boldsymbol{d}$ is given by

$$C_d(\boldsymbol{\theta}_1) = 2|\boldsymbol{d}^\top \boldsymbol{\Delta}^\top (\mathbf{L}^{-1} - \mathbf{B}_{22}) \boldsymbol{\Delta} \boldsymbol{d}|, \tag{22}$$

where

$$\mathbf{B}_{22} = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{L}_{22}^{-1} \end{pmatrix},$$

and $\mathbf{L}_{22}$ is obtained from the partition of $\mathbf{L}$ according to the partition of $\boldsymbol{\theta}$. The eigenvector $\boldsymbol{d}_{\max}$ corresponds to the largest absolute eigenvalue of the matrix $\mathbf{B} = \boldsymbol{\Delta}^\top (\mathbf{L}^{-1} - \mathbf{B}_{22}) \boldsymbol{\Delta}$. Another important direction, according to Escobar and Meeker (1992) (see also Verbeke and Molenberghs, 2000) is $\boldsymbol{d} = \boldsymbol{e}_{in}$, a vector of zeroes with a one in the i-th position. In that case, the normal curvature, called the total local influence of individual i, is given by $C_i = 2|\boldsymbol{e}_{in}^\top \mathbf{B} \boldsymbol{e}_{in}| = 2|b_{ii}|$, where b_{ii} is the ith diagonal element of $\mathbf{B}$, $i = 1, ..., n$.

In order to compare local and global influence, we use the Cook's distance (D_i) and the likelihood displacement (LD_i), which are defined, respectively, by

$$D_i = (\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}})^\top (-L)(\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}})/(3p+1), \tag{23}$$

$$LD_i = 2(l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_{(i)})), \tag{24}$$

$i = 1, ..., n$, where $\widehat{\boldsymbol{\theta}}_{(i)}$ denotes the parameter estimates (MLE) without case i. See Zhao and Lee (1998) for details.

5 Curvature Derivation for SNCCM

In this Section we derive the observed information matrix and the $\boldsymbol{\Delta}$ matrix for different perturbation schemes. The observed information matrix is derived first.

5.1 The observed information matrix

From (10) and the notation in Proposition 1, we have after some algebraic manipulation that the log-likelihood function can be, alternatively, written as:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \tag{25}$$

where $\ell_i(\boldsymbol{\theta}) \propto -\frac{1}{2}\log|\Sigma| - \frac{1}{2}d_i + \log(K_i)$ with $d_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$ and $K_i = \Phi_1(A_x a_i)$, with $A_x = \lambda_x \Lambda_x/(\phi_x + \lambda_x^2 \Lambda_x)^{1/2}$, $a_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top D^{-1}(\boldsymbol{\phi})\boldsymbol{b}$, $i = 1, \ldots, n$. The matrix of second derivatives with respect to $\boldsymbol{\theta}$ is given by

$$\mathbf{L} = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}. \tag{26}$$

The elements of this matrix are given in the Appendix B. Asymptotic confidence intervals and test on the MLEs can be obtained using this matrix, This is, if $\mathbf{J} = -\mathbf{L}$ denote the observed information matrix for the marginal log-likelihood $\ell(\boldsymbol{\theta})$ of the SNCCM, then asymptotic confidence intervals and hypotheses tests for the parameter $\boldsymbol{\theta}$ are obtained assuming that the MLE $\boldsymbol{\theta}$ has approximately a $N_{3p+1}(\boldsymbol{\theta}, \mathbf{J}^{-1})$ 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 MLE $\hat{\boldsymbol{\theta}}$.

5.2 Case weights perturbation

Notice that logarithm of the likelihood function for the model (4)-(5) is given by (10) where $\ell_i(\boldsymbol{\theta})$, is the contribution from the *i-th* observation (equally weighted) to the likelihood, $i = 1, \ldots, n$. A perturbed log-likelihood function - allowing different weights for different observations - can be defined by

$$\ell(\boldsymbol{\theta}/\boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i \ell_i(\boldsymbol{\theta}), \tag{27}$$

where, $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \mu_x, \phi_x, \lambda_x)^\top$, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ is the vector of weights of the contributions from each observation to the likelihood and $\omega_0 = 1_n = (1, \ldots, 1)^\top$ is the nonperturbation point, that is, $l(\boldsymbol{\theta}/\omega_0) = l(\boldsymbol{\theta})$. This perturbation scheme is intended to evaluate whether the contribution of the observations with differing weights affects the maximum likelihood estimator of $\boldsymbol{\theta}$. Perhaps, this is the method most commonly used to evaluate the influence of a small modification of the model. Thus, using (27) it follows, after some algebraic manipulation, that the delta matrix is given by

$$\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_n), \tag{28}$$

where $\Delta_i = \dfrac{\partial \ell_i(\theta)}{\partial \theta}$, $i = 1, \ldots, n$, with elements

$$\frac{\partial l_i(\theta)}{\partial \gamma} = -\frac{1}{2}\left[\frac{\partial \log|\Sigma|}{\partial \gamma} + \frac{\partial d_i}{\partial \gamma} - 2\frac{\partial \log(K_i)}{\partial \gamma}\right], \quad \gamma = \alpha, \beta, \phi, \mu_x, \phi_x, \lambda_x. \tag{29}$$

The components $\dfrac{\partial \log|\Sigma|}{\partial \gamma}$, $\dfrac{\partial d_i}{\partial \gamma}$ and $\dfrac{\partial \log(K_i)}{\partial \gamma}$ are presented in Appendix B. The above Δ matrix should be evaluated at the MLEs and ω_o.

5.3 Perturbation of the measurements from one instrument

In this section the measurements are obtained when one instrument is modified considering additive and multiplicative perturbation schemes. Supposing that the measurements from instrument $m = 1, \ldots, p$, are chosen to be perturbed, then the perturbed model is given by

$$Y_{mi}(\omega) = \mathbf{a} + \mathbf{b}x_i + \epsilon_i, \tag{30}$$

with

$$\mathbf{Y}_{mi}(\omega_i) = \left[\begin{array}{ll} \mathbf{Y}_i + \omega_i \mathbf{e}_m, & \text{additive perturbation;} \\ \mathbf{Y}_i \odot \mathbf{1}_m(\omega_i), & \text{multiplicative perturbation,} \end{array}\right. \tag{31}$$

where $\mathbf{e}_m$ is a p-dimensional null vector with one in the mth position, $\mathbf{1}_m(\omega_i)$ of dimension p denoting a vector of ones having the mth component replaced by ω_i and $\odot$ denotes the Hadamard (elementwise) product.

