Linear Regression Models with Finite Mixtures of Skew Heavy-Tailed Errors

Víctor Hugo Lachos Dávila

Department of Statistics
University of Connecticut, U.S.A.

Second International Conference in **Stochastic Processes and Random Phenomena and Their Applications 2018**: In Tribute to the 65th birthday of Professor Dipak K. Dey

October 03-06, Lima-Peru

Motivation

- A basic assumption of linear regression (LR) models is that the error term follows a normal distribution;
- Many extensions have been proposed:
  - Using symmetric distributions: Galea, Paula and Bolfarine (1997)[*The Statistician*, Local Influence in elliptical linear regression]
  - Using asymmetric distributions: Lachos, Bandyopadhyay and Garay (2011). [*Stat. and Prob. Letters*, Heteroscedastic nonlinear regression models based on scale mixtures of skew normal (SMSN) distributions]
  - Using semiparametric methods: Ibacache, Paula and Cysneiros (2013).[*Test*, Semiparametric additive models under symmetric distributions]
- Drawback: A single parametric family in unable to provide a satisfactory models for local variations in the observed data. McLachlan and Peel (2004)[Finite Mixture Models]

Proposals based in mixtures of distributions

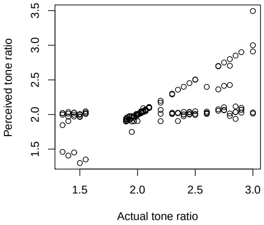
- Bartolucci and Sccacia (2005)[*Computational Statistics and Data Analysis*, The use of mixtures for dealing with non-normal regression errors]. They proposed to model the random term with finite mixtures of normal distributions.

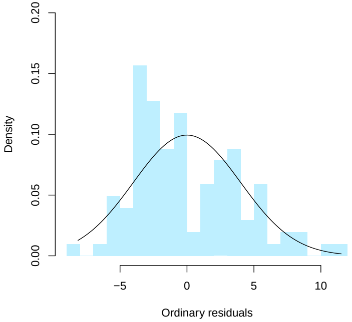
$$Y_i = \mathbf{X}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \sim \sum_{j=1}^{g} p_j \mathrm{N}(\mu_j, \sigma_j^2), \ \sum_{j=1}^{g} p_j \mu_j = 0, \ i = 1, \dots, n,$$

- Recent extensions to the multivariate context
  - Soffritti and Galimberti, G.,(2011)[*Statistics and Computing* , Multivariate linear regression with non-normal errors: a solution based on mixture models.]
  - Galimberti and Soffritti (2014)[*Computational Statistics and Data Analysis*, A multivariate linear regression analysis using finite mixtures of t-distributions]
- Drawback: Not appropriate when the error term present multimodality, heavy-tails and skewness simultaneously.
- **Our goal**: To model the random errors in regression models with finite mixtures of Scale mixture of skew-normal (SMSN) distributions introduced by Branco and Dey (2001).

Important Observation

These approaches are different of the one proposed by Zeller, Cabral and Lachos (2016)[*TEST*, Robust mixture regression modeling based on scale mixtures of skew-normal distributions], where the linear regression is modeling with different regression functions (Mixture of regressions of switching regression)





rses data) Histogram of ordinary

The linear regression model with FM-SMSN errors

Based on Bartolucci and Sccacia (2005)'s proposal, the FM-SMSN-LR model is defined by considering that the random error ϵ_i follows a g-component mixture of SMSN densities given by:

$$Y_i = \mathbf{X}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \sim \sum_{j=1}^{g} p_j SMSN(\mu_j + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu}_j), \quad p_j \geq 0, \ \sum_{j=1}^{g} p_j = 1,$$

where the $\mu_j's$ satisfy the identifiability constraint $\sum_{j=1}^{g} p_j \mu_j = 0$, so that

$$E[\epsilon_i] = 0, \ i = 1, \ldots, n.$$

$\mu_{ij} = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \mu_j = \vartheta_j + \mathbf{x}_i^\top \boldsymbol{\beta}$, where $\vartheta_j = \beta_0 + \mu_j$, $\mathbf{x}_i^\top = (x_{i1}, \ldots, x_{ip})^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$. It follows that the density of the ith observation is

$$Y_i | \boldsymbol{\Theta} \sim \sum_{j=1}^{g} p_j SMSN(y_i | \mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu}_j),$$

where $\boldsymbol{\Theta} = (\boldsymbol{\beta}^\top, (p_1, \ldots, p_{g-1})^\top, \boldsymbol{\nu}^\top, \boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_g^\top)^\top$, with $\boldsymbol{\theta}_j = (\vartheta_j, \sigma_j^2, \lambda_j)^\top$.

