Multivariate Regression Models With General Parameterization

Alexandre G. Patriota

IME-USP

Joint work with Artur Lemonte, Mário de Castro, Silvia Ferrari and Tatiane Melo

Table of contents

Some regression models

(Non)Linear regression model

Linear regression model:

$$Y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + e_i, \quad i = 1, \ldots, n$$

Non-linear regression model:

$$Y_i = f(\boldsymbol{x}_i, \boldsymbol{\beta}) + e_i, \quad i = 1, \ldots, n$$

Assumptions are typically made on $\boldsymbol{x}_i$, f and e_i to guarantee some properties of estimators and statistics.

Mixed models with non-linear fixed effects

Mixed model with linear fixed effects:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i^\top \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{e}_i, \quad i = 1, \ldots, n$$

Mixed model with non-linear fixed effects:

$$\boldsymbol{Y}_i = \boldsymbol{f}(\boldsymbol{X}_i, \boldsymbol{\beta}) + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{e}_i, \quad i = 1, \ldots, n$$

In addition, assumptions on the joint distribution of $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ (random effect and error model) are typically made.

Errors-in-variables models

A simple measurement error model:

$$z_i \quad = \quad \beta_0 + \beta_1 w_i + q_i, \quad i = 1, \ldots, n,$$

$$\begin{cases} Z_i & = & z_i + e_i, \\ W_i & = & w_i + u_i. \end{cases}$$

where

- z_i and w_i are **non-observable** response and explanatory variables,
- Z_i and W_i are the **surrogate observable** variables for z_i and w_i.

Assumptions on the joint distribution of q_i, e_i, u_i, w_i are typically made.

Distribution for the random terms

Distribution for the random terms

In general, the random terms are assumed to be **symmetric around zero**.

- Normal errors are symmetric around the mean=median. However, their kurtosis (Karl Pearson) is equal to 3.
- Other distributions are also symmetric around the median and have more flexible kurtosis.

Here, we consider the class of the **elliptical distributions**, which has the normal distribution as a particular instance.

Elliptical distributions

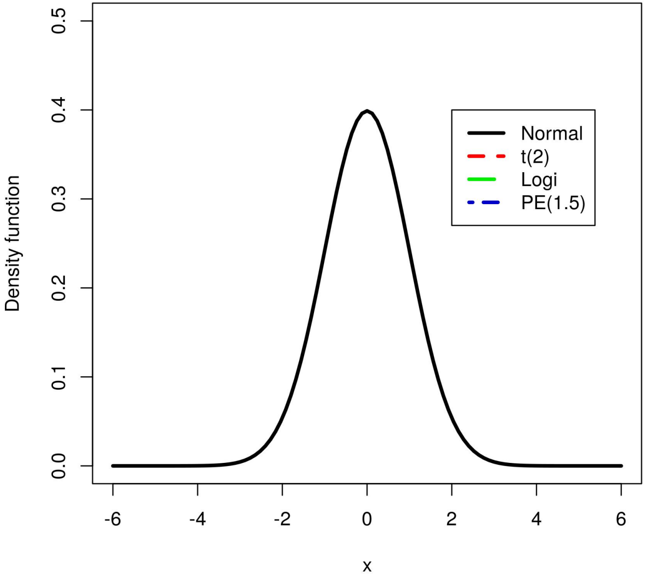
Definition: The random vector $\boldsymbol{Y}$ has an elliptical distribution if its density function exists and it is given by

