Heavy-tailed longitudinal regression models for censored data: A likelihood based perspective

Larissa Avila Matos

10th International Conference of the ERCIM WG on Computational and Methodological Statistics

Joint work with: Victor H. Lachos, Tsung-I Lin and Luis M. Castro

Support: FAPESP

10 de maio de 2019

Summary

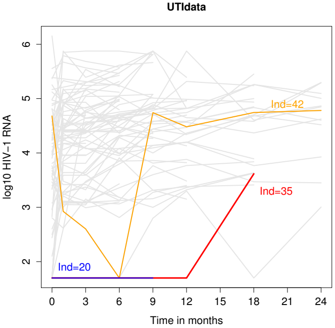
Summary

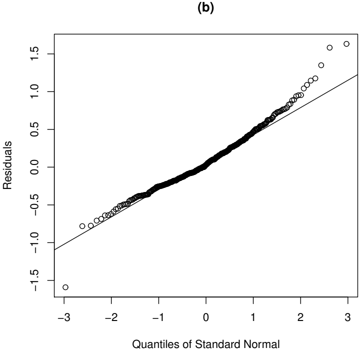
Motivation

- In AIDS studies it is quite common to observe viral load measurements that are collected irregularly over time. Moreover, these measurements can be subjected to some upper and/or lower detection limits depending on the quantification assays.
- A complication arises when these continuous repeated measures have a heavy-tailed behavior.
- For such data structures, we propose a robust censored linear model based on the class of multivariate scale mixtures of normal distributions.
- A recent proposal uses the Student-t distribution (Garay et al. (2015))

Motivating data

- Example: **Unstructured treatment interruption- UTI data**
    - The viral loads were monitored at 0, 1, 3, 6, 9, 12, 18, and 24 months after the treatment interruption
    - 72 perinatally HIV-infected children (Saitoh et al. 2008);
    - 7% of the data are left-censored (362 observations).





Motivating data

$\log_{10}$HIV-1 RNA		month 0	month 1	month 3	month 6	month 9	month 12	month 18	month 24
$\log_{10}$HIV-1 RNA	month 0		0.4877	0.4100	0.4052	0.4820	0.4435	0.3441	0.6529
	month 1	0.4877		0.9145	0.8551	0.8455	0.6978	0.7090	0.6140
	month 3	0.4100	0.9145		0.9255	0.8638	0.7209	0.7601	0.6301
	month 6	0.4052	0.8551	0.9255		0.8238	0.6490	0.6548	0.5314
	month 9	0.4820	0.8455	0.8638	0.8238		0.9185	0.7642	0.8061
	month 12	0.4435	0.6978	0.7209	0.6490	0.9185		0.6646	0.6897
	month 18	0.3441	0.7090	0.7601	0.6548	0.7642	0.6646		0.8947
	month 24	0.6529	0.6140	0.6301	0.5314	0.8061	0.6897	0.8947	

Censored longitudinal models with normal distribution

- Samson *et al.* (2006) [*Computational Statistical & Data Analysis*]
- Vaida *et al.* (2007) [*Computational Statistical & Data Analysis*]
- Vaida & Liu (2009) [*Journal of Computational and Graphical Statistics*]
- Matos *et al.* (2013b) [*Computational Statistical & Data Analysis*]

Censored longitudinal models with heavy-tailed distribution

- Lachos *et al.* (2011) [*Biometrics*]
- **Garay *et al.* (2015)** [*Statistical Methods in Medical Research*]
- **Matos *et al.* (2013a)** [*Statistica Sinica*]

Summary

Scale mixture of normal distributions (SMN)

Andrews & Mallows (1974); Lange & Sinsheimer (1993)

Stochastic representation

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

where,

- $\mu \in \mathbb{R}$ is a location vector;
- $Z \sim N(0, \mathbf{\Sigma})$;
- U is a positive random variable with cumulative distribution function (cdf) $\mathrm{H}(u|\boldsymbol{\nu})$ and probability density function (pdf) $h(u|\boldsymbol{\nu})$ independent of Z;
- $\kappa(\mathrm{U})$ is the weight function;
- Notation: $\mathbf{y} \sim \mathrm{SMN}_p(\boldsymbol{\mu}, \mathbf{\Sigma}; \mathrm{H})$.

$$\begin{aligned} \mathbf{y}|\mathrm{U} = u &\sim N(\boldsymbol{\mu}, \kappa(u)\mathbf{\Sigma}), \\ \mathrm{U} = u &\sim h(u|\boldsymbol{\nu}). \end{aligned} \tag{2}$$

Scale mixture of normal distributions (SMN)

Special cases: $\mathbf{y} \in \mathbb{R}^p$ and $\kappa(u) = 1/u$;

- **The multivariate normal distribution**
  - $P(U = 1) = 1$;
  - Distribution function: $\mathrm{N}(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

- **The multivariate Student's-t distribution**
  - $\mathrm{U} = Gamma(\nu/2, \nu/2)$;
  - Distribution function:

$$\mathrm{T}(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}} \nu^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{d}{\nu}\right)^{-(p+\nu)/2},$$

where $d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})$.