Estimation via the EM algorithm

We consider the latent indicator variable Z_{ij} such that:

$$P(Z_{ij}=1)=1-P(Z_{ij}=0)=p_j, \quad \sum_{j=1}^{g} p_j=1,$$

$$y_i|Z_{ij}=1 \sim SMSN(\mu_{ij}+b\Delta_j,\sigma_j^2,\lambda_j;H(\boldsymbol{\nu})).$$

Note that $\mathbf{Z}_1,\ldots,\mathbf{Z}_n$, with $\mathbf{Z}_i=(Z_{i1},\ldots,Z_{ig})^\top$, are independent random vectors, each one having a multinomial distribution with probability function

$$f(\mathbf{z}_i)=p_1^{z_{i1}}p_2^{z_{i2}}...(1-p_1-\ldots-p_{g-1})^{z_{ig}},$$

which we denote by $\mathbf{Z}_i \sim M(1;p_1\ldots,p_g)$.

The FM-SMSN-LR model can be represented hierarchically as

$$\begin{aligned}
y_i | u_i, t_i, Z_{ij} = 1 &\overset{\text{ind}}{\sim} N(\mu_{ij} + \Delta_j t_i, u_i^{-1}\Gamma_j), \\
T_i | u_i, Z_{ij} = 1 &\overset{\text{ind}}{\sim} NT(b, u_i^{-1}, (b, \infty)), \\
U_i | Z_{ij} = 1 &\overset{\text{ind}}{\sim} H(u_i; \boldsymbol{\nu})
\end{aligned}$$

and

$$\mathbf{Z}_i \overset{\text{iid}}{\sim} M(1; p_1 \ldots, p_g), \qquad i = 1, \ldots, n, \; j = 1, \ldots, g,$$

where

$$\Gamma_j = (1 - \delta_j^2)\sigma_j^2, \quad \Delta_j = \sigma_j \delta_j, \quad \text{and} \quad \delta_j = \frac{\lambda_j}{\sqrt{1 + \lambda_j^2}}.$$

Parameter estimation via the EM algorithm

We have that the complete-data log-likelihood function is

$$\ell_c(\boldsymbol{\Theta}) = c + \sum_{i=1}^{n}\sum_{j=1}^{g} Z_{ij}\Big(\log(p_j) - \frac{1}{2}\log\Gamma_j - \frac{u_i}{2\Gamma_j}(y_i - \mu_{ij} - \Delta_j t_i)^2 + \log(h(u_i|\boldsymbol{\nu})) + \log(NT(t_i|b, u_i^{-1}, (b,\infty))\Big),$$

where c is a constant that is independent of the parameter vector $\boldsymbol{\Theta}$. Defining the following quantities

$$\begin{aligned} \widehat{z}_{ij} &= \mathrm{E}[Z_{ij}|\widehat{\boldsymbol{\Theta}}, y_i], \\ \widehat{s}_{1ij} &= \mathrm{E}[Z_{ij}U_i|\widehat{\boldsymbol{\Theta}}, y_i], \\ \widehat{s}_{2ij} &= \mathrm{E}[Z_{ij}U_iT_i|\widehat{\boldsymbol{\Theta}}, y_i] \end{aligned}$$

and

$$\widehat{s}_{3ij} = \mathrm{E}[Z_{ij}U_iT_i^2|\widehat{\boldsymbol{\Theta}}, y_i]$$

These conditional expectations have analytical forms and can be found, for instance, in Basso et al. (2010)

Thus, the $Q-$function is given by

$$\begin{aligned}
Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(k)}) &= c + \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\widehat{z}_{ij}^{(k)}(\log(p_j) - \frac{1}{2}\log|\Gamma_j|) - \frac{1}{2\Gamma_j}\left(\widehat{s}_{1ij}^{(k)}(y_i - \mu_{ij})^2\right.\right. \\
&- \left. 2(y_i - \mu_{ij})\Delta_j\widehat{s}_{2ij}^{(k)} + \Delta_j^2\widehat{s}_{3ij}^{(k)}\right) + E[Z_{ij}\log(h(U_i|\boldsymbol{\nu}))|\widehat{\boldsymbol{\Theta}}^{(k)}, y_i] \\
&+ \left. E[Z_{ij}\log(TN(T_i|b, u_i^{-1}, (b,\infty)))|\widehat{\boldsymbol{\Theta}}^{(k)}, y_i]\right).
\end{aligned}$$

