UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação
ISSN 0103-2569

Skew-Normal Calibration Comparative Models

Vicente G. Cancho
Edwin M. Ortega
Víctor H. Lachos

Nº 270

RELATÓRIOS TÉCNICOS

São Carlos – SP
Fev./2006

Skew-Normal Calibration Comparative Models

Vicente G. Cancho*

ICMC, Universidade de São Paulo, São Carlos, SP, Brazil

Edwin. M. Ortega

ESALQ, Universidade de São Paulo, Piracicaba, SP, Brazil

Víctor H. Lachos

BAYES-FORECAST, SP, Brazil.

Abstract

In this paper we discussed inference aspects of the skew-normal calibration comparative models (SN-CCM) following both, a classical and Bayesian approach, extending the usual normal calibration comparative models (N-CCM) in order to avoid data transformation. To the proposed model we consider the maximum likelihood approach to estimation via the EM-algorithm and derive the observed information matrix allowing direct inference implementation, then we conduct the Bayesian approach via Markov chain Monte Carlo procedure. The univariate skew-normal distribution that will be used in this work was introduced by Sahu et al. (2003) which has a shape parameter that defines the direction of the asymmetric of the distribution, usually called the skewness parameter. Sahu's skew-normal distribution is attractive because estimation of the skewness parameter does not present the same difficult as is the case with Azzalini's (1985) one. The procedures are illustrated with a numerical example.

Key Words: *Sahu's skew-normal distribution; EM algorithm; calibration comparative models; skewness; MCMC.*

*Address for correspondence: Vicente Garibay Cancho, Departamento de Matemtica Aplicada e Estatística, ICMC, Universidade de São paulo, Caixa Postal 668, CEP 13560-970 São Carlos, São Paulo, Brazil. E-mail: garibay@icmc.usp.br

1 Introduction

The development of parametric families and the study of their properties have ever been a persistent theme of the statistical literature. A substantial part of the recent literature is broadly related to the skew-normal distribution, which represents a superset of the normal family and has a shape parameter that defines the direction of the asymmetry of the distribution. Advantages of using such general structures, in practice, include easiness of interpretation, as well as estimation efficiency. The main object of this paper is to consider estimation procedures, by using the EM-algorithm and MCMC methodology, in comparative calibration models where the main interest is to compare different ways of measuring the same unknown *skewed* quantity x in a group of n experimental units. The proposed model (SN-CCM) is an extension of the one (N-CCM) originally proposed by Barnett (1969) and whose application has been considered in several areas, especially in industry and biometry. 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). Examples in agriculture are consider in Fuller (1987) and in psychology and education in Dunn (1992), all those in a symmetric context. A viable alternative is to try fitting an asymmetric-normal model, which has been the subject of intense development in recent statistical literature. In this paper we use a real data studied in Barnett (1969) where the variable of interest is not symmetrically distributed and it seem better to fit a skew-normal calibration comparative model.

The model typically considered in the literature is defined as follows. Let x_i denotes the true value of the unknown quantity corresponding to individual (sample unit) i, and y_{ij}, the measurement that follows by using instrument j. A model adequate for such situations and that considering a linear relationship between y_{ij} and x_i is given (Barnett, 1969, Kimura, 1992) by $y_{ij} = \alpha_j + \beta_j x_i + \epsilon_{ij}$. The measurement errors ϵ_{ij} are assumed to be mutually independent $N_1(0, \phi_j)$, and also independent of the true values $x_1, \ldots, x_n$, which are also a random sample from a normal distribution with mean μ_x and variance ϕ_x $(N_1(\mu_x, \phi_x))$. Specifically in this work, we extend this symmetric (normal) model by considering that the unobserved quantity x_i (and consequentially the observed quantities) follow a skew-normal distribution. An interesting results of this work is that closed form expressions are obtained for all the proposed procedures, namely, the equations in the M-step of the EM-algorithm,

the conditional posterior distributions in the Gibbs sampler and also an analytical expression for the observed information matrix.

The paper is organized as follows. In Section 2 the skew-normal distribution is reviwed in a univariate context, with discussion of some of its properties. In Section 3 we present the asymmetric model, the marginal likelihood function of the observed data is derived in closed form and then the score function and the observed information matrix are also obtained. In Section 4 an EM-type algorithm for maximum likelihood estimation is developed by exploring statistical properties of the proposed model. In Section 5 a Bayesian method is considered for SN-CCM, including a discussion of model choice via a Bayes factor approach. Finally, Section 6 gives an applications to a real data set previously analyzed in the literature, including a comparison between the normal and skew-normal CCM.

2 The univariate skew-normal distribution

As considered in Sahu et al. (2003), a random variable Z is a skew-normal with location parameter μ, scale parameter σ^2 and skewness parameter λ, if the density function of Z is given by

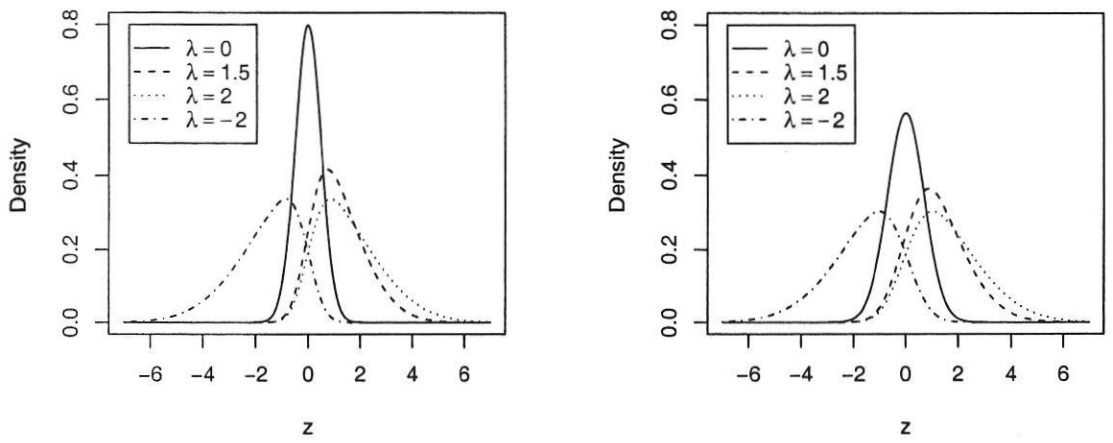
$$f_Z(z) = 2\phi_1(z|\mu, \sigma^2 + \lambda^2)\Phi_1\left(\frac{\lambda}{\sigma}\frac{(z-\mu)}{(\sigma^2+\lambda^2)^{1/2}}\right), \tag{1}$$

where $\phi_1(.|a, b^2)$ and $\Phi_1(.|a, b^2)$ denote, respectively, the p.d.f. and c.d.f of the $N_1(a, b^2)$. When $a = 0$ and $b = 1$ we denoted these functions by $\phi_1(.)$ and $\Phi_1(.)$, respectively. We use the notation $Z \sim SN_1(\mu, \sigma^2, \lambda)$ to denote this class of distribution, that will be reduced to $SN_1(\lambda)$ when it is assumed that $\mu = 0$ and $\sigma^2 = 1$. The stochastic representation of a skew-normal random variable, which can be used to simulate quickly pseudo realizations from $SN_1(\mu, \sigma^2, \lambda)$, is given by

$$Z \stackrel{\mathrm{d}}{=} \lambda|X_o| + X_1, \tag{2}$$

where $X_o \sim N_1(0, 1)$ and $X_1 \sim N_1(\mu, \sigma^2)$. It follows that the SN_1 distribution is unimodal and if $\lambda < 0$ the distribution has a negative skewness and if $\lambda > 0$ the skewness is positive. In Figure 1, we illustrate the asymmetric shape of the $SN_1(0, \sigma^2, \lambda)$ for different values of σ^2 and λ. Note that these densities can be strongly asymmetric depending on suitable choices of the parameters λ. Some important properties of a random variable $Z \sim SN_1(\mu, \sigma^2, \lambda)$ include:

Figure 1: Some members of the $SN_1(0, \sigma^2, \lambda)$ for $\lambda = 0, 1.5, 2, -2$. (a) $\sigma^2 = 0.25$ and (b) $\sigma^2 = 0.5$.