The multivariate slash distribution

- $\mathrm{U} = Beta(\nu, 1)$;
- Distribution function:

$$\mathrm{SL}(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \nu \int_0^1 u^{\nu-1} \phi_p(\mathbf{y}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) du, \quad u \in (0,1), \quad \nu > 0.$$

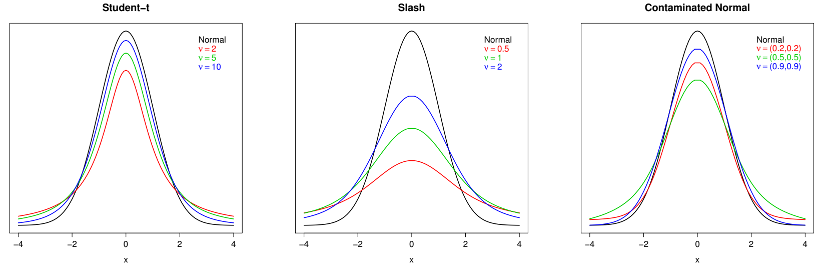
The multivariate contaminated normal distribution

- U is a discrete random variable taking one of two states and with probability function given by $h(u|\boldsymbol{\nu}) = \nu \mathbb{I}_{\{\gamma\}}(u) + (1-\nu)\mathbb{I}_{\{1\}}(u)$ and $\boldsymbol{\nu} = (\nu, \gamma)$;
- Distribution function:

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

The parameter ν can be interpreted as the proportion of outliers while γ may be interpreted as a scale factor.

Scale mixture of normal distributions (SMN)



Summary

The data can be fit using the model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n. \tag{3}$$

Considering the left censored case, we have

$$\begin{aligned} y_{ij} &\leq V_{ij} \ \text{ se } \ C_{ij} = 1, \\ y_{ij} &= V_{ij} \ \text{ se } \ C_{ij} = 0, \end{aligned}$$

where,

- $\mathbf{y}_i$ is a $n_i \times 1$ vector containing the observations for subject i measured at particular time points $\mathbf{t}_i = (t_{i1}, \ldots, t_{in_i})$;
- $\boldsymbol{\beta}$ is the vector of fixed effects of dimension $(p \times 1)$;
- $\mathbf{X}_i$ is an $n_i \times p$ design matrix;
- $\boldsymbol{\epsilon}_i$ is the vector of random errors of dimension $(n_i \times 1)$, $\boldsymbol{\epsilon}_i \overset{\text{ind.}}{\sim} \text{SMN}_{n_i}(\mathbf{0}, \boldsymbol{\Omega}_i; \text{H})$, where $\boldsymbol{\Omega}_i = \sigma^2 \text{E}_i$.

Damped exponential correlation (DEC):

$$\mathrm{E}_i = \mathrm{E}_i(\boldsymbol{\phi}, \mathbf{t}_i) = \left[\phi_1^{|t_{ij}-t_{ik}|^{\phi_2}}\right], \; i = 1, \ldots, n, \;\; j, k = 1, \ldots, n_i, \tag{4}$$

For the DEC structure, we have that:

(a) if $\phi_2 = 0$, then E_i generates the compound symmetry correlation structure;

(b) when $0 < \phi_2 < 1$, then E_i presents a decay rate between the compound symmetry structure and the first-order AR (AR (1)) model;

(c) if $\phi_2 = 1$, then E_i generates an AR(1) structure;

(d) when $\phi_2 > 1$, E_i presents a decay rate faster than the AR(1) structure; and

(e) if $\phi_2 \to \infty$, then E_i represents the first-order moving average model, MA(1).

Using the stochastic representation (1), the hierarchical representation (two-stages) of the linear regression model defined in(3) is given by

$$
\begin{aligned}
\mathbf{y}_i \mid \mathrm{U}_i = u_i &\overset{\text{ind.}}{\sim} \mathrm{N}_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \kappa(u_i)\boldsymbol{\Omega}_i), \\
\mathrm{U}_i &\overset{\text{iid.}}{\sim} h(u_i|\boldsymbol{\nu}). \qquad (5)
\end{aligned}
$$

Recall that we are interested in the case where left-censored observations can occur. That is, the observed data for the i-th subject is represented by $(\mathbf{V}_i, \mathbf{C}_i)$, where

- $\mathbf{V}_i$ is the vector of uncensored observation or limit of quantification; and
- $\mathbf{C}_i$ is the vector of censoring indicator whose value equals one if censored observation and zero if uncensored observation,

such that, considering the left censored case, we have that

$$\begin{aligned} y_{ij} &\leq V_{ij} \quad \text{se} \quad C_{ij} = 1, \\ y_{ij} &= V_{ij} \quad \text{se} \quad C_{ij} = 0, \end{aligned}$$

with $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$.

