On diagnostics in a skew version of scale mixtures of normal distribution

Ferreira, Clécio da Silva
Universidade Federal de Juiz de Fora, Departamento de Estatística
Instituto de Ciências Exatas, Campus Universitário
Juiz de Fora, CEP 36036-900,Brasil
E-mail: clecio.ferreira@ufjf.edu.br

Bolfarine, Heleno
Universidade de São Paulo, Departamento de Estatística
Rua do Matão, 1010
São Paulo, CEP 05508-090, Brasil
E-mail: hbolfar@ime.usp.br

Lachos, Victor Hugo
Universidade Estadual de Campinas, Departamento de Estatística
Rua Sérgio Buarque de Holanda, 651
Campinas, CEP 13083-859, Brasil
E-mail: hlachos@ime.unicamp.br

1. Introduction

Scale mixtures of normal distribution are often used as a challenging family for statistical procedures of symmetrical data, providing a group of thick-tailed distributions that are often used for robust inference of symmetrical data. Moreover, this class include distributions such as the Student-t, the slash, the contaminated normal, among others. However, the theory and application (through simulation or experimentation) often generate a great amount of data sets that are skewed and present heavy-tail as, for instance, data set on family income. Thus, appropriate distributions to fit and simulate these skewed and heavy tailed data are needed.

Ferreira et al. (2011) propose a new family of distributions that combine skewness with heavy tails, using the scale mixtures of normal distributions and the normal kernel, named SSMN. Moreover, this distribution is attractive because it simultaneously models the skewness with heavy tails and it has a stochastic representation that allows easy implementation of the EM-algorithm. This extension result in a flexible class of models for robust univariate models once it contains as especial cases, the skew-normal (Azzalini and Dalla-Valle, 1996) distribution and all the symmetric class of SMN distributions defined by Andrews and Mallows (1974). All these distributions have heavier tails than the skew-normal one, and thus can be used for robust inference in many type of models. The main virtue of the members of this family of distributions is that they are easy to simulate from and they also supply genuine EM algorithms for maximum likelihood estimation. For univariate skewed responses, the EM-type algorithm has been discussed with emphasis on the skew-t, skew-slash, skew-contaminated normal and skew-exponential power distributions.

In this paper, we extend the EM algorithm for linear regression models (SSMN-RM) and we developed some methods to obtain diagnostic measures on SSMN models. However, as the observed data likelihood of the SSMN-RM involves an intractable integral, it is very difficult directly to apply Cook's (1986) approach to obtain local influence measures. In this paper we apply Zhu and Lee's (2001) local influence approach to the SSMN-RM leading to closed form expression of local influence measures.

2. Skew-normal distributions

A simpler departure which defines the univariate skew-normal distribution has been proposed by Azzalini (1985), defining the following probability density function (pdf)

$$(1) \quad f(y) = 2\phi(y|\mu, \sigma^2)\Phi_1\left(\frac{\lambda(y-\mu)}{\sigma}\right), \quad y \in \mathbb{R}.$$

where $\phi(.|\mu, \sigma^2)$ stands for the pdf of the normal distribution with mean μ and variance σ^2, $\Phi_1(.)$ represents the cumulative distribution function (cdf) of the standard normal distribution. When $\lambda = 0$, the skew normal distribution reduces to the normal distribution ($y \sim \mathrm{N}(\mu, \sigma^2)$). A random variable y with pdf as in (1), will be denoted by $\mathrm{SN}(\mu, \sigma^2, \lambda)$. Its marginal stochastic representation, which can be used to derive several of its properties, is given by

$$(2) \quad y \stackrel{d}{=} \mu + \sigma(\delta|T_0| + (1-\delta^2)^{1/2}T_1), \quad \text{with} \quad \delta = \frac{\lambda}{\sqrt{1+\lambda^2}},$$

where $|T_0|$ denotes the absolute value of T_0, $T_0 \sim \mathrm{N}_1(0, 1)$ and $T_1 \sim \mathrm{N}(0, 1)$ are independent, and "$\stackrel{d}{=}$" means "distributed as". From (2) it follows that the expectation and variance of y are given, respectively, by

$$(3) \quad E[Y] = \mu + \sqrt{\frac{2}{\pi}}\sigma\delta, \quad Var[Y] = \sigma^2(1 - \frac{2}{\pi}\delta^2).$$

3. A skew version of scale mixtures of normal distribution

Lange and Sinsheimer (1993) provide a group of thick-tailed symmetric distributions which has the normal distribution as particular case.