(P1) $E[Z] = \mu + \sqrt{\frac{2}{\pi}}\lambda$, $Var[Z] = \sigma^2 + (1 - \frac{2}{\pi})\lambda^2$ and

$$E[Z^3] = 3\mu(\sigma^2 + \lambda^2) + \mu^3 + \sqrt{\frac{2}{\pi}}\lambda(3\sigma^2 + 2\lambda^2 + 3\mu^2).$$

(P2) The skewness and kurtosis indices are, respectively, given by

$$\gamma = \lambda^3\sqrt{\frac{2}{\pi}}(\frac{4}{\pi} - 1)\left(\sigma^2 + (1 - \frac{2}{\pi})\lambda^2\right)^{-3/2}, \quad \kappa = \left(8\lambda^4\frac{(\pi - 3)}{\pi^2}\right)\left(\sigma^2 + (1 - \frac{2}{\pi})\lambda^2\right)^{-2}.$$

(P3) $-Z \sim SN_1(\mu, \sigma^2, -\lambda)$.

(P4) $Z \sim SN_1(\mu, \sigma^2, \lambda)$ if and only if $Z \sim SN_A(\mu, \sigma^2 + \lambda^2, \lambda/\sigma)$, where SN_A denotes the skew-normal distribution introduced by Azzalini (1985).

(P5) Let $Z_1, \ldots, Z_n$ be a random sample from Z, then the moments estimator of μ, σ^2 and λ, are, respectively, given by

$$\hat{\mu} = \bar{Z} - \left(\frac{2}{\pi}\right)^{-1/6} m_3^{1/3}\left[\frac{4}{\pi} - 1\right]^{-1/3}, \tag{3}$$

$$\hat{\sigma}^2 = m_2 - \left(1 - \frac{2}{\pi}\right)\left(\frac{2}{\pi}\right)^{-1/3} m_3^{2/3}\left[\left(\frac{4}{\pi} - 1\right)\right]^{-2/3}, \tag{4}$$

$$\hat{\lambda} = m_3^{1/3}\left[\sqrt{\frac{2}{\pi}}(\frac{4}{\pi} - 1)\right]^{-1/3}, \tag{5}$$

where $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$, $m_2 = \frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$ and $m_3 = \frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z})^3$.

(P6) The simpler form of Z, i.e, $Z \sim SN_1(\lambda)$ has some interesting properties, for instance,

$$E[Z^{2k+1}] = \sqrt{(2/\pi)}\lambda \frac{(2k+1)!}{2^k} \sum_{j=0}^{k} \frac{j!(2\lambda)^{(2k)}}{(2j+1)!(k-j)!}, \; E[Z^{2k}] = (1+\lambda^2)^k \frac{(2k-1)!}{2^{k-1}(k-1)!}$$

and

$$P(Z \leq z) = 2\Phi_2(\mathbf{w}|\mathbf{0}, \boldsymbol{\Omega}), \text{ with } \mathbf{w} = (z, 0)^\top, \; \boldsymbol{\Omega} = \begin{pmatrix} (1+\lambda^2) & -\lambda\sqrt{1+\lambda^2} \\ -\lambda\sqrt{1+\lambda^2} & (1+\lambda^2) \end{pmatrix}$$

From (P1) it can be seen that the skewness parameter is always finite, this not is the case with Azzalini's skew-normal distribution for which estimation of the parameters is not a trivial problem. Despite its nice properties, this skew-normal family has some inferential problems. For instance, can be shown that the observed information matrix is singular when $\lambda = 0$.

3 The skew-normal calibration comparative model

As discussed in the introduction, the general comparative calibration model is defined by considering that

$$y_{ij} = \alpha_j + \beta_j x_i + \epsilon_{ij}, \tag{6}$$

$i = 1, \ldots, n$, $j = 1, \ldots, p$, with $\alpha_1 = 0$ and $\beta_1 = 1$, i.e., as in Barnett (1969) we are considering the existence of a reference instrument (instrument 1) which makes unbiased measurement. The situation where $x_i \overset{iid}{\sim} SN_1(\mu_x, \phi_x, \lambda_x)$ and $\epsilon_{ij} \overset{ind}{\sim} N_1(0, \phi_j)$, with x_i independent of ϵ_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, p$ will be called of SN-CCM. Since x_i is considered as a random variable, the model we consider is a structural model (Bolfarine and Galea-Rojas, 1995). The above model consider, for instance, that in the case of Barnett's (1969) data set, vital capacity is not symmetrically distributed in the population. On the other hand, the errors ϵ_{ij}, are related to measurement errors so that it reasonable to expected them to be normally distributed. The asymmetric parameter λ_x incorporates asymmetry in the latent variable x_i and consequently in the observed quantities y_{ij}. If $\lambda_x = 0$, then the asymmetric model reduces to the normal calibration comparative model (N-CCM) considered in Barnett (1969) which has been extensively treated in the literature.

Letting $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^\top$, $\mathbf{a} = (0, \boldsymbol{\alpha}^\top)^\top = (0, \alpha_2, \ldots, \alpha_p)^\top$, $\mathbf{b} = (1, \boldsymbol{\beta}^\top)^\top = (1, \beta_2, \ldots, \beta_p)^\top$ and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})^\top$, this model can be written in matrix form as

$$\mathbf{y}_i = \mathbf{a} + \mathbf{b}x_i + \boldsymbol{\epsilon}_i, \tag{7}$$

where

$$\boldsymbol{\epsilon}_i \overset{iid}{\sim} N_p(\mathbf{0}, D(\boldsymbol{\phi})) \text{ and } x_i \overset{iid}{\sim} SN_1(\mu_x, \phi_x, \lambda_x), \tag{8}$$

$i = 1, \ldots, n$, with $D(\boldsymbol{\phi}) = diag(\boldsymbol{\phi})$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$. Note from (2) that the regression set up defined in (7)-(8) can be written hierarchically as

$$\mathbf{y}_i \mid x_i \overset{ind}{\sim} N_p(\mathbf{a} + \mathbf{b}x_i, D(\boldsymbol{\phi})), \tag{9}$$

$$x_i \mid T_i = t_i \overset{ind}{\sim} N_1(\mu_x + \lambda_x t_i, \phi_x), \tag{10}$$

$$T_i \overset{iid}{\sim} HN_1(0, 1), \tag{11}$$

$i = 1, \ldots, n$, all independent, where $HN_1(0, 1)$ denote the standardized univariate half-normal distribution (Johnson et al., 1994). Integrating out the $x_i's$ it follows that the marginal distribution for the responses $\mathbf{y}_i$ is given by

$$f_{\mathbf{Y}_i}(\mathbf{y}_i|\boldsymbol{\theta}) = 2\phi_p(\mathbf{y}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(A_x a_i), \tag{12}$$

where $\boldsymbol{\mu} = \mathbf{a} + \mathbf{b}\mu_x$, $\boldsymbol{\Sigma} = D(\boldsymbol{\phi}) + \phi_c \mathbf{b}\mathbf{b}^\top$, $A_x = \dfrac{\lambda_x \Lambda_x}{\sqrt{\phi_x \phi_c + \lambda_x^2 \Lambda_x}}$, $a_i = \mathbf{b}^\top D^{-1}(\boldsymbol{\phi})(\mathbf{y}_i - \boldsymbol{\mu})$, $\Lambda_x = (\phi_c{}^{-1} + \mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b})^{-1} = \dfrac{\phi_c}{c}$, with $\phi_c = \phi_x + \lambda_x^2$, $c = 1 + \phi_c \mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}$ and $i = 1, \ldots, n$.

Notice that the marginal distribution in (12) is in the class of fundamental skew-normal distributions introduced by Arellano-Valle and Genton (2005). The log-likelihood function for $\boldsymbol{\theta} = (\mu_x, \boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \phi_x, \boldsymbol{\phi}^\top, \lambda_x)^\top$ given the observed sample $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ is given by

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

where $\ell_i(\theta) = \log(2) - (p/2)\log(2\pi) - \dfrac{1}{2}\log|\boldsymbol{\Sigma}| - \dfrac{1}{2}d_i + \log(K_i)$ with $d_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$ and $K_i = \Phi_1(A_x a_i)$, $i = 1, \ldots, n$.