Thus, the EM-type algorithm for ML estimation of $\boldsymbol{\Theta}$ is defined as follows:

E-step: Given a current estimate $\widehat{\boldsymbol{\Theta}}^{(k)}$, compute $\widehat{z}_{ij}$, $\widehat{s}_{1ij}$, $\widehat{s}_{2ij}$, $\widehat{s}_{3ij}$, for $i = 1, \ldots, n$ and $j = 1, \ldots, g$.

M-step: Update $\widehat{\boldsymbol{\Theta}}^{(k)}$ by maximizing $Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(k)}) = E[\ell_c(\boldsymbol{\Theta})|\mathbf{y}, \widehat{\boldsymbol{\Theta}}^{(k)}]$ over $\boldsymbol{\Theta}$, which leads to the following closed-form expressions:

$$
\begin{aligned}
\widehat{p}_j^{(k+1)} &= n^{-1}\sum_{i=1}^{n}\widehat{z}_{ij}^{(k)}, \\
\widehat{\vartheta}_j^{(k+1)} &= \frac{\sum_{i=1}^{n}[\widehat{s}_{1ij}^{(k)}(y_i - \mathbf{x}_i^\top\widehat{\boldsymbol{\beta}}) - \widehat{\Delta}_j\widehat{s}_{2ij}^{(k)}]}{\sum_{i=1}^{n}\widehat{s}_{1ij}^{(k)}}, \\
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^{n}\sum_{j=1}^{g}\frac{\widehat{s}_{1ij}^{(k)}\mathbf{x}_i\mathbf{x}_i^\top}{\widehat{\Gamma}_j}\right)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{g}\frac{1}{\widehat{\Gamma}_j}[\widehat{s}_{1ij}^{(k)}(y_i - \widehat{\vartheta}_j^{(k)}) - \widehat{\Delta}_j\widehat{s}_{2ij}^{(k)}]\mathbf{x}_i, \\
\widehat{\Delta}_j^{(k+1)} &= \frac{\sum_{i=1}^{n}(y_i - \widehat{\mu}_{ij}^{(k)})\widehat{s}_{2ij}^{(k)}}{\sum_{i=1}^{n}\widehat{s}_{3ij}^{(k)}}
\end{aligned}
$$

and

$$
\widehat{\Gamma}_j^{(k+1)} = \sum_{i=1}^{n}\left(\widehat{s}_{1ij}^{(k)}(y_i - \widehat{\mu}_{ij}^{(k)})^2 - 2(y_i - \widehat{\mu}_{ij}^{(k)})\widehat{\Delta}_j s_{2ij}^{(k)} + \widehat{\Delta}_j^2\widehat{s}_{3ij}^{(k)}\right) / \sum_{i=1}^{n}\widehat{z}_{ij}^{(k)}.
$$

CML-step: Update $\widehat{\boldsymbol{\nu}}^{(k)}$ by maximizing the current marginal log-likelihood function, obtaining

$$\boldsymbol{\nu}^{(k+1)} = \text{argmax}_{\boldsymbol{\nu}} \sum_{i=1}^{n} \log \left(\sum_{j=1}^{n} p_j \phi_{SMSN} \left(y_i | \mu_{ij}^{(k+1)} + b(\boldsymbol{\nu}) \Delta_j^{(k+1)}, \sigma_j^{2(k+1)}, \lambda_j^{(k+1)}, \boldsymbol{\nu} \right) \right).$$

We can also obtain a estimative of β_0 as

$$\widehat{\beta}_0^{(k+1)} = \sum_{j=1}^{n} \widehat{p}_j^{(k+1)} \widehat{\vartheta}_j^{(k+1)}$$

and for μ_j, $j = 1, \ldots, g$, as

$$\widehat{\mu}_j = \widehat{\vartheta}_j - \widehat{\beta}_0.$$

This process is iterated until a suitable convergence rule is satisfied $|\ell(\widehat{\boldsymbol{\Theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\Theta}}^{(k)}) - 1|$, is small enough.