Definition 1. *A random variable Y follows a scale mixtures of normal distribution with location parameter $\mu \in \mathbb{R}$ and a positive scale parameter σ^2 if its density function assumes the form*

$$(4) \quad f(y) = \int_0^\infty \phi(y|\mu, \kappa(u)\sigma^2)dH(u; \boldsymbol{\tau}),$$

where $H(.; \boldsymbol{\tau})$ is the cdf of a positive random variable U indexed by the parameter vector $\boldsymbol{\tau}$ and $\kappa(.)$ is a strictly positive function. For a random variable with a pdf as in (4), we shall use the notation $Y \sim SMN(\mu, \sigma^2, H; \kappa)$. Moreover, when $\mu = 0$ and $\sigma^2 = 1$, we denote $Y \sim SMN(H; \kappa)$.

A asymmetric version of SMN distributions has been introduced by Ferreira et al. (2011) as a challenging family for statistical procedures of asymmetric data. This new family of distributions contains all the distributions studied by Lange and Sinsheimer (1993), the so called normal/independent (NI) distribution, but with an extra parameter that regulates the skewness.

Definition 2. *A random variable Y follows an skew scale mixtures of normal distribution (SSMN) with location parameter $\mu \in \mathbb{R}$, scale factor σ^2 and skewness parameter $\lambda \in \mathbb{R}$, if its pdf is given by*

$$(5) \quad f(y) = 2\int_0^{+\infty} \phi(y|\mu, \sigma^2\kappa(u))\Phi_1(\lambda(y-\mu)/\sigma)dH(u, \boldsymbol{\tau}),$$

where U is a positive random variable with cdf $H(u; \boldsymbol{\tau})$. For a random vector with pdf as in (5), we use the notion $y \sim SSMN(\mu, \sigma^2, \lambda, H; \kappa)$. If $\mu = 0$ and $\sigma^2 = 1$ we refer to it as a standard SSMN distribution and we denote it by $SSMN(\lambda, H; \kappa)$. If $\lambda = 0$, we have the SMN distribution.

Examples of SSMN distributions

• *The Skew Student-t normal distribution, with* $\nu > 0$ *degrees of freedom,* $Y \sim StN(\mu, \sigma^2, \lambda; \nu)$.
Considering $U \sim Gamma(\nu/2, \nu/2)$, $\kappa(u) = 1/u$, Y has the density function:

$$(6) \quad f(y) = 2\frac{1}{\sigma\sqrt{\nu\pi}}\frac{\Gamma((\nu+1)/2)}{\Gamma(\frac{\nu}{2})}\left(1+\frac{d}{\nu}\right)^{-(\frac{\nu+1}{2})}\Phi\left(\lambda\frac{(y-\mu)}{\sigma}\right).$$

The skew Student-t normal distribution has been developed by Gómez et al. (2007). A particular case of the StN distribution is the skew-Cauchy distribution, when $\nu = 1$. Also, when $\nu \uparrow \infty$, we get the skew-normal distribution as the limiting case.

• *The skew-slash distribution, with the shape parameter* $\nu > 0$, $SSL(\mu, \sigma^2, \lambda; \nu)$.
With $h(u; \nu) = \nu u^{\nu-1}\mathbb{I}_{(0,1)}(u)$ and $\kappa(u) = 1/u$, we have

$$(7) \quad f(y) = 2\nu\int_0^1 u^{\nu-1}\phi\left(y|\mu, \frac{\sigma^2}{u}\right) du\Phi_1\left(\lambda\frac{(y-\mu)}{\sigma}\right), \quad y \in \mathbb{R}.$$

The skew-slash distribution reduces to the skew-normal distribution when $\nu \uparrow \infty$. See Wang and Genton (2006) for further details.