- Frequentist inference on the parameter vector $\theta = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\phi}^\top)$ is based on the marginal distribution for $\mathbf{y}_i$. For the SMN-CR model with complete data, we have that, marginally, $\mathbf{y}_i \overset{\text{ind.}}{\sim} \text{SMN}_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Omega}_i, \boldsymbol{\nu})$,

Proposition

Let $\mathbf{y}$ *be partitioned as* $\mathbf{y}_i = vec(\mathbf{y}_i^o, \mathbf{y}_i^c)$ *with* $dim(\mathbf{y}_i^o) = n_i^o$, $dim(\mathbf{y}_i^c) = n_i^c$ *and* $n_i^o + n_i^c = n_i$, *where* $vec(\cdot)$ *denotes the operator which stacks vectors or matrices of the same number of columns and* $C_{ij} = 0$ *for all elements in* $\mathbf{y}_i^o$, *and 1 for all elements in* $\mathbf{y}_i^c$. *Let* $\mathbf{V}_i$, $\mathbf{X}_i$, *and* $\boldsymbol{\Omega}_i$ *also be partitioned as follows:* $\mathbf{V}_i = vec(\mathbf{V}_i^o, \mathbf{V}_i^c)$, $\mathbf{X}_i^\top = (\mathbf{X}_i^o, \mathbf{X}_i^c)$, *and* $\boldsymbol{\Omega}_i = \begin{pmatrix} \boldsymbol{\Omega}_i^{oo} & \boldsymbol{\Omega}_i^{oc} \\ \boldsymbol{\Omega}_i^{co} & \boldsymbol{\Omega}_i^{cc} \end{pmatrix}$. *Then, we have*

$$\mathbf{y}_i \mid u_i \sim N_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \kappa(u_i)\boldsymbol{\Omega}_i),$$

where

$$\mathbf{y}_i^o \mid u_i \sim N_{n_i^o}(\mathbf{X}_i^o\boldsymbol{\beta}, \kappa(u_i)\boldsymbol{\Omega}_i^{oo}) \quad \textit{and} \quad \mathbf{y}_i^c \mid \mathbf{y}_i^o, u_i \sim N_{n_i^c}(\boldsymbol{\mu}_i, \kappa(u_i)\mathbf{S}_i), \tag{6}$$

with $\boldsymbol{\mu}_i = \mathbf{X}_i^c\boldsymbol{\beta} + \boldsymbol{\Omega}_i^{co}(\boldsymbol{\Omega}_i^{oo})^{-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta})$ *and* $\mathbf{S}_i = \boldsymbol{\Omega}_i^{cc} - \boldsymbol{\Omega}_i^{co}(\boldsymbol{\Omega}_i^{oo})^{-1}\boldsymbol{\Omega}_i^{oc}$.

- The likelihood function is given by $L(\boldsymbol{\theta}) = \prod_{i=1}^{n} L_i(\boldsymbol{\theta})$, where

Proposition

Let $\Phi_{n_i}(\mathbf{u};\mathbf{a},\mathbf{A})$ *and* $\phi_{n_i}(\mathbf{u};\mathbf{a},\mathbf{A})$ *be the cdf (left tail) and pdf, respectively, of* $N_{n_i}(\mathbf{a},\mathbf{A})$ *computed at* $\mathbf{u}$. *The likelihood function for the i-th subject is given by*

$$\begin{aligned} L_i(\boldsymbol{\theta}) &= f(\mathbf{y}_i^o \mid \boldsymbol{\theta}) f(\mathbf{y}_i^c \leq \mathbf{V}_i^c \mid \mathbf{y}_i^o, \boldsymbol{\theta}) \\ &= \int_0^{\infty} f(\mathbf{y}_i^o \mid u_i, \boldsymbol{\theta}) f(\mathbf{y}_i^c \leq \mathbf{V}_i^c \mid u_i, \mathbf{y}_i^o, \boldsymbol{\theta}) dH(u_i) \\ &= \int_0^{\infty} \phi_{n_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^o \boldsymbol{\beta}, \kappa(u_i)\boldsymbol{\Omega}_i^{oo}) \Phi_{n_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \kappa(u_i)\mathbf{S}_i) dH(u_i). \end{aligned} \tag{7}$$

- The log-likelihood function is given by $\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{n} \{\log L_i(\boldsymbol{\theta})\}$.

The likelihood function for special elements of the SMN class are given by.

(a) *(normal)* If U is degenerate in 1, i.e., $P(U=1)=1$, then

$$L_i(\boldsymbol{\theta}) \quad = \quad \phi_{n_i^o}(\mathbf{y}_i^o;\mathbf{X}_i^o\boldsymbol{\beta},\kappa(u_i)\boldsymbol{\Omega}_i^{oo})\Phi_{n_i^c}(\mathbf{V}_i^c;\boldsymbol{\mu}_i,\mathbf{S}_i).$$

(b) *(Student's-t)* If $\kappa(u)=1/u$ and U is distributed as $Gamma(\nu/2,\nu/2)$, with $\nu>0$, then

$$L_i(\boldsymbol{\theta}) \quad = \quad t_{n_i^o}(\mathbf{y}_i^o;\mathbf{X}_i^o\boldsymbol{\beta},\boldsymbol{\Omega}_i^{oo},\nu)\mathrm{T}_{n_i^c}\left(\mathbf{V}_i^c;\boldsymbol{\mu}_i,\left(\frac{\nu+\boldsymbol{\delta}}{\nu+n_i^o}\right)\mathbf{S}_i,\nu+n_i^o\right),$$

where $\boldsymbol{\delta}=(\mathbf{y}_i^o-\mathbf{X}_i^o\boldsymbol{\beta})^\top(\boldsymbol{\Omega}_i^{oo})^{-1}(\mathbf{y}_i^o-\mathbf{X}_i^o\boldsymbol{\beta})$.