3.1 The score function and the observed information matrix

Asymptotic standard errors for parameters can be calculated by inverting the expected or observed information matrix. Many authors prefer to use the observed information matrix. In the following we indicating how to compute the score vector and the observed information matrix. Thus, from (13) follows that the score function is of the form

$$U(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} U_i(\boldsymbol{\theta}), \tag{14}$$

where $U_i(\theta) = \dfrac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = (U_i(\boldsymbol{\beta})^\top, U_i(\boldsymbol{\alpha})^\top, U_i(\boldsymbol{\phi})^\top, U_i(\mu_x), U_i(\phi_x), U_i(\lambda_x))^\top$ and

$$U_i(\boldsymbol{\gamma}) = \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} = -\frac{1}{2}\frac{\partial \log|\Sigma|}{\partial \boldsymbol{\gamma}} - \frac{1}{2}\frac{\partial d_i}{\partial \boldsymbol{\gamma}} + W_{\Phi_1}(A_x a_i)\left[A_x \frac{\partial a_i}{\partial \boldsymbol{\gamma}} + a_i \frac{\partial A_x}{\partial \boldsymbol{\gamma}}\right], \tag{15}$$

$i = 1, \ldots, n$, with $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$, $u \in \mathbb{R}$, $\boldsymbol{\gamma} = \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\phi}, \mu_x, \phi_x, \lambda_x$.

The matrix of second derivatives with respect to $\boldsymbol{\theta}$ is given by

$$\mathbf{L}(\boldsymbol{\theta}) = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \sum_{i=1}^{n} \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}. \tag{16}$$

It follows from (15) that the observed, per element, information matrix is given by

$$\mathbf{J}_i = -\mathbf{L}_i(\boldsymbol{\theta}) - \left[\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top}\right], \tag{17}$$

where $\dfrac{\partial^2 \ell_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} = -\dfrac{1}{2}\dfrac{\partial^2 \log|\Sigma|}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} - \dfrac{1}{2}\dfrac{\partial^2 d_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} + \dfrac{\partial^2 \log K_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top}$, with

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

$\Delta_{\Phi_1}(u) = W'_{\Phi_1}(u) = -W_{\Phi_1}(u)(u + W_{\Phi_1}(u))$, $u \in \mathbb{R}$ and $\boldsymbol{\gamma}, \boldsymbol{\tau} = \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\phi}, \mu_x, \phi_x, \lambda_x$. Closed form expressions for the derivatives above are given in the Appendix.

In the next section we discuss the MLE of the parameters vector $\boldsymbol{\theta}$ for SN-CCM using the EM algorithm.

4 Maximum likelihood estimation

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 SN-CCM using double augmentation, by considering that

$(\mathbf{x}, \mathbf{t})$ are missing data. Thus, under the hierarchical representation (9)-(11) it follows that the complete log-likelihood function associated with $(\mathbf{y}, \mathbf{x}, \mathbf{t})$ is

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}, \mathbf{t}) = -\frac{n}{2}\log|D(\boldsymbol{\phi})| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{a} - \mathbf{b}x_i)^\top D^{-1}(\boldsymbol{\phi})(\mathbf{y}_i - \mathbf{a} - \mathbf{b}x_i)$$
$$-\frac{n}{2}\log(\phi_x) - \frac{1}{2\phi_x}\sum_{i=1}^{n}(x_i - \mu_x - \lambda_x t_i)^2 + c. \tag{18}$$

where c is a constant that is independent of the parameter vector $\boldsymbol{\theta}$.

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

$$\begin{aligned}
\widehat{t}_i &= \widehat{\mu}_{T_i} + W_{\Phi_1}(\frac{\widehat{\mu}_{T_i}}{\widehat{M}_T})\widehat{M}_T, \\
\widehat{t^2}_i &= \widehat{\mu}^2_{T_i} + \widehat{M}^2_T + W_{\Phi_1}(\frac{\widehat{\mu}_{T_i}}{\widehat{M}_T})\widehat{M}_T\widehat{\mu}_{T_i}, \\
\widehat{x}_i &= \widehat{r}_i + \widehat{s}\,\widehat{t}_i, \\
\widehat{x^2}_i &= \widehat{T^2}_x + \widehat{r}_i^2 + 2\widehat{r}_i\,\widehat{s}\,\widehat{t}_i + \widehat{s}^2\,\widehat{t^2}_i, \quad \text{and} \\
\widehat{tx}_i &= \widehat{r}_i\,\widehat{t}_i + \widehat{s}\,\widehat{t^2}_i
\end{aligned} \tag{19}$$

where $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$, $\widehat{M}^2_T = [1+\widehat{\lambda}^2_x\mathbf{b}^\top(D(\widehat{\boldsymbol{\phi}})+\widehat{\phi}_x\widehat{\mathbf{b}}\widehat{\mathbf{b}}^\top)^{-1}\widehat{\mathbf{b}}]^{-1}$, $\mu_{T_i} = \lambda_x M^2_T\widehat{\mathbf{b}}^\top(D(\boldsymbol{\phi})+\widehat{\phi}_x\widehat{\mathbf{b}}\widehat{\mathbf{b}}^\top)^{-1}(\mathbf{y}_i-\widehat{\mathbf{a}}-\widehat{\mathbf{b}}\widehat{\mu}_x)$, $\widehat{T^2}_x = \widehat{\phi}_x[1+\widehat{\phi}_x\widehat{\mathbf{b}}^\top D^{-1}(\widehat{\boldsymbol{\phi}})\widehat{\mathbf{b}}]^{-1}$, $\widehat{r}_i = \widehat{\mu}_x+\widehat{T^2}_x\widehat{\mathbf{b}}^\top D^{-1}(\widehat{\boldsymbol{\phi}})(\mathbf{y}_i-\widehat{\mathbf{a}}-\widehat{\mathbf{b}}\widehat{\mu}_x)$ and $\widehat{s} = \widehat{\lambda}_x(1 - \widehat{T^2}_x\widehat{\mathbf{b}}^\top D^{-1}(\widehat{\boldsymbol{\phi}})\widehat{\mathbf{b}})$.

Using a simple algebra we find that

$$E[\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}, \mathbf{t})|\mathbf{y}, \widehat{\boldsymbol{\theta}}] = -\frac{n}{2}\log|D(\boldsymbol{\phi})|$$
$$-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{a} - \mathbf{b}\widehat{x}_i)^\top D^{-1}(\boldsymbol{\phi})(\mathbf{y}_i - \mathbf{a} - \mathbf{b}\widehat{x}_i) - \frac{1}{2}\mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}\sum_{i=1}^{n}(\widehat{x^2}_i - \widehat{x}_i^2)$$
$$-\frac{n}{2}\log(\phi_x) - \frac{1}{2\phi_x}\sum_{i=1}^{n}(\widehat{x^2}_i + \mu_x^2 + \lambda_x^2\widehat{t^2}_i - 2\widehat{x}_i\mu_x - 2\widehat{xt}_i\lambda_x + 2\lambda_x\mu_x\widehat{t}_i) + c. \tag{20}$$

We then have the following EM algorithm:

E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, compute $\widehat{t}_i$, $\widehat{t^2}_i$, $\widehat{x}_i$, $\widehat{x^2}_i$ and $\widehat{xt}_i$ for $i = 1, \ldots, n$, using (19).

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

$$\begin{aligned}
\widehat{\alpha}_j &= \bar{y}_j - \bar{\widehat{x}}\beta_j, \;\; j = 2,\ldots,p, \\
\widehat{\beta}_j &= \frac{\sum_{i=1}^n \widehat{x}_i(y_{ij}-\bar{y}_j)}{\sum_{i=1}^n \widehat{x^2}_i - n(\bar{\widehat{x}})^2}, \;\; j = 2,\ldots,p, \\
\widehat{\phi}_1 &= \frac{1}{n}\sum_{i=1}^n (y_{i1}^2 - 2\widehat{x}_i y_{i1} + \widehat{x^2}_i), \\
\widehat{\phi}_j &= \frac{1}{n}\sum_{i=1}^n (y_{ij}^2 + \alpha_j^2 + \beta_j^2\widehat{x^2}_i - 2\alpha_j y_{ij} - 2y_{ij}\beta_j\widehat{x}_i + 2\alpha_j\beta_j\widehat{x}_i), \\
\widehat{\mu}_x &= \frac{1}{n}\sum_{i=1}^n (\widehat{x}_i - \lambda_x\widehat{t}_i), \\
\widehat{\phi}_x &= \frac{1}{n}\sum_{i=1}^n (\widehat{x^2}_i + \mu_x^2 + \lambda_x^2\widehat{t^2}_i - 2\mu_x\widehat{x}_i - 2\lambda_x\widehat{tx}_i + 2\lambda_x\mu_x\widehat{t}_i), \;\; \text{and} \\
\widehat{\lambda}_x &= \frac{\sum_{i=1}^n (\widehat{tx}_i - \mu_x\widehat{t}_i)}{\sum_{i=1}^n \widehat{t^2}_i},
\end{aligned} \tag{21}$$

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 $j = 2,\ldots,p$.