• *The skew-contaminated normal distribution,* $SCN(\mu, \sigma^2, \lambda; \nu, \gamma)$, $0 \leq \nu \leq 1$, $0 < \gamma \leq 1$. Taking $h(u; \nu) = \nu\mathbb{I}_{(u=\gamma)} + (1-\nu)\mathbb{I}_{(u=1)}$, $\boldsymbol{\tau} = (\nu, \gamma)^\top$ and $\kappa(u) = 1/u$, it follows straightforwardly that

$$(8) \quad f(y) = 2\left\{\nu\phi\left(y|\mu, \frac{\sigma^2}{\gamma}\right)\Phi_1\left(\lambda\frac{(y-\mu)}{\sigma}\right) + (1-\nu)\phi(y|\mu, \sigma^2)\Phi_1\left(\lambda\frac{(y-\mu)}{\sigma}\right)\right\}.$$

The parameter ν represents the percentage of outliers, while γ may be interpreted as a scale factor. The skew-contaminated normal distribution reduces to the skew-normal distribution when $\nu = 0$.

• *The skew-exponential power distribution,* $Y \sim SEP(\mu, \sigma^2, \lambda; \nu)$, *with a shape parameter* $0 < \nu \leq 1$.
Its pdf is given by

$$(9) \quad f(y) = 2\frac{\nu}{2^{1/2\nu}\sigma\Gamma(1/2\nu)}e^{-d^\nu/2}\Phi_1\left(\lambda\frac{(y-\mu)}{\sigma}\right), \; d = (y-\mu)^2/\sigma^2.$$

SSMN Regression Models

Suppose that we have observations on m independent individuals, $Y_1, \ldots, Y_n$, where $Y_i \sim SSMN(\mu_i, \sigma^2, \lambda)$, $i = 1, \ldots, n$. Associated with individual i we assume a known $p \times 1$ covariate vector $\mathbf{x}_i$, which we use to specify the linear predictor $\mu_i = \mathbf{x}_i^\top\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a p-dimensional vector of unknown regression coefficients.

$$(10) \quad \begin{aligned} y_i &= \beta_0 + \textstyle\sum_{k=1}^p x_{ik}\beta_k + \varepsilon_i, \quad i = 1, \cdots, n, \\ &= \mathbf{x}_i^\top\boldsymbol{\beta} + \varepsilon_i, \\ \varepsilon_i &\sim SSMN(0, \sigma^2, \lambda, H; \kappa). \end{aligned}$$

4. The local influence approach

Consider a perturbation vector $\mathbf{w} = (w_1, ..., w_n)^\top$ varying in an open region $\mathbf{\Omega} \in \mathbb{R}^n$. Let $\ell_c(\boldsymbol{\theta}, \mathbf{w}|\mathbf{y}_c)$, $\boldsymbol{\theta} \in \mathbb{R}^q$ be the complete-data log-likelihood of the perturbed model. We assume that there is a $\mathbf{w}_0$ such that $\ell_c(\boldsymbol{\theta}, \mathbf{w}_0|\mathbf{Y}_c) = \ell_c(\boldsymbol{\theta}|\mathbf{Y}_c)$ for all $\boldsymbol{\theta}$. Let $\widehat{\boldsymbol{\theta}}(\mathbf{w})$ denote the maximum of the function $Q(\boldsymbol{\theta}, \mathbf{w}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_c(\boldsymbol{\theta}, \mathbf{w}|\mathbf{Y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}]$. The influence graph is defined as $\boldsymbol{\alpha}(\mathbf{w}) = (\mathbf{w}^\top, f_Q(\mathbf{w}))^\top$ where $f_Q(\mathbf{w})$ is the Q-displacement function defined as follows:

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

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

$$C_{f_Q,\mathbf{d}} = -2\mathbf{d}^\top \ddot{Q}_{\mathbf{w}_o}\mathbf{d}, \quad -\ddot{Q}_{\mathbf{w}_0} = \Delta_{\mathbf{w}_o}^\top \left\{-\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}})\right\}^{-1} \Delta_{\mathbf{w}_0}$$