(c) *(contaminated normal)* If $\kappa(u)=1/u$ and $\mathbf{U}$ is a discrete random variable taking one of two states and with probability function given by $h(u|\boldsymbol{\nu})=\nu\mathbb{I}_{\{\gamma\}}(u)+(1-\nu)\mathbb{I}_{\{1\}}(u)$, then

$$\begin{aligned} L_i(\boldsymbol{\theta}) \quad &= \quad \nu\left[\phi_{n_i^o}(\mathbf{y}_i^o;\mathbf{X}_i^o\boldsymbol{\beta},\gamma^{-1}\boldsymbol{\Omega}_i^{oo})\Phi_{n_i^c}(\mathbf{V}_i^c;\boldsymbol{\mu}_i,\gamma^{-1}\mathbf{S}_i)\right] \\ &+ \quad (1-\nu)\left[\phi_{n_i^o}(\mathbf{y}_i^o;\mathbf{X}_i^o\boldsymbol{\beta},\boldsymbol{\Omega}_i^{oo})\Phi_{n_i^c}(\mathbf{V}_i^c;\boldsymbol{\mu}_i,\mathbf{S}_i)\right]. \end{aligned}$$

Summary

SAEM Algorithm

EM Algorithm - Dempster *et al.* (1977)

Let $\boldsymbol{\theta}$ be the parameter vector and $\boldsymbol{y}_c = (\mathbf{y}^\top, \boldsymbol{q}^\top)$ be the vector of complete data, i.e., the observed data $\mathbf{y}^\top$ and the missing/censored data (or the latent variables, depending on the situation) $\boldsymbol{q}^\top$. The EM algorithm consists basically of two steps: the expectation (E-step) and the maximization (M-step).

- **E-Step:** Calculate the conditional expectation

$$Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = E\left[\ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c) \mid \mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}\right],$$

  where $\widehat{\boldsymbol{\theta}}^{(k)}$ is the estimate of $\boldsymbol{\theta}$ at the k-th iteration.

- **M-Step:** Update $\boldsymbol{\theta}^{(k)}$ according to

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \underset{\theta}{\text{argmax}} Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}).$$

SAEM Algorithm

MCEM Algorithm - Wei & Tanner (1990)

- **E-Step: MC:**

  1. Simulation-step: Draw $\mathbf{q}^{(k,l)}$ $(l = 1, \ldots, m)$ from the conditional distribution $f(\mathbf{q}|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k-1)})$;

  2. Approximation-step: Using $\mathbf{q}^{(k,l)}$ $(l = 1, \ldots, m)$, calculate the conditional expectation $Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)})$ through the approximation,

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \frac{1}{m}\sum_{l=1}^{m} \ell_c(\boldsymbol{\theta}|\mathbf{q}^{(k,l)}, \mathbf{y}).$$

- **M-Step:** Update $\boldsymbol{\theta}^{(k)}$ according to

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \underset{\theta}{\operatorname{argmax}} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}).$$

- **E-Step:**

  1. Simulation-step: Draw $\mathbf{q}^{(k,l)}$ $(l = 1, \ldots, m)$ from the conditional distribution $f(\mathbf{q}|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k-1)})$;

  2. Stochastic-approximation-step: Update $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ according to

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k-1)}) + \delta_k \left[\frac{1}{m} \sum_{l=1}^{m} \ell_c(\boldsymbol{\theta}|\mathbf{q}^{(k,l)}, \mathbf{y}) - Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k-1)}) \right],$$

  where $\ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta} \mid \mathbf{y}_c)$ is the complete log-likelihood function and δ_k is a smoothness parameter, *i.e.*, a decreasing sequence of positive numbers such that $\sum_{k=1}^{\infty} \delta_k = \infty$ and $\sum_{k=1}^{\infty} \delta_k^2 < \infty$.

- **M-Step:** Update $\boldsymbol{\theta}^{(k)}$ according to

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \underset{\theta}{\mathrm{argmax}} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}).$$

- As proposed by Galarza *et al.* (2015), we will consider the following smoothing parameter

$$\delta_k = \begin{cases} 1, & \text{if} \quad 1 \leq k \leq cW; \\ \frac{1}{k-cW}, & \text{if} \quad cW+1 \leq k \leq W, \end{cases} \tag{8}$$

  where,

  - W is the maximum number of iterations; and
  - c is a cut point ($0 \leq c \leq 1$) which determines the percentage of the initial iterations.