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, $-\ell(\widehat{\boldsymbol{\theta}})/N + P/N$), Schawarz's Bayesian Information Criterion (BIC, $-\ell(\widehat{\boldsymbol{\theta}})/N + 0.5\log(N)P/N$), and the Hannan-Quinn Criterion (HQ, $-\ell(\widehat{\boldsymbol{\theta}})/N + \log(\log(N))P/N$), 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 N-CCM and SN-CCM fits. Note that if $\lambda_x = 0$ the M-step equations reduce to the equations given by Bolfarine and Galea-Rojas (1995).

5 Bayesian analysis

An integral part of Bayesian analysis is the assignment of prior distributions to all unknowns in the model. In order to guarantee proper posteriors, we adopt proper priors with known hyper-parameters. Thus, we take a normal prior distribution for the elements in $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$,

with densities function given by

$$\pi(\beta_j|\mu_\beta,\sigma_\beta^2) \propto \exp\left\{-\frac{1}{2\sigma_\beta^2}(\beta_j-\mu_\beta)^2\right\},$$

$$\pi(\alpha_j|\mu_\alpha,\sigma_\alpha^2) \propto \exp\left\{-\frac{1}{2\sigma_\alpha^2}(\alpha_j-\mu_\alpha)^2\right\},\ \ j=2,\ldots,p.$$

The scales parameters ϕ_j are taken to be $IG(\frac{\tau_e}{2},\frac{T_e}{2})$ (inverse gamma) with densities functions given by

$$\pi(\phi_j|\tau_e,T_e) \propto \left(\frac{1}{\phi_j}\right)^{\frac{\tau_e}{2}+1}\exp\left\{-\frac{T_e}{2\phi_j}\right\},\ \ j=1,\ldots,p.$$

The family of prior distributions for the parameters of the latent variable μ_x, ϕ_x and λ_x were chosen mainly for computational simplicity. Thus, we use, respectively, the normal, the inverse gamma and the truncated normal distributions with densities functions, respectively, given by

$$\begin{aligned}\pi(\mu_x|\varsigma,\nu^2) &\propto \exp\left\{-\frac{1}{2\nu^2}(\mu_x-\varsigma)^2\right\},\\ \pi(\phi_x|\tau_x,T_x) &\propto \left(\frac{1}{\phi_x}\right)^{\frac{\tau_x}{2}+1}\exp\left\{-\frac{T_x}{2\phi_x}\right\},\end{aligned}$$

and

$$\pi(\lambda_x|\mu_\lambda,\sigma_\lambda^2) \propto \exp\left\{-\frac{1}{2\sigma_\lambda^2}(\lambda_x-\mu_\lambda)^2\right\}\mathbb{I}_{\{\lambda_x>0\}}.$$

where $\mathbb{I}_{\{A\}}$ is the indicator of A. Similar prior distribution for λ_x can be defined in the situation where we have negative skewness.

Assuming prior independence among the parameters, let $\pi(\boldsymbol{\theta})$ denote a prior distribution for $\boldsymbol{\theta}=(\mu_x,\boldsymbol{\alpha}^\top,\boldsymbol{\beta}^\top,\phi_x,\boldsymbol{\phi}^\top,\lambda_x)^\top$, with

$$\pi(\theta) \propto \prod_{j=1}^{p}\left(\pi(\beta_j|\mu_\beta,\sigma_\beta^2)\pi(\alpha_j|\mu_\alpha,\sigma_\alpha^2)\pi(\phi_j|\tau_e,T_e)\right)\pi(\mu_x|\varsigma,\nu^2)\pi(\phi_x|\tau_x,T_x)\pi(\lambda_x|\mu_\lambda,\sigma_\lambda^2),$$

then from the hierarchical representation given in (9)-(11) the join density of $\boldsymbol{\theta}$, $\mathbf{y}=(\mathbf{y}_1^\top,\ldots,\mathbf{y}_n^\top)^\top$, $\mathbf{x}=(x_1,\ldots,x_n)^\top$ and $\mathbf{t}=(t_1,\ldots,t_m)^\top$, is proportional to

$$\pi(\theta)\prod_{i=1}^{n}\phi_p(\mathbf{y}_i|\mathbf{a}+\mathbf{b}x_i,D(\boldsymbol{\phi}))\phi_1(x_i|\mu_x+\lambda_x t_i,\phi_x)\phi_1(t_i|0,1)\mathbb{I}_{\{t_i>0\}}. \tag{22}$$

Distribution (22) is not tractable analytically but MCMC methods such as the Gibbs sampler, can be used to draw samples, from which features of marginal posterior distributions

of interest can be inferred. In situations where those distributions are simple to sample from, as in the present case, the approach is easily implemented. In other situations, the more complex Metropolis-Hastings approach needs to be considered (see Gamerman, 1997). All the full conditional distributions required to implement the Gibbs sampler are straightforward to derive by working the complete joint posterior using, for example, Lemma 2 in Arellano-Valle et al. (2005a). These have closed form and are given by

$$\beta_j|... \sim N_1(A_{\beta_j}^{-1}\varphi_{\beta_j}, A_{\beta_j}^{-1}), \quad j = 2, \ldots, p, \tag{23}$$

with $A_{\beta_j} = \dfrac{1}{\sigma_\beta^2} + \dfrac{\sum_{i=1}^n x_i^2}{\phi_j}$ and $\varphi_{\beta_j} = \dfrac{\mu_\beta}{\sigma_\beta^2} + \dfrac{\sum_{i=1}^n (y_{ij} - \alpha_j)x_i}{\phi_j}$;

$$\alpha_j|... \sim N_1(A_{\alpha_j}^{-1}\varphi_{\alpha_j}, A_{\alpha_j}^{-1}), \quad j = 2, \ldots, p, \tag{24}$$

with $A_{\alpha_j} = \dfrac{1}{\sigma_\alpha^2} + \dfrac{n}{\phi_j}$ and $\varphi_{\alpha_j} = \dfrac{\mu_\alpha}{\sigma_\alpha^2} + \dfrac{\sum_{i=1}^n (y_{ij} - \beta_j x_i)}{\phi_j}$;

$$\mu_x|... \sim N_1(A_{\mu_x}^{-1}\varphi_{\mu_x}, A_{\mu_x}^{-1}), \tag{25}$$

with $A_{\mu_x} = \dfrac{1}{\nu^2} + \dfrac{n}{\phi_x}$ and $\varphi_{\mu_x} = \dfrac{\varsigma}{\nu^2} + \dfrac{\sum_{i=1}^n (x_i - \lambda_x t_i)}{\phi_x}$;

$$\lambda_x|... \sim N_1(A_{\lambda_x}^{-1}\varphi_{\lambda_x}, A_{\lambda_x}^{-1})\mathbb{I}_{\{\lambda_x > 0\}}, \tag{26}$$

with $A_{\lambda_x} = \dfrac{1}{\sigma_\lambda^2} + \dfrac{\sum_{i=1}^n t_i^2}{\phi_x}$ and $\varphi_{\lambda_x} = \dfrac{\mu_\lambda}{\sigma_\lambda^2} + \dfrac{\sum_{i=1}^n (x_i - \mu_x)t_i}{\phi_x}$;

$$\phi_j|... \sim IG(\frac{\tau_e + n}{2}, \frac{T_e + \sum_{i=1}^n (y_{ij} - \alpha_j - \beta_j x_i)^2}{2}), \quad j = 1, \ldots, p; \tag{27}$$

$$\phi_x|... \sim IG(\frac{\tau_x + n}{2}, \frac{T_x + \sum_{i=1}^n (x_i - \mu_x - \lambda_x t_i)^2}{2}); \tag{28}$$

$$x_i|... \sim N_1(\varphi_{x_i}, A_x^{-1}), \quad i = 1, \ldots, n, \tag{29}$$

with $A_x = \dfrac{1}{\phi_x} + \mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}$ and $\varphi_{x_i} = \mu_x + \lambda_x t_i + A_x^{-1}\mathbf{b}D^{-1}(\boldsymbol{\phi})\left(\mathbf{y}_i - \mathbf{a} - \mathbf{b}(\mu_x + \lambda_x t_i)\right)$;

$$t_i|... \sim N_1(A_t^{-1}\varphi_{t_i}, A_{t_i}^{-1})\mathbb{I}_{\{t_i > 0\}}, \quad i = 1, \ldots, n, \tag{30}$$

with $A_t = 1 + \dfrac{\lambda_x^2}{\phi_x}$ and $\varphi_{t_i} = \dfrac{\lambda_x}{\phi_x}(x_i - \mu_x)$;

The notation "$\beta_j|...$" is used to denote the conditional distribution of β_j, $j = 2, \ldots, p$, given all the other parameters and similarly for all the other conditional distributions. The algorithm starts with initial values for all variables considered and cycles, generating samples from the above conditional distributions, until convergence, which can be verified using the approach developed by Gelman and Rubin (1992), for example, which requires running several parallel chains.