Approximated standard errors

In this work to approximate the asymptotic covariance matrix of $\widehat{\boldsymbol{\Theta}}$, where $\boldsymbol{\Theta} = (\boldsymbol{\beta}^\top, (p_1, \ldots, p_{g-1})^\top, \boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_g^\top)^\top$ with $\boldsymbol{\theta}_j = (\vartheta_j, \sigma_j^2, \lambda_j)^\top$, we use an alternative method suggested by Basford et al. (1997), which consists of approximating the inverse of the covariance matrix by:

$$\mathbf{I}_o(\widehat{\boldsymbol{\Theta}}) = \sum_{i=1}^{n} \widehat{\mathbf{s}}_i \widehat{\mathbf{s}}_i^\top,$$

where

$$\widehat{\mathbf{s}}_i = \frac{\partial}{\partial \boldsymbol{\Theta}} \log \left(\sum_{j=1}^{g} \phi_{SMSN}(y_i | \mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu}) \right) \Bigg|_{\boldsymbol{\Theta} = \widehat{\boldsymbol{\Theta}}},$$

where

$$\widehat{\mathbf{s}}_i = (\widehat{s}_{i,\boldsymbol{\beta}}, \widehat{s}_{i,p_1}, \ldots, \widehat{s}_{i,p_{g-1}}, \widehat{s}_{i,\vartheta_1}, \ldots, \widehat{s}_{i,\vartheta_g}, , \widehat{s}_{i,\sigma_1^2}, \ldots, \widehat{s}_{i,\sigma_g^2}, \widehat{s}_{i,\lambda_1}, \ldots, \widehat{s}_{i,\lambda_g})^\top.$$

Expressions for the elements $\widehat{s}_{i,\boldsymbol{\beta}}, \widehat{s}_{i,p_j}, \widehat{s}_{i,\vartheta_j}, \widehat{s}_{i,\sigma_j^2}, \widehat{s}_{i,\lambda_j}$, $j = 1, \ldots, g$, are given as follows:

$$\begin{aligned} \widehat{s}_{i,\boldsymbol{\beta}} &= \frac{\sum_{j=1}^{G} p_j D_{\boldsymbol{\beta}}(y_i; \boldsymbol{\Theta}_j)}{f(y_i; \boldsymbol{\Theta})}, \quad \widehat{s}_{i,\vartheta_j} = \frac{p_r D_{\vartheta_j}(y_i; \boldsymbol{\Theta}_j)}{f(y_i; \boldsymbol{\Theta})}, \\ \widehat{s}_{i,\sigma_j^2} &= \frac{p_r D_{\sigma_j^2}(y_i; \boldsymbol{\Theta}_j)}{f(y_i; \boldsymbol{\Theta})}, \widehat{s}_{i,\lambda_j} = \frac{p_r D_{\lambda_j}(y_i; \boldsymbol{\Theta}_j)}{f(y_i; \boldsymbol{\Theta})} \end{aligned}$$

and

$$\widehat{s}_{i,p_j} = \frac{1}{f(y_i; \boldsymbol{\Theta})} \left(\phi_{SMSN}(y_i | \mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu}) - \phi_{SMSN}(y_i | \mu_{ig} + b\Delta_g, \sigma_g^2, \lambda_g, \boldsymbol{\nu}) \right),$$

where

$$D_{\vartheta_j}(y_i; \boldsymbol{\theta}_j) = \frac{\partial}{\partial \vartheta_j} (\phi_{SMSN}(y_i | \mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu})).$$

Following Basso et al.(2010), we define

$$I_{ij}^{\Phi}(w) \quad = \quad \int_0^{\infty} \kappa^{-w}(u_i)\exp\{-\frac{1}{2}\kappa^{-1}(u_i)d_{ij}\}\Phi(\kappa^{-1/2}(u_i)A_{ij})dH(u_i)$$

and

$$I_{ij}^{\phi}(w) \quad = \quad \int_0^{\infty} \kappa^{-w}(u_i)\exp\{-\frac{1}{2}\kappa^{-1}(u_i)d_{ij}\}\phi\left(\kappa^{-1/2}(u_i)A_{ij}\right)dH(u_i),$$

where

$$d_{ij} = \frac{(y_i - \mu_{ij} - b\Delta_j)^2}{\sigma_j^2} \quad \text{and} \quad A_{ij} = \lambda_j\frac{(y_i - \mu_{ij} - b\Delta_j)^2}{\sigma_j}, \quad i = 1, \dots, n, \; j = 1, \dots, g.$$