Let $\omega = (\omega_1, \ldots, \omega_n)^\top$. The no perturbation case follows by taking $\omega_o = \mathbf{0}$ in the additive case and $\omega_o = \mathbf{1}$ in the multiplicative case. The perturbed log-likelihood follows from (25) with $\mathbf{y}_{mi}(\omega_i)$ replacing $\mathbf{y}_i, i = 1, \ldots, n$. This is,

$$\ell(\theta|\omega) = \sum_{i=1}^{n} \ell_i(\theta|\omega_i), \tag{32}$$

where $\ell_i(\theta|\omega_i) \propto -\frac{1}{2}\log|\Sigma| - \frac{1}{2}d_{mi}(\omega_i) + \log(K_{mi}(\omega_i))$ with $d_{mi}(\omega_i) = (\mathbf{y}_{mi}(\omega_i) - \mu)^\top \Sigma^{-1}(\mathbf{y}_{mi}(\omega_i) - \mu)$, $K_{mi} = \Phi_1(A_x a_{mi}(\omega_i))$ and $a_{mi}(\omega_i) = (\mathbf{y}_{mi}(\omega_i) - \mu)^\top D^{-1}(\phi)\mathbf{b}$.

Differentiating $\ell(\theta|\omega)$ with respect to ω and θ leads to:

$$\Delta = (\Delta^m_{1\theta}, \ldots, \Delta^m_{n\theta}), \tag{33}$$

where

$$\begin{aligned} \Delta^m_{i\theta} &= \frac{\partial T_{mi}(\omega_i)}{\partial \theta} + W_{\Phi_1}(A_x a_{mi}(\omega_i))\left[\frac{\partial A_x}{\partial \theta}S_{mi} + A_x\frac{\partial S_{mi}(\omega_i)}{\partial \theta}\right] \\ &\quad + A_x W'_{\Phi_1}(A_x a_{mi}(\omega_i))S_{mi}\left[A_x\frac{\partial a_m(\omega_i)}{\partial \theta} + a_{mi}(\omega_i)\frac{\partial A_x}{\partial \theta}\right], \end{aligned} \tag{34}$$

with $T_{mi}(\omega_i) = -(\mathbf{y}_{mi}(\omega_i) - \mu)^\top \Sigma^{-1}\dfrac{\partial \mathbf{y}_{mi}(\omega_i)}{\partial \omega_i}$, $S_{mi}(\omega_i) = \mathbf{b}^\top D^{-1}(\phi)\dfrac{\partial \mathbf{y}_{mi}(\omega_i)}{\partial \omega_i}$, $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$, $W'_{\Phi_1}(u) = -W_{\Phi_1}(u)(u + W_{\Phi_1}(u))$, $u \in \mathbb{R}$, and $\dfrac{\partial a_{mi}(\omega_i)}{\partial \theta}$ as in the unperturbed

Table 1: Results of fitting normal and skew-normal calibration comparative models (NCCM and SNCCM) to the Barnett data set. SE are the estimated asymptotic standard errors based in the information matrix given in the Appendix A.

	NCCM		SNCCM	
Parameter	Estimate	SE	Estimate	SE
α_2	-2.0447	1.0532	-2.0719	1.0457
α_3	-5.2857	1.2278	-5.2687	1.2334
α_4	-4.3726	1.2406	-4.3669	1.2431
β_2	1.0597	0.0446	1.0609	0.0443
β_3	1.1919	0.0521	1.1912	0.0523
β_4	1.1306	0.0526	1.1304	0.0527
ϕ_1	5.0247	1.0006	5.0399	0.9959
ϕ_2	1.9151	0.5846	1.7957	0.5799
ϕ_3	2.9236	0.7893	3.0425	0.8103
ϕ_4	3.8845	0.8744	3.9338	0.8961
μ_x	22.4611	0.9008	13.0562	1.2599
σ_x^2	53.3983	9.7174	141.8462	34.1031
λ_x	-	-	5.1780	2.9403
- log-likelihood	788.0931		783.5044	
AIC	2.7781		2.7656	
BIC	2.8544		2.8483	
HQ	2.7726		2.7596	

case, replacing $\mathbf{y}_i$ by $\mathbf{y}_{mi}(\omega_i)$ (see Appendix B). Note that, $\dfrac{\partial \mathbf{y}_{mi}(\omega_i)}{\partial \omega_i} = \mathbf{e}_m$ in the additive case and $\dfrac{\partial \mathbf{y}_{mi}(\omega_i)}{\partial \omega_i} = \mathbf{1}_m(y_{im})$ in the multiplicative case, $i = 1, \ldots, n$. The above $\boldsymbol{\Delta}$ matrix should be evaluated at the MLEs and $\boldsymbol{\omega}_o$. Expression for the derivatives $\dfrac{\partial T_{mi}(\omega_i)}{\partial \boldsymbol{\theta}}$ and $\dfrac{\partial S_{mi}(\omega_i)}{\partial \boldsymbol{\theta}}$ are given in the Appendix C.

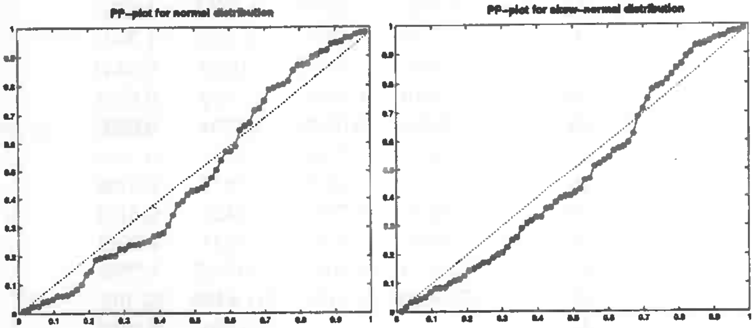
6 Applications

In this section we analyze two real data sets. We will focus only on the parameter set $\boldsymbol{\theta}$.

Barnett (1969) data set. Two instruments used for measuring the vital capacity of the human lung and operated by skilled and unskilled operators were compared on a common group of 72 patients. We consider the measurements divided by 100 in order to achieve numerical stability. Data transformation was used to improve the normal fitting (Bolfarine et al., 2002). We compare in the sequel NCCM (extensively treated in the literature) and SNCCM fitting for this data set. Resulting parameter estimates are given in Table 1. Note that using AIC, BIC and HQ values shown in the bottom of the Table 1 favors SNCCM, supporting the contention of the departure from normality. A more emphatic conclusion in this direction is achieved by considering a parametric test for normality, where the null hypothesis is $\lambda_x = 0$, such as the likelihood ratio test for which the observed value of the test statistic in this case is 9.1773 and the associated critical level of the $\chi^2(1)$ at 5% is 3.84. This conclusion is also corroborated by the PP-plot (Azzalini, 2005) reported in Figure 1.

Figure 1: Barnett data set: Healy's type plot for (a) Normal CCM, (b) Skew-normal CCM.

(a) (b)

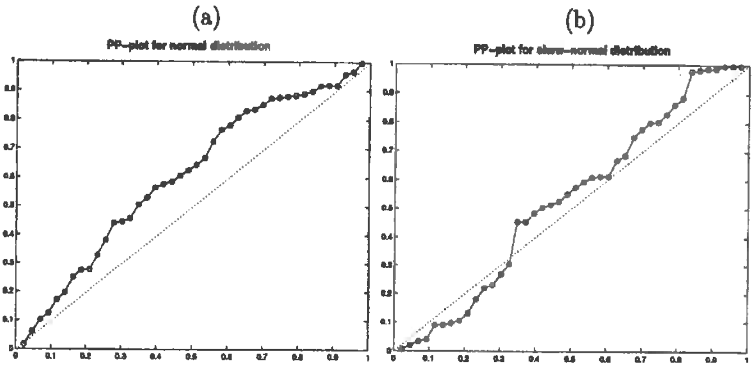


With this data set, a local influence study conducted by the authors has led to the same conclusions as the ones resulting by using the asymmetric normal model and reported in Galea-Rojas et al. (2002) so that they are not presented here.