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

As in Cook (1986), the expression $-\ddot{Q}_{\mathbf{w}_0}$ is the fundamental equation for detecting influential observations. A clear picture of $-\ddot{Q}_{\mathbf{w}_0}$ (a symmetric matrix), is given by its spectral decomposition

$$-2\ddot{Q}_{\mathbf{w}_o} = \sum_{k=1}^{n} \xi_k \mathbf{e}_k \mathbf{e}_k',$$

where $(\xi_1, \mathbf{e}_1), \ldots, (\xi_n, \mathbf{e}_n)$ are the eigenvalue-eigenvector pairs of the matrix $-2\ddot{Q}_{\mathbf{w}_o}$ with $\xi_1 \geq \ldots \geq \xi_q, \xi_{q+1} = \ldots = \xi_n = 0$ and $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are elements of the associated orthonormal basis. Zhu and Lee (2001) proposed to inspect all eigenvectors corresponding to nonzero eigenvalues for more revealing information but it can be computationally intensive for large n.

The perturbation schemes included in this work are: Case weights, responde variable perturbations, explanatory variable perturbation and perturbation of the scale parameter σ^2.

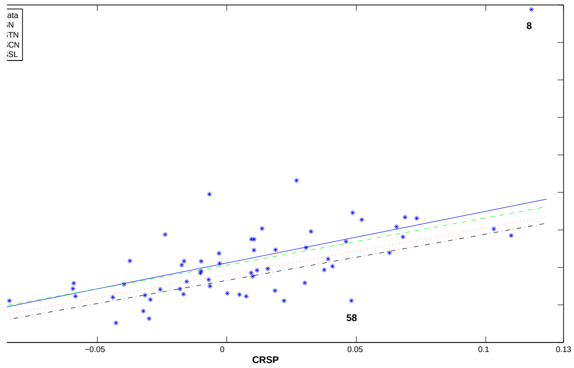
5. Application

This section we consider a analysis of a data set taken from Table 1 of Butler et al. (1990). This data set was also studied by Azzalini and Capitanio (2003), under asymmetrical model. On the basis of arguments presented by them, a linear regression is introduced:

$$y = \beta_0 + \beta_1 CRSP + \epsilon$$

where y is the excess rate of the Martin Marietta company, $CRSP$ is an index of the excess rate of return for the New York market as a whole and ϵ is an error term which in our case is taken to be distributed as $SSMN(0, \sigma^2, \lambda, H; \kappa)$. Data over a period of $n = 60$ consecutive months are available. Table 1 presents the ML estimates of the parameters for the normal symmetric normal and the SSMN models (SN, StN, SEP, SSL and SCN), together with their corresponding standard errors calculated via the observed information matrix. The log-likelihood values (see row $\ell(\widehat{\boldsymbol{\theta}})$) indicate that the SMSN distributions with heavy tails presents the best fit than the SN model, with the SSL, StN and SCN ones significantly better (the LR test is comparing all others distributions with SN model).

Figure 1: SSMN regression models for the Martin Marietta data, with influence points.



Since $E[\varepsilon] \neq 0$, the residual for y_i is given by $e_i = y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}} - \widehat{\varepsilon}_i$, where $\widehat{\epsilon_i} = \widehat{E[\epsilon_i]}$. The estimated regression lines (Figure 1) are obtained as $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 CRSP_i + \widehat{\epsilon}_i$.

As an example, Figure 2 presents the index graphs of $M(0)$ for the SN, StN, SSL and SCN models for case weights perturbation. As can be seen in this figures, the observation 58 is influential in all models, while observation 8 (above the estimated straight) is influential in the MLEs of the model SN (actually, in all forms of perturbation, but not showed here).

Table 1: MLEs of the five models fitted on the Martin Marietta data. The values in parenthesis are estimated asymptotic standard errors, using information matrix.