5.1 Bayesian model choice

Model determination is a fundamental issue in statistics. The literature on the model assessment or checking and model selection presents many approaches, beginning with the Bayes factor approach. Several modifications of the Bayes factor are presented in the literature (see for example, Aitkin, 1981; Berger and Perichi, 1996). Geisser and Eddy (1979) took a predictive approach based on cross validation methods to obtain the pseudo-Bayes factor. This approach is described next.

Consider a choice between two parametric models denoted by the joint density $f(\mathbf{t}; \boldsymbol{\theta}_i, M_i)$ or the likelihood function $L(\boldsymbol{\theta}_i; \mathbf{t}, M_i)$, $i = 1, 2$. Suppose that w_i is the prior probability of selecting the model M_i $i = 1, 2$ and $f(\mathbf{t}|M_i)$ is the predictive distribution for the model M_i, that is,

$$f(\mathbf{t}|M_i) = \int f(\mathbf{t}|\boldsymbol{\theta}_i, M_i)\pi(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta} \tag{31}$$

where $\pi(\theta_i|M_i)$ is the prior distribution under model M_i. If $\mathbf{t}_0$ denotes the observed data, then we choose the model yielding the larger $w_i f(\mathbf{t}_0|M_i)$.

Often we set $w_i = 0.5$, $i = 1, 2$ and compute the Bayes factor of the M_1 with respect to M_2 as

$$B_{12} = \frac{f(\mathbf{t}_0|M_1)}{f(\mathbf{t}_0|M_i)}. \tag{32}$$

The predictive distribution (31) can be approximated by its Monte Carlo estimate using S generated samples from the prior density $\pi(\boldsymbol{\theta}_i|M_i)$ that is,

$$\hat{f}(\mathbf{t}|M_i) = \frac{1}{S}\sum_{s=1}^{S} f(\mathbf{t}|\boldsymbol{\theta}_i^{(s)}, M_i) \tag{33}$$

where $\boldsymbol{\theta}_i^{(s)}$ denotes the vector for $\boldsymbol{\theta}_i$ draw in the sth sample. Some modifications of the estimate (35) for the predictive density are proposed in the literature (see, e.g. Newton and Raptery,

1994; Gelfand and Dey, 1994). For the model selection we could also consider the conditional predictive ordinate (CPO) (see, e.g. Gelfand and Dey, 1994), given by

$$f(t_r|\mathbf{t}, M_i) = \int f(t_r|\boldsymbol{\theta}_i, \mathbf{t}_{(r)}, M_i)\pi(\boldsymbol{\theta}_i|\mathbf{t}_{(r)}, M_i)d\boldsymbol{\theta}_i \tag{34}$$

where $\mathbf{t}_{(r)} = (t_1, \ldots, t_{r-1}, t_{r+1}, \ldots, t_n)$.

Using the generated Gibbs samples, equation (34) can be approximated by its Monte Carlo estimate,

$$\hat{f}(t_r|\mathbf{t}_{(r)}M_i) = \frac{1}{S}\sum_{s=1}^{S} f(t_r|\mathbf{t}_{(r)}, \boldsymbol{\theta}_i^{(s)}, M_i) \tag{35}$$

where $\boldsymbol{\theta}_i^{(s)}$ denotes the vector for $\boldsymbol{\theta}_i$ draw in the sth Gibbs sample. We can use the obtained estimates $c_r(l) = \hat{f}(t_r|\mathbf{t}_{(r)}M_i)$ in model selection. In this way, we consider plots of $c_r(l)$ versus r $(r = 1, \ldots, n)$ for different models; large values (on average) indicate the better model. An alternative is to choose the model for which $c(l) = \sum_{r=1}^{n} \log(c_r(l))$ (l index models). We choose the model with the largest $c(l)$ value. Geisser and Eddy (1979) suggest that the product of the predictive densities $\prod_{r=1}^{n} f(t_r|\mathbf{t}_{(r)}, M_i)$ could be used in model selection as a relative indicator. If we have two models M_1 and M_2, we have the ratio, which is called the pseudo Bayes factor,

$$PSBF = \frac{\prod_{r=1}^{n} f(t_r|\mathbf{t}_{(r)}, M_1)}{\prod_{r=1}^{n} f(t_r|\mathbf{t}_{(r)}, M_2)}. \tag{36}$$

as an approximation to the Bayes factor. Using samples generated by the MCMC process, the pseudo Bayes factor (36) can be estimated by using equation (35).

6 Application to the Barnett data set

In this section we apply the methods developed earlier to the Barnett (1969) data set where 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 improve numerical stability. In the literature, a data transformation was used to justify the use of the normal distribution (Bolfarine et al., 2002), the need for this can be seen in Figure 6. We compare N-CCM (extensively treated in

the literature), and SN-CCM fitting for this data set based on maximum likelihood estimation and Bayesian methods.

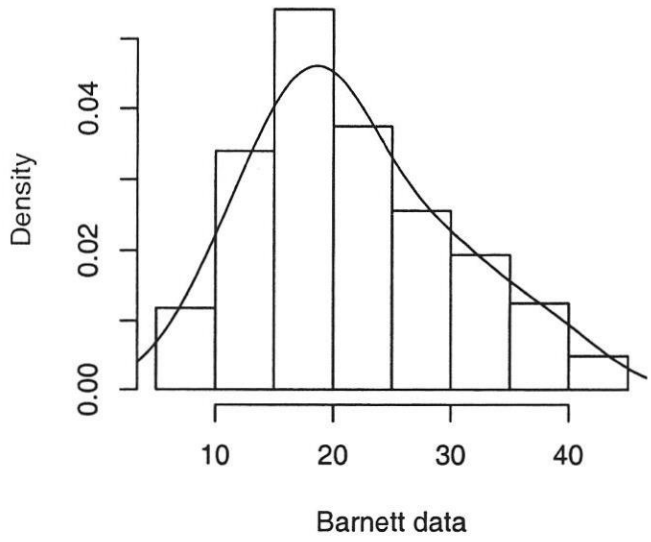


Figure 2: Barnett Data set. Histogram with kernel density estimated to the observed measurement.

In Table 1, we give the MLE of the parameters for the normal and skew-normal CCM. Note that the AIC, BIC and HQ values shown at the bottom of this table favor SN-CCM, supporting the observed departure from normality. A conclusion in this direction is also achieved by considering a parametric test for normality. The null hypothesis is $\lambda_x = 0$, and the likelihood ratio test of this hypothesis has the tests statistic 9.1773 . The associated critical level of $\chi^2(1)$ at 5% is 3.84.

The Bayes estimates were obtained using the software WinBUGS and the following independent prior specifications:

$$\alpha_j \sim N(0, 10^2),\ \beta_j \sim N(0, 10^2),\ j = 2, 3, 4,\ \phi_j \sim IG(0.01, 0.01),\ j = 1, \ldots, 4,$$
$$\mu_x \sim N(0, 10^2),\ \phi_x \sim IG(0.01, 0.01),\ \lambda_x \sim N(0, 10^2)\mathbb{I}_{\{\lambda_x > 0\}}.$$

Considering different initial values, two parallel independent runs of the Gibbs sampler chain

with size 30000 were run for each parameter, discarding the first 10000 iterations. To eliminate the effect of the initial values and to avoid correlation problems, we considered a spacing of size 10, obtaining a sample of size 2000 from each chain. To monitor the convergence of the Gibbs samples we used the between and within sequence information, following the approach developed in Gelman and Rubin (1992) to obtain the potential scale reduction , $\hat{R}$. In all cases, these values were close to one, indicating the convergence of the chain. In order to compare estimates with the N-CCM fit ($\lambda_x = 0$), In Table 2 we report posterior summaries for the parameters of the N-CCM and SN-CCM.