Chipkevitch et al. (1996) data set. Measurement of the testicular volume of 42 adolescents were made in a certain sequence by using five different techniques: ultrasound (US), graphical method proposed by the authors (I), dimensional measurement (II), prader orchidometer (III) and ring orchidometer (IV), with the ultrasound approach assumed to be the reference measurement device. A histogram of the measurement (see figure 5(b)) shows certain asymmetry in the data set such that Galea et al. (2002) (see also Bolfarine et al., 2002) use a cubic root transformation to better achieve normality. Although such method (variable transformations) may give reasonable empirical results, it should be avoided if a more suitable theoretical model can be found (see Azzalini and Dalla-valle, 1996). By introducing a more flexible parametric distribution capable to accommodating such departures, we use a SNCCM to fitting this data set. Thus, under the SNCCM, parameter estimates (standard errors) are $\hat{\mu}_x = 4.0011(1.3951)$, $\hat{\alpha}_2 = 0.1022(0.5655)$, $\hat{\alpha}_3 = 0.0097(0.6217)$, $\hat{\alpha}_4 = 0.0481(0.6277)$, $\hat{\alpha}_5 = 1.5390(0.6337)$, $\hat{\beta}_2 = 0.8838(0.0509)$, $\hat{\beta}_3 = 0.9495(0.0559)$, $\hat{\beta}_4 = 1.1419(0.0565)$, $\hat{\beta}_5 = 1.0826(0.0570)$, $\hat{\phi}_x = 59.1993(21.5063)$, $\hat{\phi}_1 = 1.3385(0.3714)$, $\hat{\phi}_2 = 1.3285(0.3480)$, $\hat{\phi}_3 = 1.6737(0.4322)$, $\hat{\phi}_4 = 1.1577(0.3710)$, $\hat{\phi}_5 = 1.4104(0.3994)$ and $\hat{\lambda}_x = 4.7619(4.7495)$. The PP-plot reported in Figure 2 shows a visible improvement in the adequacy of the fit under the SNCCM relative to consider a NCCM for the observed measurement.

In the perturbation of case weights scheme we have $C_{dmax} = 6.8629$ and adolescents 31 and 32 stand out, as depicted in Figure 3(a). The same adolescents are found influential for the estimation of the complete parameter vector $\boldsymbol{\theta}$ using the total local influence C_i as shown in figure 3(b). Figures 3(c) and 3(d) gives the index plot of LD_i and LD_i for the SNCCM. Once again adolescents 31 and 32 are prominent.

Figure 2: Chipkevitch's data set: Healy's type plot for (a) Normal CCM, (b) Skew-normal CCM.



Examining the effects of perturbing the measurement taken by the five techniques as the previous example. The values of C_{dmax} for additive perturbations are 2.4273(US), 2.5756(I), 2.0696(II), 2.4854(III) and 2.9440(IV), whereas for multiplicative perturbations the C_{dmax} values are 427.6429(US), 400.9534(I), 460.4602(II), 563.5952(III) and 702.7542(IV). Figures 4 and 5(a) illustrate the differences between the perturbation schemes. Technique IV is chosen because it presents the largest C_{dmax} values.

In order to compare with the normal case in Figure (6) we show results of local influence under the NCCM using cubic root transformation as in Galea et al. (2002). Note that the conclusions are different because under the skew-normal model observations 31 and 32 are the most influent while under the symmetric normal model (after transforming the data) the most influent one is 20.

7 Final Conclusions

Paper deals with a skew normal comparative calibration model (SNCCM) with the normal comparative calibration model (NCCM) as a special case. Closed form expression is obtained for the likelihood function of the observed measurements which can be maximized by using existing statistical software. An EM-type algorithm is also developed by exploring statistical properties of the model considered. A local influence study is also conducted by developing model curvature for two perturbation schemes. Two data sets exhibiting skewness features are studied under the assumption of a skew normal distribution. In both cases, the SNCCM seems to present a better fit. We point out that the results and methods provided in this paper is not available elsewhere in the literature and the approaches used here can be used easily extended in treating other multivariate models.

Figure 3: Chipkevitch's data set. Index plot of (a) $|dmax|$. (b) C_i for perturbation of case weights for θ (c) Likelihood Displacement LD_i (d) Cook's distance D_i.

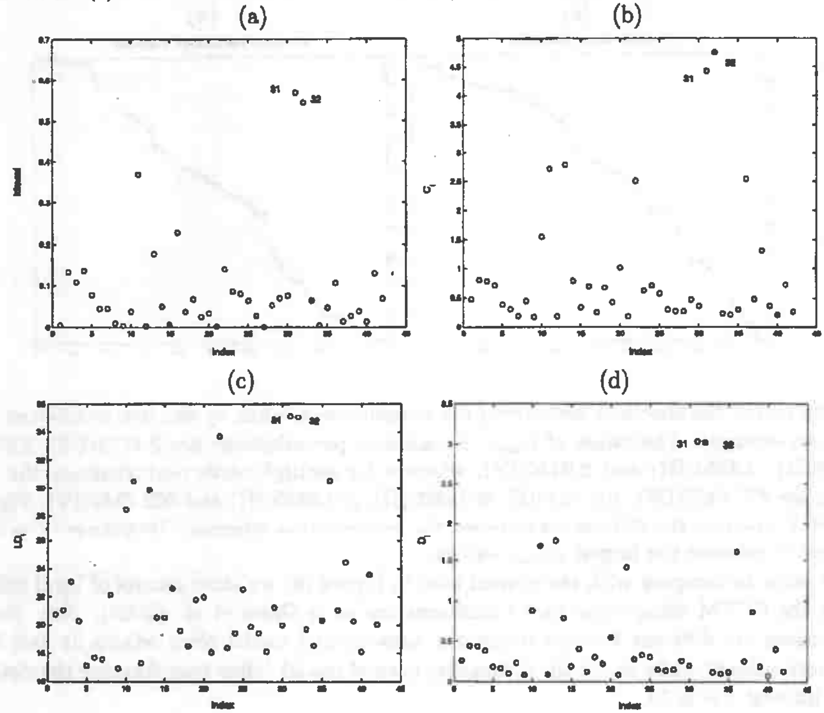


Figure 4: Chipkevitch's data set. Plot of $|d_{max}|$ versus the measurements taken by technique IV (a) Additive perturbation. (b) multiplicative perturbation.