	SN	SSL	SEP	StN	SCN
α	-0.093(0.013)	-0.042(0.031)	-0.050(0.006)	-0.042(0.024)	-0.040(0.014)
β	1.379(0.241)	1.247(0.193)	1.228(0.191)	1.235(0.190)	1.263(0.177)
σ^2	0.019(0.004)	0.002(0.001)	0.001(0.001)	0.003(0.002)	0.004(0.001)
λ	3.916(1.405)	0.547(0.636)	0.552(0.165)	0.717(0.649)	0.709(0.392)
ν	-	1.101	0.511	2.8	0.089
γ	-	-	-	2.8	0.05
$\ell(\widehat{\boldsymbol{\theta}})$	66.017	74.014	71.247	73.547	74.694
LR	-	15.993	10.461	15.060	17.355
p-value	-	0	0.001	0	0

Acknowledgments
The first author acknowledges the partial financial support from Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) and CNPq - Brazil.

REFERENCES (RÉFERENCES)

Andrews, D. R. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society*, Series B, **36**, 99-102.

Arellano-Valle, R.B., Ozan, S., Bolfarine, H. and Lachos, V.H. (2005). Skew-Normal Measurement Error Models. *Journal of Multivariate Analysis*, **96**, 93-116.

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

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *Journal of the Royal Stat. Society, B*, **65**, 367-389.

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

Butler, R. L., McDonald, J. B., Nelson, R. D. and White, S. B. (1990). Robust and partly adaptive estimation of regression models. *Rev. Econ. Statist*, **72**, 321-327.

Cook, R.D. (1986). Assessment of local influence. *Journal of Royal Statistical Society*, 48, 133-169.

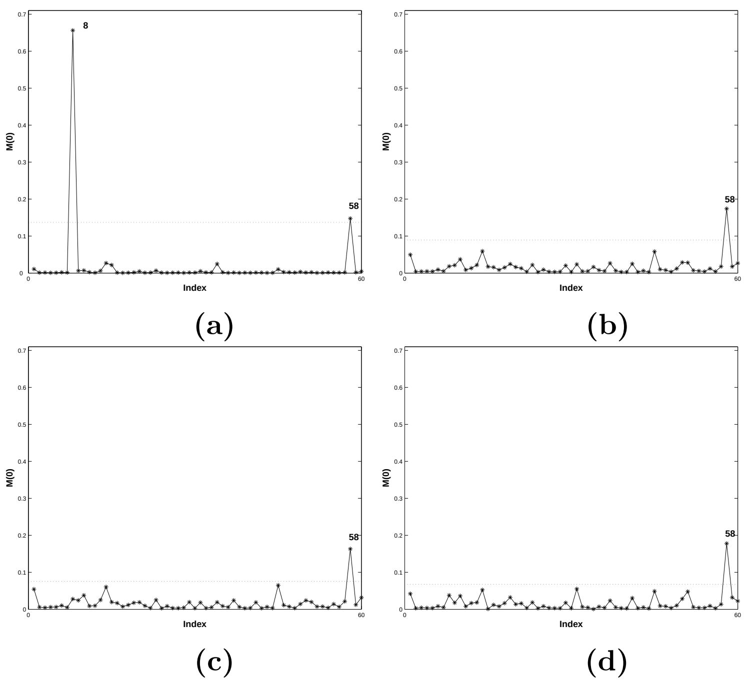
Ferreira, C. S., Bolfarine, H. and Lachos, V. H. (2011). Skew Scale Misture of normal distributions: Properties and Estimation. *Statistical Methodology*, 8, Issue 2, 154-171.

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

Lange, K. and Sinsheimer, J.S. (1993). Normal/independent distributions and their applications in robust regression.*Journal of Computational and Graphical Statistics*, 2, 175-198.

Zhu, H. and Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society*, Ser. **B**, 63, 111-126.

Figure 2: Martin Marietta's data set. Index plots of M(0) for case weights perturbation. (a) Skew normal, (b) skew slash, (c) skew t-normal and (d) skew contaminated normal models. Dotted lines are the bench-mark for M(0) with $c^* = 3$.



RÉSUMÉ (ABSTRACT) — optional

Ferreira et. al (2011) have defined a skewed version of the scale mixtures of normal distributions (SSMN) and derived several of its probabilistic and inferential properties, including estimation via EM algorithm. In this paper, we extend the EM algorithm for linear regression models and we develop diagnostic analysis via local influence for linear regression models under SSMN models, following Zhu and Lee's (2001) approach. Finally, results obtained for a real data set are reported, illustrating the usefulness of the proposed methodology.