We also have be used some existing Bayesian measures of fit such as the DIC (deviance information criterion) as described in Spiegelhalter et al. (2002). In Table 2 we present estimates for DIC based on the simulated Gibbs samples. Note that SN-CCM is the best model based on the DIC criteria (smallest value). Further, it can be shown that the pseudo Bayes factor (36), of the SN-CCM (M_1) with respect to N-CCM (M_2) is given by $PSFB_{12} = 1.418572 \times 10^{12}$, indicating strong support to the SN-CCM.

7 Conclusions

The paper considers an extension of the symmetric structural calibration comparative model by considering that the true covariate follows a skew-normal distribution. Inference is approached via maximum likelihood and Bayesian methodology. Both approaches can be computationally implemented by using simple and accessible software as R and WinBUGS. The model in Barnett (1969) is obtained as a special case. The result of this paper can be extended in several directions. One possibility is to study hypothesis of the form $H_1 : \beta_2 = \ldots = \beta_p, \alpha_2 = \ldots = \alpha_p = 0$, $H_2 : \beta_2 = \ldots = \beta_p$ or $H_3 : \alpha_2 = \ldots = \alpha_p = 0$ by using likelihood ratio, score and Wald statistics test in order to compare the behavior of these statistics test under the N-CCM and SN-CCM. The illustration show that the approach improves Bayesian and maximum likelihood inferences under the incorrect structural N-CCM.

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

In this appendix the observed information matrix is obtained for the SN-CCM. Using the notation presented in Section 3, it follows from (15) and some results related to vector derivatives (see Nel, 1980) that $U_i(\boldsymbol{\theta})$, $i = 1, \ldots, n$, have elements given by

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

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

$$
\begin{aligned}
\frac{\partial d_i}{\partial \mu_x} &= -2\boldsymbol{b}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i, \\
\frac{\partial d_i}{\partial \boldsymbol{\alpha}} &= -2\mathbb{I}_{(p)}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i, \\
\frac{\partial d_i}{\partial \boldsymbol{\beta}} &= -2q_i D^{-1}(\boldsymbol{\psi})\boldsymbol{W}_{2i} + 2c^{-1}\phi_c a_i q_i D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}, \\
\frac{\partial d_i}{\partial \phi_x} &= -c^{-2}a_i^2, \\
\frac{\partial d_i}{\partial \boldsymbol{\phi}} &= -D^{-2}(\boldsymbol{\phi}) D(\boldsymbol{X}_i)\boldsymbol{X}_i + 2c^{-1}\phi_c a_i D^{-2}(\boldsymbol{\phi}) D(\boldsymbol{b})\boldsymbol{X}_i - c^{-2}\phi_c^2 a_i^2 D^{-2}(\boldsymbol{\phi}) D(\boldsymbol{b})\boldsymbol{b}, \\
\frac{\partial d_i}{\partial \lambda_x} &= -2\lambda_x c^{-2}a_i^2,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial a_i}{\partial \mu_x} &= -\boldsymbol{b}^\top D^{-1}(\boldsymbol{\phi})\boldsymbol{b}, \\
\frac{\partial a_i}{\partial \boldsymbol{\alpha}} &= -D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}, \\
\frac{\partial a_i}{\partial \boldsymbol{\beta}} &= D^{-1}(\boldsymbol{\psi})\boldsymbol{W}_{2i} - \mu_x D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}, \\
\frac{\partial a_i}{\partial \phi_x} &= 0, \\
\frac{\partial a_i}{\partial \boldsymbol{\phi}} &= -D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi})\boldsymbol{X}_i, \\
\frac{\partial a_i}{\partial \lambda_x} &= 0,
\end{aligned}
$$

where $\boldsymbol{X}_i = \boldsymbol{Y}_i - \boldsymbol{a} - \boldsymbol{b}\mu_x$, $\boldsymbol{\psi} = (\phi_2, ..., \phi_p)^\top$, $\boldsymbol{W}_{2i} = \boldsymbol{y}_{2i} - \boldsymbol{\alpha} - \boldsymbol{\beta}\mu_x$, $\mathbb{I}_{(p)} = [\mathbf{0}, \mathbf{I}_{p-1}]$ of dimension $p-1 \times p$ and $\boldsymbol{y}_{i2} = (y_{i2}, ..., y_{ip})^\top$.

From (17) it follows that $\dfrac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top}$ have elements with components given by