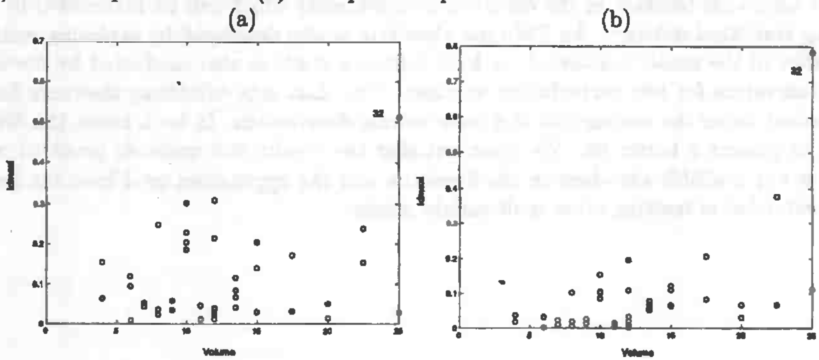
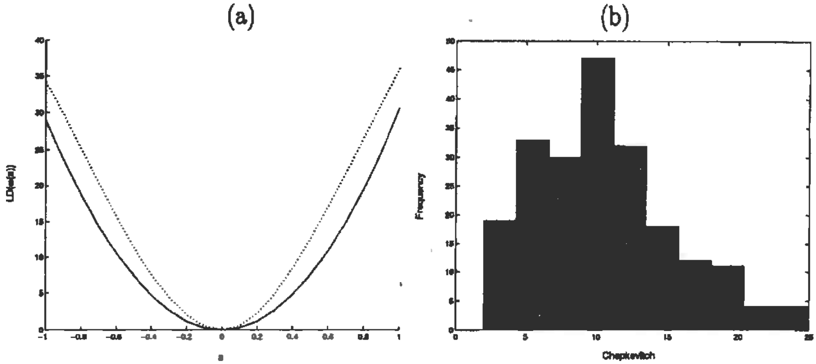


Figure 5: Chipkevitch's data set. (a) Plot of the likelihood displacement $LD(\omega(a))$ versus a for direction d_{max} imposing additive (solid-line) and multiplicative (dotted-line) perturbations on the measurements taken by techniques IV (b) Histogram for the observed measurements.



Appendix A: Proofs

Before proving the main results we consider the following lemmas. The notation used is that of Section 2.

Lemma A1. *Let* $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Then for any fixed k-dimensional vector* $\mathbf{a}$ *and* $k \times n$ *matrix* $\mathbf{B}$,

$$E[\Phi_k(\mathbf{a} + \mathbf{B}\mathbf{Y}|\boldsymbol{\eta}, \boldsymbol{\Omega})] = \Phi_k(\mathbf{a}|\boldsymbol{\eta} - \mathbf{B}\boldsymbol{\mu}, \boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top).$$

Proof. See Arellano-Valle et al. (2005a,b). □

Lemma A2. *Let* $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and* $\mathbf{X} \sim N_q(\boldsymbol{\eta}, \boldsymbol{\Omega})$. *Then,*

$$\begin{aligned} \phi_p(\mathbf{y}|\boldsymbol{\mu} + \mathbf{A}\mathbf{x}, \boldsymbol{\Sigma})\phi_q(\mathbf{x}|\boldsymbol{\eta}, \boldsymbol{\Omega}) &= \phi_p(\mathbf{y}|\boldsymbol{\mu} + \mathbf{A}\boldsymbol{\eta}, \boldsymbol{\Sigma} + \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^\top) \\ &\times \phi_q(\mathbf{x}|\boldsymbol{\eta} + \boldsymbol{\Lambda}\mathbf{A}^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu} - \mathbf{A}\boldsymbol{\eta}), \boldsymbol{\Lambda}), \end{aligned}$$

where $\boldsymbol{\Lambda} = (\boldsymbol{\Omega}^{-1} + \mathbf{A}^\top\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}$.

Proof. See Arellano-Valle et al. (2005a). □

Proof of Preposition 1:

From (4), (5) and the definition in (2) with $k = 1$, it follows that the marginal density of $\mathbf{Y}_i$ is obtained by computing the following integral:

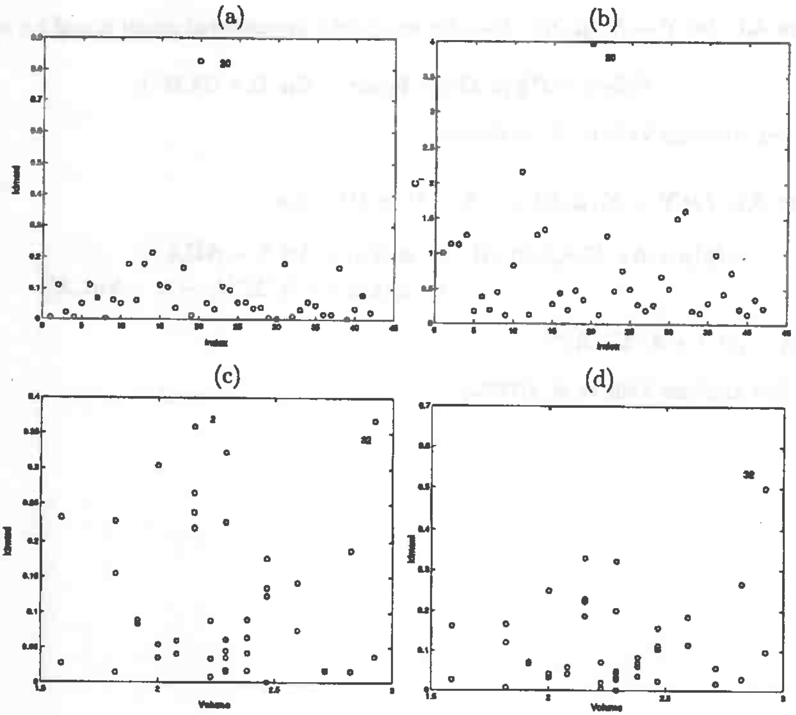
$$f_{\mathbf{Y}_i}(\mathbf{y}_i|\theta) = \int_{\mathbb{R}} f(\mathbf{y}_i|x_i,\theta) f(x_i|\theta) dx_i$$

$$= \int_{\mathbb{R}} 2\phi_p(\mathbf{y}_i|\mathbf{a} + \mathbf{b}x_i, D(\phi))\phi_1(x_i|\mu_x,\phi_x)\Phi_1\left(\frac{\lambda_x}{\phi_x^{1/2}}(x_i - \mu_x)\right) dx_i \tag{A1}$$

which, by using Lemma 2 is equal to:

$$f_{\mathbf{Y}_i}(\mathbf{y}_i|\theta) = 2\phi_p(\mathbf{y}|\mathbf{a} + \mathbf{b}\mu_x, \Sigma)E[\Phi_1(-\frac{\lambda_x\mu_x}{\phi_x^{1/2}} + \frac{\lambda_x Z_i}{\phi_x^{1/2}}|0,1)], \tag{A2}$$

Figure 6: Chipkevitch's data set diagnostic under the NCCM using cubic root transformation. Perturbation of case weights: Index plot of (a) $|dmax|$ and (b) C_i. Plot of $|d_{max}|$ versus the measurements taken using technique IV under: (c) Additive perturbation (d) multiplicative perturbation.



where $Z_i \sim N_1(\mu_x + \Lambda_x \mathbf{b}^\top D^{-1}(\phi)(\mathbf{y}_i - \mathbf{a} - \mathbf{b}\mu_x), \Lambda_x)$, with $\Lambda_x = (\phi_x{}^{-1} + \mathbf{b}^\top D^{-1}(\phi)\mathbf{b})^{-1}$. The proof is concluded by using Lema A1 and noting that $\Lambda_x \mathbf{b} D^{-1}(\phi) = \mathbf{b}\Sigma^{-1}$.