After some algebraic manipulation, we obtain:

$$
\begin{aligned}
D_{\boldsymbol{\beta}}(y_i;\Theta_j) &= \frac{2}{\sqrt{2\pi\sigma_j^2}}\left[\sigma^{-2}(y_i-\mu_{ij}-b\Delta_j)I_{ij}^{\Phi}(3/2)-\sigma_j^{-1}\lambda_j I_{ij}^{\phi}(1)\right]\mathbf{x}_i,\\
D_{\vartheta_j}(y_i;\Theta_j) &= \frac{2}{\sqrt{2\pi\sigma_j^2}}\left[\sigma_j^{-2}(y_i-\mu_{ij}-b\Delta_j)I_{ij}^{\Phi}(3/2)-\sigma_j^{-1}\lambda_j I_{ij}^{\phi}(1)\right],\\
D_{\sigma_j^2}(y_i;\Theta_j) &= \frac{1}{\sqrt{2\pi\sigma_j^2}}\left[-\sigma_j^{-2}I_{ij}^{\Phi}(1/2)+\sigma_j^{-4}(y_i-\mu_{ij}-b\Delta_j)^2 I_{ij}^{\Phi}(3/2)\right.\\
&+ \left.\sigma_j^{-4}(y_i-\mu_{ij}-b\Delta_j)b\Delta_j I_{ij}^{\Phi}(3/2)-\lambda_j\sigma_j^{-3}(y_i-\mu_{ij})I_{ij}^{\phi}(1)\right],\\
D_{\lambda_j}(y_i;\Theta_j) &= \frac{2}{\sqrt{2\pi}\sigma_j^2}\left[\frac{(y_i-\mu_{ij}-b\Delta_j)b}{(1+\lambda_j^2)^{(3/2)}}I_{ij}^{\Phi}(3/2)+\left((y_i-\mu_{ij}-b\Delta_j)-\frac{b\Delta_j}{1+\lambda_j^2}I_{ij}^{\phi}(1)\right)\right]
\end{aligned}
$$

Original Parameterization

Using the delta method, we find the Hessian matrix for the original parameter vector $\boldsymbol{\Theta}^* = (\boldsymbol{\beta}^\top, (p_1, \ldots, p_{g-1})^\top, \beta_0, \mu_1, \ldots, \mu_g, \sigma_1^2, \ldots, \sigma_g^2, \lambda_1, \ldots, \lambda_g)^\top$, as

$$\mathbf{I}_o(\widehat{\boldsymbol{\Theta}}^*) = \mathbf{J}(\boldsymbol{\Theta}^*|\boldsymbol{\Theta})\mathbf{I}_o(\widehat{\boldsymbol{\Theta}})^{-1}\mathbf{J}(\boldsymbol{\Theta}^*|\boldsymbol{\Theta})^\top,$$

where $\mathbf{J}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^*)$ is the Jacobian matrix of order $(p+4g-1) \times (p+4g)$, defined by:

$$J(\boldsymbol{\Theta}^*|\boldsymbol{\Theta}) = \frac{\partial \boldsymbol{\Theta}^*}{\partial \boldsymbol{\Theta}} = \begin{pmatrix} \mathbf{I}_p & \mathbf{0}_{p\times(g-1)} & \mathbf{0}_{p\times g} & \mathbf{0}_{p\times g} & \mathbf{0}_{p\times g} \\ \mathbf{0}_{(g-1)\times p} & \mathbf{I}_{(g-1)\times(g-1)} & \mathbf{0}_{(g-1)\times g} & \mathbf{0}_{(g-1)\times g} & \mathbf{0}_{(g-1)\times g} \\ \mathbf{0}_{1\times p} & \mathbf{0}_{1\times(g-1)} & \mathbf{1}_{1\times g} & \mathbf{0}_{1\times g} & \mathbf{0}_{1\times g} \\ \mathbf{0}_{g\times p} & \boldsymbol{\gamma} & \mathbf{I}_g & \mathbf{0}_{g\times g} & \mathbf{0}_{g\times g} \\ \mathbf{0}_{g\times p} & \mathbf{0}_{g\times(g-1)} & \mathbf{0}_{g\times g} & \mathbf{I}_g & \mathbf{0}_{g\times g} \\ \mathbf{0}_{g\times p} & \mathbf{0}_{g\times(g-1)} & \mathbf{0}_{g\times g} & \mathbf{0}_{g\times g} & \mathbf{I}_g \end{pmatrix},$$

where $\boldsymbol{\gamma} = \dfrac{\partial \boldsymbol{\mu}}{\partial \mathbf{p}^\top} = \mathbf{A}(\mathbf{p}^{-1})^\top$ is a matrix of dimension $g \times (g-1)$, with $\mathbf{A} = (\mu_g - \mu_1, \ldots, \mu_g - \mu_{g-1})^\top$ and $\mathbf{p}^{-1} = (1/p_1, \ldots, 1/p_{g-1})^\top$.