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

$$
\begin{aligned}
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= 2c^{-1}\phi_c[D^{-1}(\boldsymbol{\psi}) - 2c^{-1}\phi_c \boldsymbol{M}_1] \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\beta} \partial \phi_x} &= 2c^{-2}D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}, \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\beta} \partial \boldsymbol{\phi}^\top} &= -2c^{-1}\phi_c[D_1(\boldsymbol{\beta}) - c^{-1}\phi_c D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}\boldsymbol{b}^\top D(\boldsymbol{b})]D^{-2}(\boldsymbol{\phi}), \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\beta} \partial \lambda_x} &= 4\lambda_x c^{-2}D^{-1}(\boldsymbol{\phi})\boldsymbol{\beta} \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \phi_x \partial \phi_x} &= -\frac{1}{c^2\phi_c^2}(c-1)^2, \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \phi_x \partial \boldsymbol{\phi}^\top} &= -c^{-2}\boldsymbol{b}^\top D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi}), \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \phi_x \partial \lambda_x} &= -\frac{2\lambda_x}{c^2\phi_c^2}(c-1)^2, \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^\top} &= -D^{-2}(\boldsymbol{\phi}) - c^{-2}\phi_c^2 D(\boldsymbol{b})D^{-1}(\boldsymbol{\phi})\boldsymbol{M}D^{-1}(\boldsymbol{\phi})D(\boldsymbol{b}) + 2c^{-1}\phi_c D^2(\boldsymbol{b})D^{-3}(\boldsymbol{\phi}), \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\phi} \partial \lambda_x} &= -\frac{2}{c^2}\lambda_x D(\mathbf{b})D^{-2}(\boldsymbol{\phi})\mathbf{b}, \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \lambda_x \partial \lambda_x} &= \frac{2}{c^2}\mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}(1 + (\phi_x - \lambda_x^2)\mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 A_x}{\partial \boldsymbol{\tau}\partial\boldsymbol{\gamma}^\top} &= 0, \quad \boldsymbol{\tau} = \mu_x, \boldsymbol{\alpha}; \quad \boldsymbol{\gamma} = \mu_x, \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi_x, \boldsymbol{\phi}, \lambda_x \\
\frac{\partial^2 A_x}{\partial \boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= -(\frac{4}{\lambda_x^2}\phi_c\phi_x A_x^3 - \frac{3}{\lambda_x^4}(2c\phi_x + \lambda_x^2)^2 A_x^5)\mathbf{M}_1 - \frac{1}{\lambda_x^2}(2c\phi_x + \lambda_x^2)A_x^3 D^{-1}(\boldsymbol{\psi}), \\
\frac{\partial^2 A_x}{\partial \boldsymbol{\beta}\partial\phi_x} &= -[\frac{2}{\lambda_x^2}(c + \phi_x \mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}) + \frac{3A_x^2}{2\lambda_x^4\phi_c^2}(2c\phi_x + \lambda_x^2)(2c\phi_x + \lambda_x^2 - c^2(\phi_x + \phi_c))] \times \\
&\quad A_x^3 D^{-1}(\boldsymbol{\phi})\boldsymbol{\beta}, \\
\frac{\partial^2 A_x}{\partial \boldsymbol{\beta}\partial\boldsymbol{\phi}^\top} &= \frac{1}{2\lambda_x^4}(\frac{2}{\lambda_x^2}\phi_c\phi_x A_x^3 - 3(2c\phi_x + \lambda_x^2)^2 A_x^5)D^{-1}(\boldsymbol{\phi})\boldsymbol{\beta}\mathbf{b}^\top D(\mathbf{b})D^{-2}(\boldsymbol{\phi}) \\
&\quad + \frac{1}{\lambda_x^2}(2c\phi_x + \lambda_x^2)A_x^3 \mathbb{I}_{(p)} D(\mathbf{b})D^{-2}(\boldsymbol{\phi}), \\
\frac{\partial^2 A_x}{\partial \boldsymbol{\beta}\partial\lambda_x} &= -\frac{4}{\lambda_x^3}[\phi_x(\lambda_x^2 \mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b} - c)]A_x^3 D^{-1}(\boldsymbol{\phi})\boldsymbol{\beta} \\
&\quad - \frac{3A_x^5}{\lambda_x^5\Lambda_x^2}(2c\phi_x + \lambda_x^2)(\phi_x\phi_c + \lambda_x^2(2c^{-1}\phi_x + c^{-2}\lambda_x^2 - \phi_x))D^{-1}(\boldsymbol{\phi})\boldsymbol{\beta}, \\
\frac{\partial^2 A_x}{\partial \phi_x\partial\phi_x} &= [\frac{1}{\lambda_x^2\phi_c^2}(\mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}(\phi_x - c(\phi_c + \phi_x)) + c(1-c)) \\
&\quad - \frac{1}{\lambda_x^2\phi_c^3}(2c\phi_x + \lambda_x^2 - c^2(\phi_c + \phi_x))]A_x^3 + \frac{3}{4\lambda_x^4\phi_c^4}[2c\phi_x + \lambda_x^2 - c^2(\phi_c + \phi_x)]^2 A_x^5,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 A_x}{\partial \phi_x\partial\boldsymbol{\phi}^\top} &= [\frac{1}{2\lambda_x^2\phi_c}(-2\phi_x + 2c(\phi_c + \phi_x)) + \frac{3}{4\lambda_x^4\phi_c^2}(2c\phi_x + \lambda_x^2)(2c\phi_x + \lambda_x^2 - c^2(\phi_c + \phi_x))A_x^2] \times \\
&\quad A_x^3 \mathbf{b}^\top D(\mathbf{b})D^{-2}(\boldsymbol{\phi}), \\
\frac{\partial^2 A_x}{\partial \phi_x\partial\lambda_x} &= \frac{1}{\lambda_x^3\phi_c^3}[2\lambda_x^2\phi_c(\phi_x - c(\phi_c + \phi_x))\mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b} - (\lambda_x^2(c^2\phi_c + 4c\phi_x + 2\lambda_x^2) + 2c\phi_x\phi_c - \\
&\quad (\phi_c + \phi_x)(c^2\phi_c + 2c^2\lambda_x^2))]A_x^3 + \frac{3}{2\lambda_x^5\phi_c^2\Lambda_x^2}[2c\phi_x + \lambda_x^2 - c^2(\phi_c\phi_x)] \times \\
&\quad [\phi_x\phi_c + \lambda_x^2(2c^{-1}\phi_x + c^{-2}\lambda_x^2 - \phi_x)]A_x^5, \\
\frac{\partial^2 A_x}{\partial \boldsymbol{\phi}\partial\boldsymbol{\phi}^\top} &= [-\frac{\phi_x\phi_c}{\lambda_x^2}A_x^3 + \frac{3(2c\phi_x + \lambda_x^2)^2}{2\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\phi_x + \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} &= [2\frac{\phi_x}{\lambda_x^3}(\lambda_x^2 \mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b} - c)A_x^3]D(\mathbf{b})D^{-2}(\boldsymbol{\phi})\mathbf{b} \\
&\quad + 3[\phi_x\phi_c + \frac{\lambda_x^2}{\lambda_x^3\Lambda_x^2}(2c^{-1}\phi_x + c^{-2}\lambda_x^2 - \phi_x)A_x^5]D(\mathbf{b})D^{-2}(\boldsymbol{\phi})\mathbf{b}, \\
\frac{\partial^2 A_x}{\partial \lambda_x\partial\lambda_x} &= \frac{1}{\lambda_x^6\Lambda_x^4}[4c^{-1}\lambda_x^4\Lambda_x^2(\phi_x - c^{-1}\lambda_x^2\phi_x \mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b} - c^{-2}\lambda_x^4 \mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b} + c^{-1}\lambda_x^2) \\
&\quad - (\phi_x\phi_c + \lambda_x^2(2c^{-1}\phi_x + c^{-2}\lambda_x^2 - \phi_x))(3\lambda_x^2\Lambda_x^2 + 4c^{-2}\lambda_x^4\Lambda_x)]A_x^3 + \\
&\quad \frac{3}{\lambda_x^6\Lambda_x^4}[\phi_x\phi_c + \lambda_x^2(2c^{-1}\phi_x + c^{-2}\lambda_x^2 - \phi_x)]^2 A_x^5,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 d_i}{\partial \mu_x \partial \mu_x} &= 2\boldsymbol{b}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{b}, \\
\frac{\partial^2 d_i}{\partial \mu_x \partial \boldsymbol{\alpha}^\top} &= 2\boldsymbol{b}^\top \boldsymbol{\Sigma}^{-1}\mathbb{I}_{(p)}^\top, \\
\frac{\partial^2 d_i}{\partial \mu_x \partial \boldsymbol{\beta}^\top} &= -2c^{-1}\boldsymbol{A}_i, \\
\frac{\partial^2 d_i}{\partial \mu_x \partial \phi_x} &= 2\frac{(c-1)}{c^2\phi_c}a_i, \\
\frac{\partial^2 d_i}{\partial \mu_x \partial \boldsymbol{\phi}^\top} &= 2c^{-1}\boldsymbol{X}_i^\top \boldsymbol{\Sigma}^{-1} D^{-1}(\boldsymbol{\phi}) D(\boldsymbol{b}), \\
\frac{\partial^2 d_i}{\partial \mu_x \partial \lambda_x} &= 4\lambda_x \frac{(c-1)}{c^2\phi_c}a_i, \\
\frac{\partial^2 d_i}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top} &= 2\mathbb{I}_{(p)}\boldsymbol{\Sigma}^{-1}\mathbb{I}_{(p)}^\top, \\
\frac{\partial^2 d_i}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top} &= 2q_i[\boldsymbol{D}^{-1}(\boldsymbol{\psi}) - 2c^{-1}\phi_c \boldsymbol{M}_1] + 2c^{-1}\phi_x \boldsymbol{D}^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}(\boldsymbol{Y}_{i2} - \boldsymbol{\alpha})^\top \boldsymbol{D}^{-1}(\boldsymbol{\psi}),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 d_i}{\partial \boldsymbol{\alpha} \partial \phi_x} &= 2c^{-2}a_i D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}, \\
\frac{\partial^2 d_i}{\partial \boldsymbol{\alpha} \partial \lambda_x} &= 2\mathbb{I}_{(p)}\Sigma^{-1}D^{-1}(\boldsymbol{\phi})[D(\boldsymbol{X}_i) - c^{-1}\phi_c\, a_i D(\boldsymbol{b})], \\
\frac{\partial^2 d_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= 4\frac{\phi_c^2}{c^2}a_i[D^{-1}(\boldsymbol{\psi})(\boldsymbol{Y}_{i2} - \boldsymbol{\alpha} - 2\boldsymbol{\beta}\mu_x)\boldsymbol{\beta}^\top D^{-1}(\boldsymbol{\psi}) + D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}(\boldsymbol{Y}_{i2} - \boldsymbol{\alpha} - 2\boldsymbol{\beta}\mu_x)^\top D^{-1}(\boldsymbol{\psi})] \\
&\quad -2c^{-1}\phi_c \boldsymbol{D}^{-1}(\boldsymbol{\psi})(\boldsymbol{Y}_{i2} - \boldsymbol{\alpha} - 2\boldsymbol{\beta}\mu_x)(\boldsymbol{Y}_{i2} - \boldsymbol{\alpha} - 2\boldsymbol{\beta}\mu_x)^\top \boldsymbol{D}^{-1}(\boldsymbol{\psi}) \\
&\quad +2\mu_x(q_i + c^{-1}\phi_c a_i)D^{-1}(\boldsymbol{\psi}) + 2\frac{\phi_c^2}{c^2}a_i^2[D^{-1}(\boldsymbol{\psi}) - 4\frac{\phi_c}{c}\boldsymbol{M}_1], \\
\frac{\partial^2 d_i}{\partial \boldsymbol{\beta} \partial \phi_x} &= -2c^{-2}a_i \boldsymbol{A}_i^\top, \\
\frac{\partial^2 d_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\phi}^\top} &= 2[q_i \mathbb{I}_{(p)} D(\boldsymbol{Y}_i - \boldsymbol{a} - q_i\boldsymbol{b}) + c^{-1}\phi_c \boldsymbol{A}_i^\top (\boldsymbol{Y}_i - \boldsymbol{a} - \boldsymbol{b}q_i)^\top D(\boldsymbol{b})]D^{-2}(\boldsymbol{\phi}), \\
\frac{\partial^2 d_i}{\partial \boldsymbol{\beta} \partial \lambda_x} &= -4\lambda_x c^{-2}a_i \mathbf{A}_i^\top, \\
\frac{\partial^2 d_i}{\partial \phi_x \partial \phi_x} &= 2\frac{c^{-3}}{\phi_c}(c-1)a_i^2, \\
\frac{\partial^2 d_i}{\partial \phi_x \partial \boldsymbol{\phi}^\top} &= (-2c^{-3}\phi_c a_i^2 D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi})\boldsymbol{b} + 2c^{-2}a_i D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi})\boldsymbol{X}_i)^\top, \\
\frac{\partial^2 d_i}{\partial \phi_x \partial \lambda_x} &= 4\lambda_x \frac{c^{-3}}{\phi_c}(c-1)a_i^2,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 d_i}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top} &= 2D^{-3}(\boldsymbol{\phi})D^2(\boldsymbol{X}_i) - 4c^{-1}\phi_c a_i D^{-3}(\boldsymbol{\phi})D(\boldsymbol{b})D(\boldsymbol{X}_i) \\
&\quad -2c^{-1}\phi_c D^{-2}(\boldsymbol{\phi})D(\boldsymbol{b})\boldsymbol{X}_i\boldsymbol{X}_i^\top D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi}) \\
&\quad +2c^{-2}\phi_c^2 D^{-2}(\boldsymbol{\phi})D(\boldsymbol{b})\boldsymbol{X}_i\boldsymbol{X}_i^\top \boldsymbol{M}D^{-1}(\boldsymbol{\phi})D(\boldsymbol{b}) \\
&\quad +2c^{-2}\phi_c^2 a_i^2 D^{-3}(\boldsymbol{\phi})D^2(\boldsymbol{b}) - 2c^{-3}\phi_c^3 a_i^2 D^{-1}(\boldsymbol{\phi})D(\boldsymbol{b})\boldsymbol{M}D(\boldsymbol{b})D^{-1}(\boldsymbol{\phi}) \\
&\quad +2c^{-2}\phi_c^2 D^{-1}(\boldsymbol{\phi})D(\boldsymbol{b})\boldsymbol{M}\boldsymbol{X}_i\boldsymbol{X}_i^\top D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi}), \\
\frac{\partial^2 d_i}{\partial\boldsymbol{\phi}\partial\lambda_x} &= 2\lambda_x[-2c^{-3}(\phi_x+\lambda_x^2)a_i^2 D(\mathbf{b})D^{-2}(\boldsymbol{\phi})\mathbf{b} + 2c^{-2}a_i D(\mathbf{b})D^{-2}(\boldsymbol{\phi})\mathbf{X}_i],
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{\partial^2 a_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= -2\mu_x D^{-1}(\boldsymbol{\psi}), \\
\frac{\partial^2 a_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\phi}^\top} &= -\mathbb{I}_{(p)}D(\boldsymbol{Y}_i - \boldsymbol{a} - 2\mu_x\boldsymbol{b})D^{-2}(\boldsymbol{\phi}), \\
\frac{\partial^2 a_i}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top} &= 2D(\boldsymbol{X}_i)D(\boldsymbol{b})D^{-3}(\boldsymbol{\phi})
\end{aligned}
$$