Summary

- The complete data log-likelihood function:
  - $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\phi}^\top)^\top$ (ν known);
  - Augmenting data: $\mathbf{y}_c = (\mathbf{V}^\top, \mathbf{C}^\top, \mathbf{y}^\top, \boldsymbol{u}^\top)^\top$;
  - $[\ \mathbf{V} \quad \mathbf{C} \quad \mathbf{y}\ ] \Rightarrow [\ \mathbf{y}\ ]$.

$$
\begin{aligned}
\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) &= \sum_{i=1}^{n} \{\log f(\mathbf{y}_i|u_i) + \log h(u_i|\boldsymbol{\nu})\} \\
&= -\frac{N}{2}\log\sigma^2 - \sum_{i=1}^{n}\frac{1}{2}\log|\mathrm{E}_i| - \sum_{i=1}^{n}\frac{\kappa^{-1}(u_i)}{2\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \mathrm{E}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \\
&+ \sum_{i=1}^{n}\log h(u_i|\nu) + C,
\end{aligned}
$$

with C being a constant that does not depend on the parameter vector $\boldsymbol{\theta}$ and $\sum_{i=1}^{n} n_i = N$.

- Q-function: For the i-th subject,

$$
\begin{aligned}
Q_i\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}\right) &= -\frac{1}{2\sigma^2}E\left[\kappa^{-1}(u_i)(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^\top \mathrm{E}_i^{-1}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})|\mathbf{V},\mathbf{C},\widehat{\boldsymbol{\theta}}^{(k)}\right]\\
&- \frac{n_i}{2}\log\sigma^2-\frac{1}{2}\log|\mathrm{E}_i|\\
&= -\frac{n_i}{2}\log\widehat{\sigma^2}^{(k)}-\frac{1}{2}\log|\widehat{\mathrm{E}}_i^{(k)}|-\frac{1}{2\widehat{\sigma^2}^{(k)}}\left[tr\left(\widehat{\kappa\mathbf{y}_i^2}^{(k)}\widehat{\mathrm{E}}_i^{-1(k)}\right)\right.\\
&+ \left.\widehat{\kappa_i}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\widehat{\mathrm{E}}_i^{-1(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}-2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\widehat{\mathrm{E}}_i^{-1(k)}\widehat{\kappa\mathbf{y}_i}^{(k)}\right],
\end{aligned}
$$

with

$$
\widehat{\kappa\mathbf{y}_i^2}^{(k)} = E\{\kappa^{-1}(u_i)\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\}, \tag{9}
$$

$$
\widehat{\kappa\mathbf{y}_i}^{(k)} = E\{\kappa^{-1}(u_i)\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\}, \tag{10}
$$

$$
\widehat{\kappa_i}^{(k)} = E\{\kappa^{-1}(u_i)|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\}. \tag{11}
$$

- **Simulation-step:** Gibbs Sampler

Step 1. Sample $\mathbf{y}_i^c$ from $f(\mathbf{y}_i^c|\mathbf{V}_i^c, \mathbf{y}_i^o, u_i, \boldsymbol{\theta}^{(k)})$, where

$$f(\mathbf{y}_i^c|\mathbf{V}_i^c, \mathbf{y}_i^o, u_i, \boldsymbol{\theta}^{(k)}) \sim \mathrm{TN}_{n_i^c}(\boldsymbol{\mu}_i, \kappa(u_i)\mathbf{S}_i; \mathbb{A}_i),$$

with $\boldsymbol{\mu}_i = \mathbf{X}_i^c\boldsymbol{\beta} + \boldsymbol{\Omega}_i^{co}(\boldsymbol{\Omega}_i^{oo})^{-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta})$ and $\mathbf{S}_i = \boldsymbol{\Omega}_i^{cc} - \boldsymbol{\Omega}_i^{co}(\boldsymbol{\Omega}_i^{oo})^{-1}\boldsymbol{\Omega}_i^{oc}$. Then, $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i^c}, y_{n_i^c+1}, \ldots, y_{n_i})$ is a sample generated for the n_i^c censored cases and the observed values (uncensored cases).

Step 2. Sample u_i from $f(u_i|\mathbf{y}_i, \boldsymbol{\theta}^{(k)})$.

SMN-CR model

SAEM - E-step

(a) Student's-t, $U_i \sim Gamma(\frac{\nu}{2}, \frac{\nu}{2})$ and $\kappa(u_i) = \frac{1}{u_i}$,

$$f(u_i|\mathbf{y}_i, \boldsymbol{\theta}^{(k)}) \sim Gamma\left(\frac{\nu + n_i}{2}, \frac{\nu + (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})}{2}\right);$$

(b) Slash, $U_i \sim Beta(\nu, 1)$ and $\kappa(u_i) = \frac{1}{u_i}$,

$$f(u_i|\mathbf{y}_i, \boldsymbol{\theta}^{(k)}) \sim TGamma\left(\nu + \frac{n_i}{2}, \frac{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})}{2}, 1\right);$$