Appendix B: The observed information matrix in the skew-normal structural calibration comparative model

In this appendix the observed information matrix is obtained for the SNCCM. From (10), it follows that

$$\frac{\partial \ell_i(\theta)}{\partial \gamma} = -\frac{1}{2}\frac{\partial \log|\Sigma|}{\partial \gamma} - \frac{1}{2} d_{i\gamma} + \frac{\partial \log K_i}{\partial \gamma}, \tag{B1}$$

where

$$\frac{\partial \log K_i}{\partial \gamma} = W_{\Phi_1}(A_x a_i)[A_x \frac{\partial a_i}{\partial \gamma} + a_i \frac{\partial A_x}{\partial \gamma}],$$

with $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$, $u \in \mathbb{R}$, $A_x = \lambda_x \Lambda_x/(\phi_x + \lambda_x^2 \Lambda_x)^{1/2}$, $\Lambda_x = \phi_x/c$, $a_i = \boldsymbol{X}_i^\top D^{-1}(\phi)\boldsymbol{b}$, $d_{i\gamma} = \frac{\partial d_i}{\partial \gamma}, \gamma = \mu_x, \alpha, \beta, \phi_x, \phi, \lambda_x$ and $d_i = \boldsymbol{X}_i^\top \Sigma^{-1} \boldsymbol{X}_i$, $\boldsymbol{X}_i = \boldsymbol{Y}_i - \boldsymbol{a} - \boldsymbol{b}\mu_x$, $c = 1 + \phi_x \mathbf{b}^T D^{-1}(\phi)\mathbf{b}$, $i = 1, ..., n$. Further, using results in Nel (1980) related to vector derivatives it follows that,

$$\begin{aligned}
\frac{\partial \log|\Sigma|}{\partial \gamma} &= 0, \quad \gamma = \mu_x, \alpha, \lambda_x, \\
\frac{\partial \log|\Sigma|}{\partial \beta} &= 2\frac{\phi_x}{c} D^{-1}(\psi)\beta, \\
\frac{\partial \log|\Sigma|}{\partial \phi_x} &= c^{-1}\frac{c-1}{\phi_x}, \\
\frac{\partial \log|\Sigma|}{\partial \phi} &= -\frac{\phi_x}{c} D(b) D^{-2}(\phi) b + D^{-1}(\phi) 1_p,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial A_x}{\partial \gamma} &= 0, \quad \gamma = \mu_x, \alpha, \\
\frac{\partial A_x}{\partial \beta} &= -\frac{(2c + \lambda_x^2)}{\lambda_x^2} A_x^3 D^{-1}(\psi)\beta, \\
\frac{\partial A_x}{\partial \phi} &= \frac{(2c + \lambda_x^2)}{2\lambda_x^2} A_x^3 D(b) D^{-2}(\phi) b, \\
\frac{\partial A_x}{\partial \phi_x} &= \frac{(2c + \lambda_x^2 - c^2)}{2\phi_x^2 \lambda_x^2} A_x^3, \\
\frac{\partial A_x}{\partial \lambda_x} &= \frac{\phi_x}{\Lambda_x^2 \lambda_x^3} A_x^3,
\end{aligned}$$

$$\begin{aligned}
d_{i\mu_x} &= -2\boldsymbol{b}^\top \Sigma^{-1} \boldsymbol{X}_i, \\
d_{i\alpha} &= -2\mathbb{I}_{(p)} \Sigma^{-1} \boldsymbol{X}_i,
\end{aligned}$$

$$
\begin{aligned}
d_{i\beta} &= -2q_i D^{-1}(\psi)W_{2i} + 2c^{-1}\phi_x a_i q_i D^{-1}(\psi)\beta, \\
d_{i\phi_x} &= -c^{-2}a_i^2, \\
d_{i\phi} &= -D^{-2}(\phi)D(X_i)X_i + 2c^{-1}\phi_x a_i D^{-2}(\phi)D(b)X_i - c^{-2}\phi_x^2 a_i^2 D^{-2}(\phi)D(b)b, \\
d_{i\lambda_x} &= 0, \\
\frac{\partial a_i}{\partial \mu_x} &= -b^\top D^{-1}(\phi)b, \\
\frac{\partial a_i}{\partial \alpha} &= -D^{-1}(\psi)\beta, \\
\frac{\partial a_i}{\partial \beta} &= D^{-1}(\psi)W_{2i} - \mu_x D^{-1}(\psi)\beta, \\
\frac{\partial a_i}{\partial \phi_x} &= 0, \\
\frac{\partial a_i}{\partial \phi} &= -D(b)D^{-2}(\phi)X_i, \\
\frac{\partial a_i}{\partial \lambda_x} &= 0,
\end{aligned}
$$

where $M = D^{-1}(\phi)bb^\top D^{-1}(\phi)$, $\psi = (\phi_2, ..., \phi_p)^\top$, $W_{2i} = y_{2i} - \alpha - \beta\mu_x$, $\mathbb{I}_{(p)} = [0, \mathrm{I}_{p-1}]$ of dimension $p-1 \times p$ and $y_{i2} = (y_{i2}, ..., y_{ip})^\top$.

From (B.1) it follows that the observed, per element, information matrix is given by

$$
\mathrm{J}_i = -\mathrm{L}_i(\theta|y_i) = -\left(\frac{\partial^2 l_i(\theta)}{\partial\gamma\partial\tau^\top}\right), \tag{B2}
$$

where $\frac{\partial^2 \ell_i}{\partial\gamma\partial\tau^\top} = -\frac{1}{2}\frac{\partial^2 \log|\Sigma|}{\partial\gamma\partial\tau^\top} - \frac{1}{2}d_{i\gamma\tau^\top} + \frac{\partial^2 \log K_i}{\partial\gamma\partial\tau^\top}$, with

$$
\begin{aligned}
\frac{\partial^2 \log K_i}{\partial\gamma\partial\tau^\top} &= W_{\Phi_1}(A_x a_i)[\frac{\partial A_x}{\partial\gamma}\frac{\partial a_i}{\partial\tau^\top} + A_x\frac{\partial^2 a_i}{\partial\gamma\partial\tau^\top} + \frac{\partial a_i}{\partial\gamma}\frac{\partial A_x}{\partial\tau^\top} + a_i\frac{\partial^2 A_x}{\partial\gamma\partial\tau^\top}] \\
&\quad + \Delta_{\Phi_1}(A_x a_i)[A_x\frac{\partial a_i}{\partial\gamma} + a_i\frac{\partial A_x}{\partial\gamma}][A_x\frac{\partial a_i}{\partial\tau^\top} + a_i\frac{\partial A_x}{\partial\tau^\top}],
\end{aligned}
$$