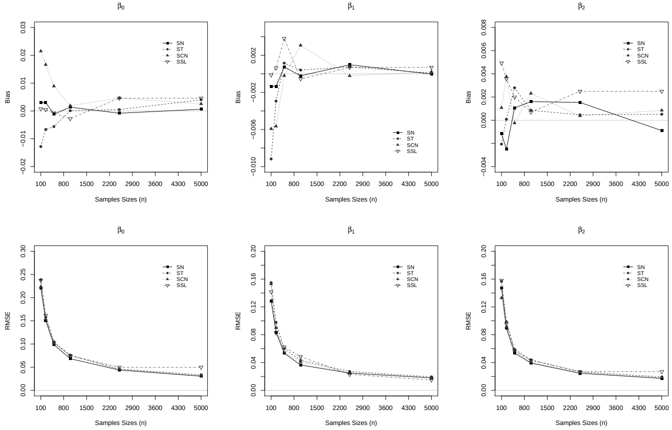
Simulations study: 3 scenarios

- The computational procedures were implemented using the R package `FMsmsnReg()`
- **Scenario 1:** consistency of the approximate standard errors
- **Scenario 2:** asymptotic properties of the EM estimates
- **Scenario 3:** performance of the estimates for FM-SMSN-LR models in the presence of outliers

Consistency of the approximate standard errors (simulation study 1)

		Scenario 1: $(\sigma_1^2 = 0.2,\ \sigma_2^2 = 0.4)$				Scenario 2: $(\sigma_1^2 = \sigma_2^2 = 2)$			
Par.		SN	ST	SCN	SSL	SN	ST	SCN	SSL
β_1	Mean	-4.0002	-3.9985	-3.9996	-3.9947	-3.9949	-3.9958	-3.9963	-4.0005
	IM SE	0.0368	0.0418	0.0402	0.0423	0.0889	0.1021	0.0974	0.0985
	MC Sd	0.0365	0.0426	0.0403	0.0449	0.0899	0.1076	0.0950	0.1031
	COV	94.7%	94.2%	95.5%	95.0%	95.0%	92.9%	95.4%	93.3%
β_2	Mean	-3.0012	-2.9998	-3.0014	-2.9938	-2.9994	-2.9989	-2.9967	-3.0013
	IM SE	0.0374	0.0424	0.0410	0.0432	0.0859	0.1005	0.0975	0.1020
	MC Sd	0.0370	0.0442	0.0413	0.0430	0.0836	0.1046	0.0977	0.1109
	COV	95.6%	93.7%	94.0%	96.0%	96.2%	94.4%	94.2%	92.0%
μ_1	Mean	-4.0026	-3.9945	-4.0040	-4.0166	-4.0295	-3.9806	-4.0899	-3.9924
	IM SE	0.0853	0.0800	0.0894	0.0854	0.1396	0.2782	0.1896	0.2531
	MC Sd	0.0691	0.0876	0.0744	0.0859	0.1111	0.3161	0.2483	0.2202
	COV	98.2%	99.8%	98.6%	98.6%	97.3%	92.3%	84.5%	94.8 %
μ_2	Mean	0.9992	1.0012	1.0007	0.9945	0.9990	1.0103	1.0391	0.9955
	IM SE	0.0837	0.0878	0.0862	0.0873	0.0744	0.1098	0.0861	0.0983
	MC Sd	0.0630	0.0625	0.0656	0.0625	0.0692	0.1000	0.1060	0.0813
	COV	98.3%	99.7%	98.4%	99.0%	96.7%	96.7%	86.4%	97.7%

Asymptotic properties of the EM estimates (simulation study 2)