(c) Contaminated normal, $h(u|\boldsymbol{\nu}) = \nu\mathbb{I}_{\{\rho\}}(u) + (1-\nu)\mathbb{I}_{\{1\}}(u)$, and $\kappa(u_i) = \frac{1}{u_i}$, $f(u_i|\mathbf{y}_i, \boldsymbol{\theta}^{(k)})$, is a discrete distribution taking values γ with probability $\frac{p_1}{p_1+p_2}$ and 1 with probability $\frac{p_2}{p_1+p_2}$, where

$$p_1 = \nu\gamma^{\frac{n_i}{2}} \exp\left(-\frac{\gamma}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right),$$

$$p_2 = (1-\nu)\exp\left(-\frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right).$$

- **Stochastic-approximation-step:** Since the sequence $(\mathbf{y}_i^{(k,l)}, u_i^{(k,l)})$ for $l = 1, \ldots, m$ is collected at the k-th iteration, we replace the conditional expectations given in (9) –(11) by the following stochastic approximations:

$$\widehat{\kappa \mathbf{y}_i^2}^{(k)} = \widehat{\kappa \mathbf{y}_i^2}^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{l=1}^{m} \mathbf{y}_i^{(k,l)} (\mathbf{y}_i^{(k,l)})^\top - \widehat{\kappa \mathbf{y}_i^2}^{(k-1)} \right], \quad (12)$$

$$\widehat{\kappa \mathbf{y}_i}^{(k)} = \widehat{\kappa \mathbf{y}_i}^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{l=1}^{m} \mathbf{y}_i^{(k,l)} u_i^{(k,l)} - \widehat{\kappa \mathbf{y}_i}^{(k-1)} \right], \quad (13)$$

$$\widehat{\kappa_i}^{(k)} = \widehat{\kappa_i}^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{l=1}^{m} u_i^{(k,l)} - \widehat{\kappa_i}^{(k-1)} \right]. \quad (14)$$

Update $\widehat{\boldsymbol{\theta}}^{(k)}$ by the maximization of $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$, which leads to the following expressions:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^{n} \widehat{\kappa_i}^{(k)} \mathbf{X}_i^\top \widehat{\mathrm{E}}_i^{-1(k)} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^\top \widehat{\mathrm{E}}_i^{-1(k)} \widehat{\kappa \mathbf{y}_i}^{(k)}, \\
\widehat{\sigma^2}^{(k+1)} &= \frac{1}{N} \sum_{i=1}^{n} \left[tr\left(\widehat{\kappa \mathbf{y}_i^2}^{(k)} \widehat{\mathrm{E}}_i^{-1(k)}\right) - 2\widehat{\boldsymbol{\beta}}^{(k)\top} \mathbf{X}_i^\top \widehat{\mathrm{E}}_i^{-1(k)} \widehat{\kappa \mathbf{y}_i}^{(k)} \right. \\
&+ \left. \widehat{\kappa_i}^{(k)} \widehat{\boldsymbol{\beta}}^{(k)\top} \mathbf{X}_i^\top \widehat{\mathrm{E}}_i^{-1(k)} \mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k)} \right], \\
\widehat{\boldsymbol{\phi}}^{(k+1)} &= \underset{\boldsymbol{\phi} \in (0,1)\times\mathbb{R}+}{\operatorname{argmax}} \left(-\frac{1}{2\widehat{\sigma^2}^{(k)}} \left[tr\left(\widehat{\kappa \mathbf{y}_i^2}^{(k)} \widehat{\mathrm{E}}_i^{-1(k)}\right) - 2\widehat{\boldsymbol{\beta}}^{(k)\top} \mathbf{X}_i^\top \widehat{\mathrm{E}}_i^{-1(k)} \widehat{\kappa \mathbf{y}_i}^{(k)} \right. \right. \\
&+ \left. \left. \widehat{\kappa_i}^{(k)} \widehat{\boldsymbol{\beta}}^{(k)\top} \mathbf{X}_i^\top \widehat{\mathrm{E}}_i^{-1(k)} \mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k)} \right] - \frac{1}{2} \log(|\widehat{\mathrm{E}}_i^{-1(k)}|) \right).
\end{aligned}
$$

- Let $\mathbf{y}_i^c$ be the true unobserved response vector for the censored components. The predictions of the censored components, denoted by $\tilde{\mathbf{y}}_i^{c(k)}$, are calculated as

$$\tilde{\mathbf{y}}_i^{c(k)} = E\{\mathbf{y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\}, \quad i = 1, \ldots, n,$$

where

$$\tilde{\mathbf{y}}_i^{c(k)} = \tilde{\mathbf{y}}_i^{c(k-1)} + \delta_k \left[\frac{1}{m} \sum_{l=1}^{m} \mathbf{y}_i^{c(k,l)} - \tilde{\mathbf{y}}_i^{c(k)} \right]. \tag{15}$$

- The $\mathbf{y}_i^{c(k,l)}$'s are obtained without computational effort from the **Step 1** of the proposed SAEM algorithm.