$\Delta_{\Phi_1}(u) = W'_{\Phi_1}(u) = -W_{\Phi_1}(u)(u + W_{\Phi_1}(u))$, $u \in \mathbb{R}$, $d_{i\gamma\tau^\top} = \frac{\partial^2 d_i}{\partial\gamma\partial\tau^\top}$, $\gamma, \tau = \mu_x, \alpha, \beta, \phi_x, \phi, \lambda_x$, where

$$
\frac{\partial^2 \log|\Sigma|}{\partial\tau\partial\gamma^\top} = 0, \quad \tau = \mu_x, \alpha, \lambda_x; \quad \gamma = \mu_x, \alpha, \beta, \phi_x, \phi, \lambda_x
$$

$$
\begin{aligned}
\frac{\partial^2 \log|\Sigma|}{\partial\beta\partial\phi_x} &= 2c^{-2}D^{-1}(\psi)\beta, \\
\frac{\partial^2 \log|\Sigma|}{\partial\phi_x\partial\phi_x} &= -\frac{1}{c^2\phi_x^2}(c-1)^2, \\
\frac{\partial^2 \log|\Sigma|}{\partial\beta\partial\phi^\top} &= -2c^{-1}\phi_x[D_1(\beta) - c^{-1}\phi_x D^{-1}(\psi)\beta b^\top D(b)]D^{-2}(\phi),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= 2c^{-1}\phi_x[D^{-1}(\boldsymbol{\psi}) - 2c^{-1}\phi_x \boldsymbol{M}_1],\\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial\phi_x\partial\boldsymbol{\phi}^\top} &= -c^{-2}\boldsymbol{b}^\top D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top} &= -D^{-2}(\boldsymbol{\phi}) - c^{-2}\phi_x^2 D(\boldsymbol{b})D^{-1}(\boldsymbol{\phi})\boldsymbol{M}D^{-1}(\boldsymbol{\phi})D(\boldsymbol{b}) + 2c^{-1}\phi_x D^2(\boldsymbol{b})D^{-3}(\boldsymbol{\phi}),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 A_x}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= -(4\frac{\phi_x}{\lambda_x^2}A_x^3 - \frac{3(2c+\lambda_x^2)^2}{\lambda_x^4}A_x^5)\boldsymbol{M}_1 - \frac{2c+\lambda_x^2}{\lambda_x^2}A_x^3 D^{-1}(\boldsymbol{\psi}),\\
\frac{\partial^2 A_x}{\partial\boldsymbol{\beta}\partial\phi_x} &= -[\frac{2(c-1)}{\lambda_x^2\phi_x}A_x^3 + \frac{3(2c+\lambda_x^2)(2c+\lambda_x^2-c^2)}{2\lambda_x^4\phi_x^2}A_x^5]D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta},\\
\frac{\partial^2 A_x}{\partial\boldsymbol{\beta}\partial\boldsymbol{\phi}^\top} &= [2\frac{\phi_x}{\lambda_x^2}A_x^3 - \frac{3(2c+\lambda_x^2)^2}{2\lambda_x^4}A_x^5]D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}\boldsymbol{b}^\top D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi})\\
&\quad + \frac{2c+\lambda_x^2}{\lambda_x^2}A_x^3 \mathbb{I}_{(p)} D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2 A_x}{\partial\boldsymbol{\beta}\partial\lambda_x} &= \frac{\phi_x A_x^3}{\lambda_x^5\Lambda_x^2}(-3A_x^2(2c+\lambda_x^2) + 4\lambda_x^2\Lambda_x)D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta},\\
\frac{\partial^2 A_x}{\partial\phi_x\partial\phi_x} &= -\frac{\lambda_x^2+1}{\lambda_x^2\phi_x^3}A_x^3 + \frac{3(2c+\lambda_x^2-c^2)^2}{4\lambda_x^4\phi_x^4}A_x^5,\\
\frac{\partial^2 A_x}{\partial\phi_x\partial\boldsymbol{\phi}^\top} &= [\frac{(c-1)}{\lambda_x^2\phi_x}A_x^3 + \frac{3(2c+\lambda_x^2)(2c+\lambda_x^2-c^2)}{4\lambda_x^4\phi_x^2}A_x^5]\boldsymbol{b}^\top D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2 A_x}{\partial\phi_x\partial\lambda_x} &= \frac{c-2}{\lambda_x^3\Lambda_x\phi_x}A_x^3 + \frac{3(2c+\lambda_x^2-c^2)}{2\lambda_x^5\Lambda_x^2\phi_x}A_x^5,\\
\frac{\partial^2 A_x}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top} &= [-\frac{\phi_x}{\lambda_x^2}A_x^3 + \frac{3(2c+\lambda_x^2)^2}{4\lambda_x^4}A_x^5]D(\boldsymbol{b})D^{-1}(\boldsymbol{\phi})\boldsymbol{M}D^{-1}(\boldsymbol{\phi})D(\boldsymbol{b})\\
&\quad - \frac{2c+\lambda_x^2}{\lambda_x^2}D^2(\boldsymbol{b})D^{-3}(\boldsymbol{\phi})A_x^3,\\
\frac{\partial^2 A_x}{\partial\boldsymbol{\phi}\partial\lambda_x} &= \frac{\phi_x A_x^3}{2\lambda_x^5\Lambda_x^2}[3A_x^2(2c+\lambda_x^2) - 4\lambda_x^2\Lambda_x]D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi})\boldsymbol{b},\\
\frac{\partial^2 A_x}{\partial\lambda_x\partial\lambda_x} &= -\frac{3\phi_x}{\lambda_x^4\Lambda_x^2}A_x^3 + \frac{3\phi_x^2}{\lambda_x^6\Lambda_x^4}A_x^5,
\end{aligned}
$$

$$
\begin{aligned}
d_{i\mu_x\mu_x} &= 2\boldsymbol{b}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{b},\\
d_{i\mu_x\boldsymbol{\alpha}^\top} &= 2\boldsymbol{b}^\top\boldsymbol{\Sigma}^{-1}\mathbb{I}_{(p)}^\top,\\
d_{i\mu_x\boldsymbol{\beta}^\top} &= -2c^{-1}\boldsymbol{A}_i,\\
d_{i\mu_x\phi_x} &= 2\frac{(c-1)}{c^2\phi_x}a_i,\\
d_{i\mu_x\boldsymbol{\phi}^\top} &= 2c^{-1}\boldsymbol{X}_i^\top\boldsymbol{\Sigma}^{-1}D^{-1}(\boldsymbol{\phi})D(\boldsymbol{b}),\\
d_{i\boldsymbol{\alpha}\boldsymbol{\alpha}^\top} &= 2\mathbb{I}_{(p)}\boldsymbol{\Sigma}^{-1}\mathbb{I}_{(p)}^\top,
\end{aligned}
$$