where $M = D^{-1}(\phi)\boldsymbol{b}\boldsymbol{b}^\top D^{-1}(\boldsymbol{\phi})$ $\boldsymbol{A}_i(\boldsymbol{Y}_{i2} - \boldsymbol{\alpha} - 2q_i\boldsymbol{\beta})^\top D^{-1}(\boldsymbol{\psi})$, $\boldsymbol{M}_1 = D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}\boldsymbol{\beta}^\top D^{-1}(\boldsymbol{\psi})$, $D_1(\boldsymbol{\beta}) = [0\ D(\boldsymbol{\beta})]$ of dimension $p-1\times p$, $q_i = \mu_x + c^{-1}a_i\phi_c$ and $i = 1, ..., n$.

References

Aitkin, M (1991). Posterior factor. *Journal of the Royal Statistical Society,* B, **53**, 111-142.

Arellano-Valle, R.B. and Genton, M. G. (2005). Fundamental skew distributions. *Journal of Multivariate Analysis,* **96**, 93-116.

Arellano-Valle R.B., Bolfarine, H. and Lachos, V.H. (2005a). Skew-normal linear mixed models. *Journal of Data Science,* **3(4)**, 415-438.

Arellano-Valle, R.B., Ozan S., Bolfarine, H. and Lachos, V.H. (2005b). Skew normal measurement error models. *Journal of Multivariate Analysis,* **96**, 265-281.

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

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.

Berger, J. O. and Pericchi, L. O. (1992). The intrinsic Bayes factor. Technical Report. Department of Statistics, Pardue University, West Lafayett.

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 Statistical Computation and Simulation*, **72**, 2, 125-140.

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

Fuller, W. A. (1987). *Measurement Error Models.* Wiley, New York.

Gamermann, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* . London, Chapman & Hill.

Geisser, S. and Eddy, W. (1979). A predictive approach to model selection.*J. Amer. Statist. Association.,* **79**, 153-160.

Gelfand, A. E. and Dey, D.K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society,* **B**, **56** 501-514.

Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457-511.

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

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.

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.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society,* B , **64**, 583-639.

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

	N-CCM		SN-CCM	
Parameter	Estimate	SE	Estimate	SE
α_2	-2.0447	1.0532	-2.0632	1.0463
α_3	-5.2857	1.2278	-5.2588	1.2340
α_4	-4.3726	1.2406	-4.3576	1.2438
β_2	1.0597	0.0446	1.0605	0.0443
β_3	1.1919	0.0521	1.1907	0.0523
β_4	1.1306	0.0526	1.1300	0.0527
ϕ_1	5.0247	1.0006	5.0392	0.9960
ϕ_2	1.9151	0.5846	1.7954	0.5800
ϕ_3	2.9236	0.7893	3.0431	0.8102
ϕ_4	3.8845	0.8744	3.9345	0.8961
μ_x	22.4611	0.9008	13.0527	1.2596
σ_x^2	53.3983	9.7174	5.1023	4.9130
λ_x	-	-	11.6979	1.5647
- log-likelihood	788.0931		783.5044	
AIC	2.7781		2.7656	
BIC	2.8544		2.8483	
HQ	2.7726		2.7596	

Table 2: Posterior summaries results of fitting normal and skew-normal calibration comparative models (N-CCM and SN-CCM) to the Barnett data set.

	N-CCM				SN-CCM			
Parameter	Mean	Sd	2.5%	97.5%	Mean	Sd	2.5%	97.5%
α_2	-2.137	1.068	-4.322	-0.112	-2.288	1.033	-4.302	-0.338
α_3	-5.373	1.254	-7.870	-2.960	-5.497	1.207	-7.935	-3.207
α_4	-4.455	1.274	-7.085	-2.073	-4.55	1.246	-7.002	-2.102
β_2	1.064	0.045	0.976	1.155	1.071	0.044	0.988	1.157
β_3	1.196	0.053	1.095	1.155	1.202	0.052	1.103	1.308
β_4	1.135	0.055	1.034	1.244	1.139	0.053	1.036	1.245
ϕ_1	5.317	1.103	3.542	7.872	5.360	1.103	3.601	7.911
ϕ_2	1.988	0.655	0.877	3.405	1.827	0.634	0.708	3.249
ϕ_3	3.213	0.916	1.682	5.270	3.345	0.934	1.793	5.493
ϕ_4	4.230	1.002	2.634	6.497	4.297	1.022	2.674	6.730
μ_x	22.450	0.911	20.690	24.240	12.370	1.274	10.450	15.190
σ_x^2	55.380	10.400	38.190	78.480	3.843	6.305	0.009	19.490
λ_x	-	-	-	-	12.390	1.742	9.136	16.030
DIC	2045.000				1967.000			
$c(l)$	-1022.479				-994.498			