- Following Wang (2013) and Garay *et al.* (2015), **the best linear predictor** of $\mathbf{y}_{i,pred}$ (with respect to the minimum mean squared error) is the conditional expectation of $\mathbf{y}_{i,pred}$ given $\mathbf{y}_{i,obs^*}$, namely

$$\widehat{\mathbf{y}}_{i,pred}(\boldsymbol{\theta}) = \mathbf{X}_{i,pred}\boldsymbol{\beta} + \boldsymbol{\Omega}_i^{pred,obs^*}\boldsymbol{\Omega}_i^{obs^*,obs^*-1}\left(\mathbf{y}_{i,obs^*} - \mathbf{X}_{i,obs^*}\boldsymbol{\beta}\right), \quad (16)$$

where, $\bar{\mathbf{X}}_i = (\mathbf{X}_{i,obs}, \mathbf{X}_{i,pred})$ be the $(n_{i,obs} + n_{i,pred}) \times p$ design matrix corresponding to $\bar{\mathbf{y}}_i = (\mathbf{y}_{i,obs}^\top, \mathbf{y}_{i,pred}^\top)$,

$$\bar{\mathbf{y}}_i^* = \left(\mathbf{y}_{i,obs^*}^\top, \mathbf{y}_{i,pred}^\top\right)^\top \sim SMN_{n_{i,obs}+n_{i,pred}}\left(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Omega}_i; \mathrm{H}\right),$$

with $\boldsymbol{\Omega}_i = \begin{pmatrix} \boldsymbol{\Omega}_i^{obs^*,obs^*} & \boldsymbol{\Omega}_i^{obs^*,pred} \\ \boldsymbol{\Omega}_i^{pred,obs^*} & \boldsymbol{\Omega}_i^{pred,pred} \end{pmatrix}$.

Summary

Example - UTI data

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, 72$$

- y_{ij} is the $\log_{10}$ HIV RNA for subject i at time t_j,
- 362 observations,
- 7% of the observation were below the detection limits,
- $t_1 = 0$, $t_2 = 1$, $t_3 = 3$, $t_4 = 6$, $t_5 = 9$, $t_6 = 12$, $t_7 = 18$, $t_8 = 24$ months before the interruption.

- This data set was analyzed previously by Vaida *et al.* (2007), Vaida & Liu (2009), Matos *et al.* (2013b), Matos *et al.* (2013a) e and Garay *et al.* (2015).

Example - UTI data

Information criteria for the SMN-CR models under DEC structure:

	Criteria	DEC	AR(1)	MA(1)	SYM	UNC
T	ℓ_{max}	-363.08	-406.98	-468.31	-364.21	-473.92
	AIC	748.15	833.96	956.62	748.43	965.84
	BIC	790.96	872.87	995.53	787.34	1000.86
	ν	2.3	2.1	2.1	2.3	2.1
SL	ℓ_{max}	-359.72	-403.08	-470.46	-360.90	-476.12
	AIC	741.44	826.15	960.92	741.79	970.24
	BIC	784.25	865.07	999.84	780.71	1005.26
	ν	0.8	0.7	1.0	0.8	1.0
CN	ℓ_{max}	**-351.32**	-396.56	-481.87	-353.37	-487.92
	AIC	**724.64**	813.12	983.74	726.75	993.83
	BIC	**767.44**	852.04	1022.66	765.66	1028.86
	ν	**(0.2,0.1)**	(0.3,0.1)	(0.1,0.1)	(0.2,0.1)	(0.1,0.1)
N	ℓ_{max}	-411.93	-463.05	-516.52	-412.06	-524.17
	AIC	845.87	946.11	1053.03	844.11	1066.34
	BIC	888.68	985.02	1091.95	883.03	1101.37
	ν	-	-	-	-	-

Example - UTI data

ML estimates with standard errors for the SMN-CR models under DEC structure:

	T	SL	CN	N
Parameter	Estimative (SE)	Estimative (SE)	Estimative (SE)	Estimative (SE)
β_1	4.040 (0.096)	4.020 (0.096)	**3.993 (0.097)**	3.625 (0.136)
β_2	4.321 (0.107)	4.312 (0.107)	**4.303 (0.111)**	4.185 (0.178)
β_3	4.354 (0.111)	4.344 (0.115)	**4.332 (0.119)**	4.259 (0.212)
β_4	4.533 (0.115)	4.498 (0.117)	**4.487 (0.119)**	4.375 (0.201)
β_5	4.675 (0.130)	4.649 (0.129)	**4.638 (0.122)**	4.579 (0.223)
β_6	4.670 (0.147)	4.646 (0.141)	**4.623 (0.139)**	4.582 (0.243)
β_7	4.688 (0.136)	4.670 (0.140)	**4.657 (0.152)**	4.688 (0.218)
β_8	4.871 (0.183)	4.842 (0.189)	**4.791 (0.206)**	4.806 (0.378)
σ^2	0.544 (0.139)	0.282 (0.065)	**0.543 (0.100)**	1.090 (0.134)
ϕ_1	0.812 (0.040)	0.820 (0.038)	**0.823 (0.038)**	0.700 (0.043)
ϕ_2	0.094 (0.083)	0.096 (0.082)	**0.121 (0.085)**	0.028 (0.071)