$$
\begin{aligned}
d_{i\alpha\beta^\top} &= 2q_i[D^{-1}(\psi) - 2c^{-1}\phi_x M_1] + 2c^{-1}\phi_x D^{-1}(\psi)\beta(Y_{i2}-\alpha)^\top D^{-1}(\psi),\\
d_{i\alpha\phi_x} &= 2c^{-2}a_i D^{-1}(\psi)\beta,\\
d_{i\alpha\phi^\top} &= 2\mathbb{I}_{(p)}\Sigma^{-1}D^{-1}(\phi)[D(X_i) - c^{-1}\phi_x a_i D(b)],
\end{aligned}
$$

$$
\begin{aligned}
d_{i\beta\beta^\top} &= 4\frac{\phi_x^2}{c^2}a_i[D^{-1}(\psi)(Y_{i2}-\alpha-2\beta\mu_x)\beta^\top D^{-1}(\psi) + D^{-1}(\psi)\beta(Y_{i2}-\alpha-2\beta\mu_x)^\top D^{-1}(\psi)]\\
&\quad -2c^{-1}\phi_x D^{-1}(\psi)(Y_{i2}-\alpha-2\beta\mu_x)(Y_{i2}-\alpha-2\beta\mu_x)^\top D^{-1}(\psi)\\
&\quad +2\mu_x(q_i + c^{-1}\phi_x a_i)D^{-1}(\psi) + 2\frac{\phi_x^2}{c^2}a_i^2[D^{-1}(\psi) - 4\frac{\phi_x}{c}M_1],\\
d_{i\beta\phi_x} &= -2c^{-2}a_i A_i^\top,\\
d_{i\beta\phi^\top} &= 2[q_i\mathbb{I}_{(p)}D(Y_i - a - q_i b) + c^{-1}\phi_x A_i^\top (Y_i - a - bq_i)^\top D(b)]D^{-2}(\phi),\\
d_{i\phi_x\phi_x} &= 2\frac{c^{-3}}{\phi_x}(c-1)a_i^2,\\
d_{i\phi_x\phi^\top} &= (-2c^{-3}\phi_x a_i^2 D(b)D^{-2}(\phi)b + 2c^{-2}a_i D(b)D^{-2}(\phi)X_i)^\top,\\
d_{i\phi\phi^\top} &= 2D^{-3}(\phi)D^2(X_i) - 4c^{-1}\phi_x a_i D^{-3}(\phi)D(b)D(X_i)\\
&\quad -2c^{-1}\phi_x D^{-2}(\phi)D(b)X_i X_i^\top D(b)D^{-2}(\phi)\\
&\quad +2c^{-2}\phi_x^2 D^{-2}(\phi)D(b)X_i X_i^\top M D^{-1}(\phi)D(b)\\
&\quad +2c^{-2}\phi_x^2 a_i^2 D^{-3}(\phi)D^2(b) - 2c^{-3}\phi_x^3 a_i^2 D^{-1}(\phi)D(b)MD(b)D^{-1}(\phi)\\
&\quad +2c^{-2}\phi_x^2 D^{-1}(\phi)D(b)MX_i X_i^\top D(b)D^{-2}(\phi),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 a_i}{\partial\gamma\partial\tau^\top} &= 0, \gamma = \mu_x,\alpha,\phi_x,\lambda_x \quad \tau = \mu_x,\alpha,\phi_x,\lambda_x; \qquad \frac{\partial^2 a_i}{\partial\gamma\partial\tau^\top} = 0, \quad \gamma = \beta,\phi, \quad \tau = \phi_x,\lambda_x,\\
\frac{\partial^2 a_i}{\partial\mu_x\partial\beta^\top} &= -2\beta^\top D^{-1}(\psi),\\
\frac{\partial^2 a_i}{\partial\mu_x\partial\phi^\top} &= b^\top D(b)D^{-2}(\phi),\\
\frac{\partial^2 a_i}{\partial\alpha\partial\beta^\top} &= -D^{-1}(\psi),\\
\frac{\partial^2 a_i}{\partial\alpha\partial\phi^\top} &= \mathbb{I}_{(p)}D(b)D^{-2}(\phi),\\
\frac{\partial^2 a_i}{\partial\beta\partial\beta^\top} &= -2\mu_x D^{-1}(\psi),\\
\frac{\partial^2 a_i}{\partial\beta\partial\phi^\top} &= -\mathbb{I}_{(p)}D(Y_i - a - 2\mu_x b)D^{-2}(\phi),\\
\frac{\partial^2 a_i}{\partial\phi\partial\phi^\top} &= 2D(X_i)D(b)D^{-3}(\phi)
\end{aligned}
$$

where $A_i = (Y_{i2} - \alpha - 2q_i\beta)^\top D^{-1}(\psi)$, $M_1 = D^{-1}(\psi)\beta\beta^\top D^{-1}(\psi)$, $q_i = \mu_x + c^{-1}\phi_x a_i$ and $i = 1, ..., n$.

Appendix C: Computing the derivatives of $T_{mi}(\omega_i)$ and $S_{mi}(\omega_i)$.

We have that $T_{mi}(\omega_i) = -(\mathbf{y}_{mi}(\omega_i) - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{y}_{mi}(\omega_i)}{\partial \omega_i}$, $S_{mi}(\omega_i) = \mathbf{b}^\top D^{-1}(\boldsymbol{\phi}) \frac{\partial \mathbf{y}_{mi}(\omega_i)}{\partial \omega_i}$, then

$$
\begin{aligned}
\frac{\partial T_{mi}(\omega_i)}{\partial \mu_x} &= \mathbf{b}^{-1} \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{y}_{mi}(\omega_i)}{\partial \omega_i} \\
\frac{\partial T_{mi}(\omega_i)}{\partial \boldsymbol{\alpha}} &= \mathbf{I}_{(p)} \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{y}_{mi}(\omega_i)}{\partial \omega_i} \\
\frac{\partial T_{mi}(\omega_i)}{\partial \boldsymbol{\beta}^\top} &= q_{mi}(\omega_i) \frac{\partial \mathbf{y}_{mi}(\omega_i)}{\partial \omega_i}^\top (D^{-1}(\boldsymbol{\phi}) - 2c^{-1}\phi_x \mathbf{M}) \mathbf{I}_{(p)}^\top \\
&\quad + c^{-1}\phi_x \frac{\partial \mathbf{y}_{mi}(\omega_i)}{\partial \omega_i}^\top D^{-1}(\boldsymbol{\phi}) \mathbf{b} (\mathbf{y}_{mi}(\omega_i) - \mathbf{a})^\top D^{-1}(\boldsymbol{\phi}) \mathbf{I}_{(p)}^\top \\
\frac{\partial T_{mi}(\omega_i)}{\partial \phi_x} &= c^{-2} a_{mi}(\omega_i) \frac{\partial \mathbf{y}_{mi}(\omega_i)}{\partial \omega_i}^\top D^{-1}(\boldsymbol{\phi}) \mathbf{b} \\
\frac{\partial T_{mi}(\omega_i)}{\partial \boldsymbol{\phi}^\top} &= \frac{\partial \mathbf{y}_{mi}(\omega_i)}{\partial \omega_i}^\top \boldsymbol{\Sigma}^{-1} [D(\mathbf{X}_{mi}(\omega_i)) - c^{-1}\phi_x a_{mi}(\omega_i) D(\mathbf{b}))] \\
\frac{\partial T_{mi}(\omega_i)}{\partial \lambda_x} &= 0. \\
\frac{\partial S_{mi}(\omega_i)}{\partial \gamma} &= 0, \quad \gamma = \mu_x, \boldsymbol{\alpha}, \boldsymbol{\phi}, \phi_x, \\
\frac{\partial S_{mi}(\omega_i)}{\partial \boldsymbol{\beta}} &= \mathbf{I}_{(p)} D^{-1}(\boldsymbol{\phi}) \frac{\partial \mathbf{y}_{ki}(\omega_i)}{\partial \omega_i} \\
\frac{\partial S_{mi}(\omega_i)}{\partial \boldsymbol{\phi}} &= -D(\mathbf{b}) D^{-2}(\boldsymbol{\phi}) \frac{\partial \mathbf{y}_{ki}(\omega_i)}{\partial \omega_i}.
\end{aligned}
$$