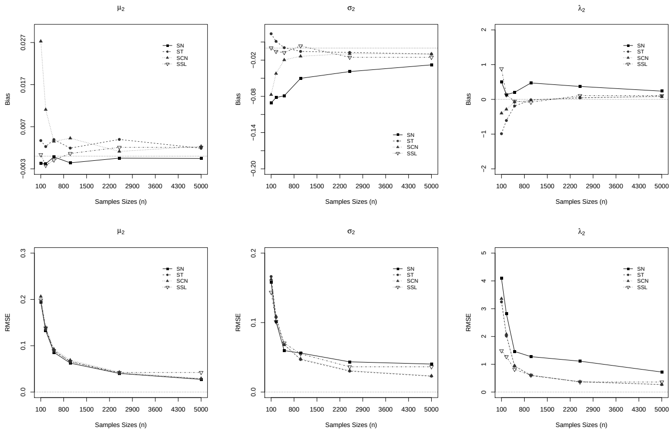


Figure: Simulation study 2. Average bias (first row) and average RMSE (second row) of the estimates of μ_2,σ_2, λ_2.

Application

The data set comes from the Australian Institute of Sport (AIS) and consists of measurements taken on 202 athletes. Here, we focus on percent body fat (Bfat), which is assumed to be explained by the name sum of skin folds (ssf) and Height in cm (Ht). Thus, we consider the following FM-SMSN-LR model:

$$Bfat_i = \beta_0 + \beta_1 ssf_i + \beta_2 Ht_i + \varepsilon_i,$$

where ε_i belongs to the FM-SMSN family for $i = 1, \ldots, 202$.

Package '*FMsmsnReg*'

Application:

```
library(FMsmsnReg)
data(AIS)
attach(AIS)
x1< − cbind(1,SSF,HT)
y < − Bfat
```

Fits a linear Regression Model with Finite Mixtures of Skew Contaminated Normal

```
parCN = FMsmsnReg(y, x1, g=2, get.init = TRUE, criteria = TRUE, group = FALSE, family = "Skew.cn", error = 0.00004, iter.max = 5000,obs.prob= FALSE,kmeans.param = NULL,show.converge=FALSE,cp=0.5)
```

Fits a linear Regression Model with Finite Mixtures of Skew normal

```
parSN = FMsmsnReg(y, x1, g=2, get.init = TRUE, criteria = TRUE, group = FALSE,family = "Skew.normal", error = 0.00004, iter.max = 5000,obs.prob= FALSE, kmeans.param = NULL,show.converge=FALSE,cp=0.5)
```

Model comparison criteria

Model	g	m	log-lik	AIC	BIC
FM-N	1	5	-367.2395	744.7850	745.1792
FM-N	2	8	-359.2902	735.3265	735.7009
FM-N	3	11	-355.2892	733.9679	734.1192
FM-T	1	6	-363.9525	738.2111	738.6053
FM-T	2	9	-358.2494	733.2449	733.6194
FM-T	3	12	-356.3237	736.0369	736.1881
FM-SN	1	6	-363.0346	738.5001	738.9097
FM-SN	2	10	-356.3079	733.7675	734.0164
FM-SN	3	14	-354.1438	738.5336	738.2486
FM-SN	4	18	-353.1388	746.0152	744.7987
FM-SN	5	22	-352.2579	754.1695	751.5973
FM-ST	1	7	-360.7632	736.1038	736.5070
FM-ST	2	11	-353.9696	**731.3286**	**731.4799**
FM-ST	3	15	-353.8492	740.2790	739.7994
FM-ST	4	19	-352.3138	746.8034	745.2888
FM-ST	5	23	-351.7865	755.7752	752.7944
FM-SCN	1	8	-357.0375	738.5001	738.9097
FM-SCN	2	12	-353.7235	733.0978	733.1278
FM-SCN	3	16	-354.1656	743.2717	742.5722
FM-SCN	4	20	-352.0380	748.7169	746.8773
[illegible]	[illegible]	[illegible]	-352.8184	760.4164	756.9983

Application results

Table: AIS data. Parameter estimates of the FM-SMSN- LR models with $g = 2$. SE denotes the corresponding standard errors obtained via the information-based matrix.