Example - UTI data

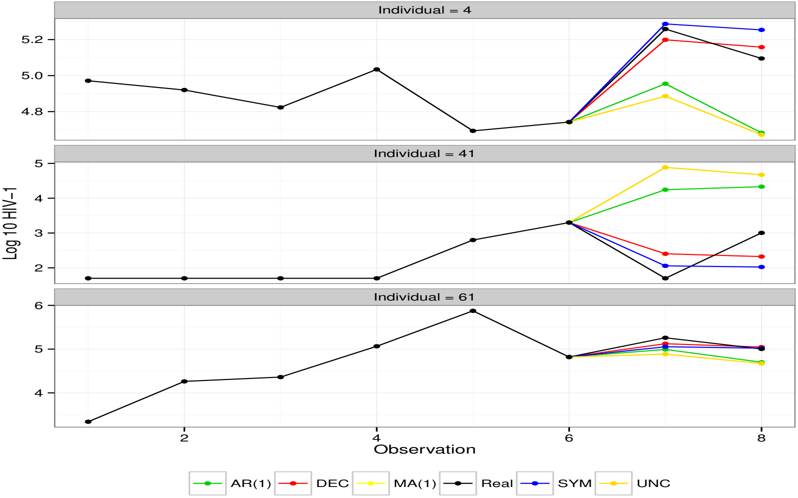
Prediction accuracy for the SMN-CR models under DEC correlation structure:

	T	SL	CN	N
MSE	0.219	0.227	**0.197**	0.240
MAE	0.357	0.361	**0.340**	0.383

$$\mathrm{MAE} = \frac{1}{58}\sum_{i,j}|y_{ij} - y^*_{ij}| \quad \text{and} \quad \mathrm{MSE} = \frac{1}{58}\sum_{i,j}(y_{ij} - y^*_{ij})^2,$$

Example - UTI data

Prediction performance for three random subjects.



Summary

Conclusions

- We have proposed a robust approach to linear regression models with censored observations based on the class of multivariate SMN distribution, called the SMN-CR model.

- To model the autocorrelation among irregularly observed measures, a damped exponential correlation structure was adopted, as proposed by Muñoz et al. (1992).

- A novel SAEM algorithm to obtain the ML estimates is developed by exploring statistical properties of the SMN class of distribution.

- We applied our methods to a recent AIDS study (freely downloadable from `R`), concluding that when the antiretroviral therapy is interrupted, the HIV-1 RNA levels in blood increase consistently along the period of evaluation.

Main reference

L.A Matos, Victor H. Lachos, Luis M. Castro and T-I Lin. (2015) Heavy-tailed longitudinal regression models for censored data: A likelihood based perspective. *Submitted*

References

Andrews, D. F. & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 99–102.

Delyon, B., Lavielle, M. & Moulines, E. (1999). Convergence of a stochastic approximation version of the em algorithm. *Annals of Statistics*, pages 94–128.

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B,*, **39**, 1–38.

Garay, A. M., Castro, L. M., Leskow, J. & Lachos, V. H. (2015). Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. *Statistical methods in medical research*, page 0962280214551191.

Lachos, V. H., Bandyopadhyay, D. & Dey, D. K. (2011). Linear and nonlinear mixed-effects models for censored hiv viral loads using normal/independent distributions. *Biometrics*, **67**, 1594–1604.

Lange, K. L. & Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *J. Comput. Graph. Stat*, **2**, 175–198.

Matos, L., Prates, M., Chen, M.-H. & Lachos, V. (2013a). Likelihood based inference for linear and nonlinear mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica*, **23**, 1323–1345.

Matos, L. A., Lachos, V. H., Balakrishnan, N. & Labra, F. V. (2013b). Influence diagnostics in linear and nonlinear mixed-effects models with censored data. *Computational Statistics & Data Analysis*, **57**(1), 450–464.

Muñoz, A., Carey, V., Schouten, J. P., Segal, M. & Rosner, B. (1992). A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics*, **48**, 733–742.

Samson, A., Lavielle, M. & Mentré, F. (2006). Extension of the saem algorithm to left-censored data in nonlinear mixed-effects model: application to hiv dynamics model. *Computational statistics & data analysis*, **51**(3), 1562–1574.

Vaida, F. & Liu, L. (2009). Fast Implementation for Normal Mixed Effects Models With Censored Response. *Journal of Computational and Graphical Statistics*, **18**(4), 797–817.

Vaida, F., Fitzgerald, A. & DeGruttola, V. (2007). Efficient hybrid EM for linear and nonlinear mixed effects models with censored response. *Computational Statistics & Data Analysis*, **51**(12), 5718–5730.

Wang, W.-L. (2013). Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. *Biometrical Journal*, **55**(4), 554–571.

Wei, G. C. & Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, **85**(411), 699–704.

Acknowledgments

Thank you!