Acknowledgements

The authors acknowledges the partial financial support from CNPq and Fapesp, Brasil and Projects Fondecyt 1030588, FANDES C-13955/10, CIMFAV, Chile.

References

Arellano-Valle, R.B. and Genton, M. G. (2005). Fundamental skew distributions. To appear in *Journal of Multivariate Analysis.*

Arellano-Valle R.B., Bolfarine, H. and Lachos, V.H. (2005a). Skew-normal linear mixed models. To appear in *Journal of Data Science.*

Arellano-Valle, R.B., Ozan S., Bolfarine, H. and Lachos, V.H. (2005b). Skew normal measurement error models. To appear in *Journal of Multivariate Analysis.*

Arnold, B.C., Beaver, R. J., Groeneveld, R. A. and Meeker, W. Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika,* 58, 471-478.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics,* 12, 171-178.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica,* XLVI, 199-208.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. Submitted.

Azzalini, A. and Capitanio, A. (1999). Statistical Applications of the Multivariate Skew Normal Distributions. *Journal Royal Statistics Society,* 61, 579-602.

Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika,* 83, 715-726.

Barnett, V.D. (1969). Simultaneous pairwise linear structural relationships. *Biometrics,* 25, 129-142.

Bolfarine, H. and Galea-Rojas, M.(1995). Maximum likelihood estimation of simultaneous pairwise linear structural ralationships. *Biometrical Journal,* 37, 673-689.

Bolfarine, H., Cabral, C.R.B. and Paula, G.A. (2002). Distance tests under nonregular conditions: Applications to the comparative calibration model. *Journal of Statistics Computation and Simulation,* 72, 2, 125-140.

Branco, M. and Dey, D. (2001). A general class of multivariate skew-elliptical distribution. *Journal of Multivariate Analysis,* 79, 93-113.

Chatterjee, S. Hadi, A.S. (1988). *Sensivity Analysis in Linear Regression*, John Wiley. New York.

Chipkevitch, E., Nishimura, R., Tu, D. and 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, R.D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society*, B, 48, 133-169.

Cook, R.D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman and Hall. London.

Dempster, A.P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, B, 39, 1-22.

Diciccio, T. and Monti, A. C. (2004). Inferential aspects of the skew exponential power distribution. *Journal of the American Statistical Association*, 99, 439-450.

Escobar, E. and Meeker, W. (1992). Assessing influence in regression analysis with censored data. *Biometrics*, 48, 507-528.

Galea, M., Bolfarine, H. and de Castro, M. (2002). Local influence in comparative calibration models. *Biometrical Journal*, 44, 59-81.

Genton, M. (2004). *Skew-Elliptical Distributions and their Applications. A Journey Beyond Normality*, Chapman and Hall. London.

Henze, N., (1986). A probabilistic representation of the "skew normal" distribution. *Scand. J. Statist.*, 13, 271-275.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. 1. New York: Wiley.

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

Kelly, G. (1985). Use of the structural equations model in assessing the reliability of a new measurement technique. *Applied Statistics*, 34, 258-263.

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

Nel, D. G. (1980). On matrix differentiation in statistics. *South African Statistical Journal*, 14, 137-193.

O'Hagan, A. and Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 16, 201-202.

Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, 27, 759-770.

Sahu, S.K., Dey, D.K., and Branco, M. (2003). A new class of multivariate distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics*, 29, 129-150.

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

Tanaka, Y., Watadani, S. and Moon, S. (1991). Influence in covariance structure analysis with an application to confirmatory factor analysis. *Communications in Statistics - Theory and Methods*, 20, 3805-3821.

Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer. New York.

Wu, X. and Luo, Z. (1993). Second-order approach to local influence. *Journal of the Royal Statistical Society,* B, 55, 929-936.

Zhao, Y. and Lee, A. (1998). Influence diagnostics for simultaneous equations models. *Australian and New Zealand Journal Statistics*, 40, 345-357.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2005-01 - DE SOUZA BORGES,W., GUSTAVO ESTEVES, L., WECHSLER, S. Process Parameters Estimation in The Taguchi On-Line Quality Monitoring Procedure For Attributes. 2005.19p. (RT-MAE-2005-01)

2005-02 - DOS ANJOS,U.,KOLEV, N. Copulas with Given Nonoverlapping Multivariate Marginals. 2005.09p. (RT-MAE-2005-02)

2005-03 - DOS ANJOS,U.,KOLEV, N. Representation of Bivariate Copulas via Local Measure of Dependence. 2005.15p. (RT-MAE-2005-03)

2005-04 - BUENO, V. C., CARMO, I. M. A constructive example for active redundancy allocation in a k-out-of-n:F system under dependence conditions. 2005. 10p. (RT-MAE-2005-04)

2005-05 - BUENO, V. C., MENEZES, J.E. Component importance in a modulated Markov system. 2005. 11p. (RT-MAE-2005-05)

2005-06 - BASAN, J. L., BRANCO, M.D´E., BOLFARINE, H. A skew item response model. 2005. 20p. (RT-MAE-2005-06)

2005-07 - KOLEV, N., MENDES, B. V. M., ANJOS, U. Copulas: a Review and recent developments. 2005. 46p. (RT-MAE-2005-07)

2005-08 - VENEZUELA, M. K., BOTTER, D. A., SANDOVAL, M. C. Diagnostic techniques in generalized estimating equations. 2005. 16p. (RT-MAE-2005-08)

2005-09 - BOLFARINE, H., LACHOS, V.H. Skew-probit measurement error models. 2005. 10p. (RT-MAE-2005-09)

2005-10 - SALINAS, V.H., ROMEO, J.S., PEÑA, A. On Bayesian estimation of a survival curve: Comparative study and examples. 2005. 14p. (RT-MAE-2005-10)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05311-970 - São Paulo, Brasil