Parameter	FM-SN		FM-ST		FM-SCN		FM-SSL	
	ML	SE	ML	SE	ML	SE	ML	SE
β_0	14.7241	0.0001	14.51593	0.00253	14.6622	0.0025	14.7475	0.0025
β_1	0.1799	0.0012	0.17972	0.00850	0.1805	0.0089	0.1796	0.0091
β_2	-0.0757	0.1302	-0.07536	0.19264	-0.0757	0.1458	-0.0754	0.1513
p_1	0.1543	0.9295	0.15418	1.04192	0.1483	1.0841	0.1514	1.0393
μ_1	2.5504	2.2932	1.93244	4.00942	2.3654	3.8355	2.3891	3.9553
μ_2	-0.4652	1.8546	-0.35226	2.94875	-0.4120	2.5091	-0.4263	2.6266
σ_1^2	0.8483	0.5074	3.80681	1.57056	2.2957	1.6255	2.3158	1.6615
σ_2^2	2.2793	0.4021	1.06550	11.56693	1.1240	7.1021	0.9740	7.0029
λ_1	0.1624	0.8467	-5.70438	0.52991	-3.5415	0.4408	-4.8612	0.3724
λ_2	-2.2318	1.7509	-0.62860	9.52263	-1.0111	7.9389	-1.0144	11.9961
ν	-	-	7.45874	-	0.2270	-	2.3036	-
γ	-	-	-		0.3075	-	-	

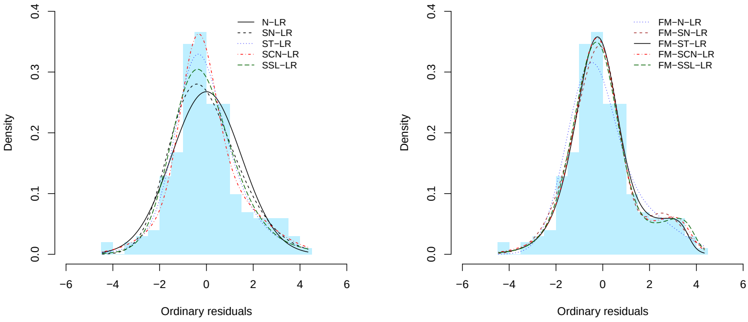


Figure: AIS dataset. Histogram of ordinary residuals superimposed for FM-SMSN-LR residual density model for a) $g = 1$ and b) $g = 2$ components.

Conclusions

- Instead semiparametric methods based on kernel density estimators, we propose parametric method based in mixtures of SMSN distributions to model the unknown form of the error term.
- This approach allows us to model data with great flexibility, accommodating simultaneously multimodality, skewness and heavy tails for the random error in linear regression models.
- The proposed methods are implemented using the R package `FMsmsnReg()`, providing practitioners with a convenient tool for further applications in their domain.
- The proposed methods can be extended to multivariate settings using the multivariate SMSN class of distributions (Branco and Dey, 2001; Cabral, Lachos and Prates, 2012), such as the recent proposals of Soffritti and Galimberti (2011) and Galimberti and Soffritti (2014).
- The proposed methods can be also extended to mixtures of SMSN distributions in the context of censored responses based on recent approaches by Caudill (2012) and Karlsson and Laitila (2014).

References

- Basso, R. M., Lachos, V. H., Cabral, C. R. B., Ghosh, P., 2010. Robust mixture modeling based on scale mixtures of skew-normal distributions. Computational Statistics & Data Analysis 54 (12), 2926–2941.
- Basford, K., Greenway, D., McLachlan, G., Peel, D., 1997. Standard errors of fitted component means of normal mixtures. Computational Statistics 12, 1–18.
- Benites, L., Maehara, R., Lachos, V. H., 2016. CensMixReg: Censored Linear Mixture Regression Models. R package version 1.0. URL http://CRAN.R-project.org/package=FMsmsnReg
- Cabral, C. R. B., Lachos, V. H., Prates, M. O., 2012b. Multivariate mixture modeling using skew-normal independent distributions. Computational Statistics & Data Analysis 56 (1), 126–142.
- Caudill, S. B., 2012. A partially adaptive estimator for the censored regression model based on a mixture of normal distributions. Statistical Methods & Applications 21, 121–137.
- Galimberti, G., Soffritti, G., 2014. A multivariate linear regression analysis using finite mixtures of t distributions. Computational Statistics and Data Analysis 71, 138–150.
- Karlsson, M., Laitila, T., 2014. Finite mixture modeling of censored regression models. Statistical Papers 55, 627–642.
- Soffritti, G., Galimberti, G., 2011. Multivariate linear regression with non-normal errors: a solution based on mixture models. Statistics and Computing, 21, 523–536.
- Zeller, C. B., Cabral, C. R., Lachos, V. H., 2016. Robust mixture regression modeling based on scale mixtures of skew-normal distributions. TEST, 25, 375-396.

Thank you!