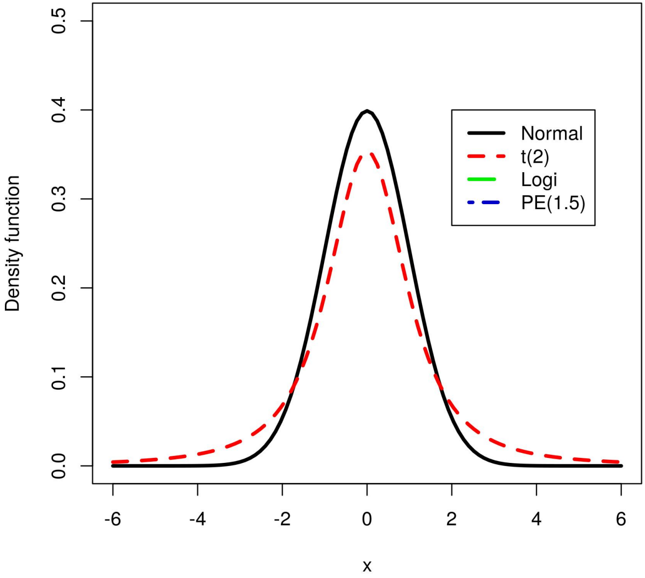
$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = |\boldsymbol{\Sigma}|^{-1/2} g\big[(\boldsymbol{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\big], \quad \boldsymbol{y} \in \mathbb{R}^d,$$

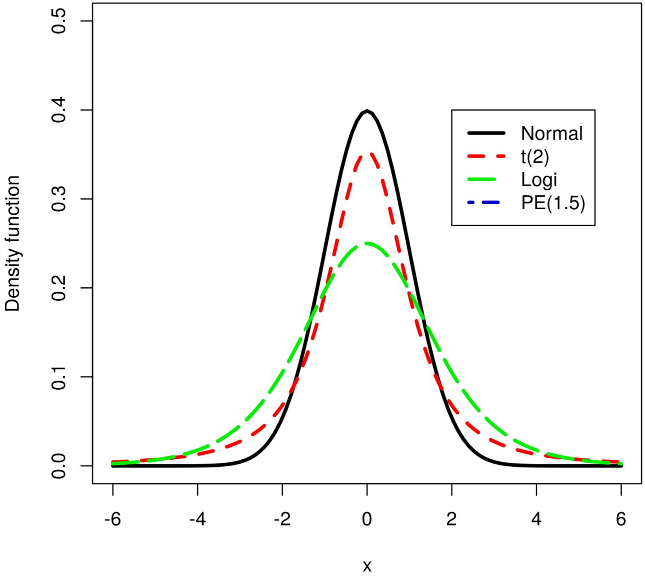
where $g : [0,\infty) \to [0,\infty)$ is such that $\int_0^\infty u^{\frac{d}{2}-1} g(u) du < \infty$.

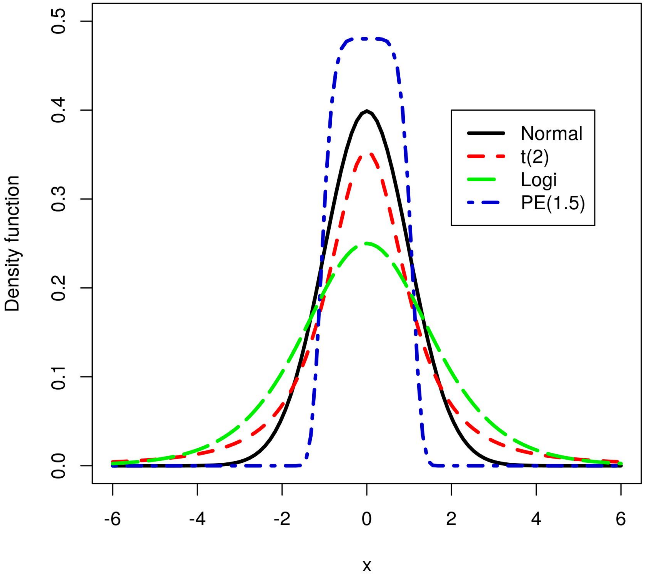
The function g is known as the generator density function. It is sufficiently smooth and does not contain extra unknown parameters.

Notation: $\boldsymbol{Y} \sim \mathcal{E}_d\big(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g\big)$, or, simply, $\boldsymbol{Y} \sim \mathcal{E}_d\big(\boldsymbol{\mu}, \boldsymbol{\Sigma}\big)$.


Normal
t(2)
Logi
PE(1.5)
Density function
x
0.0
0.1
0.2
0.3
0.4
0.5
-6
-4
-2
0
2
4
6


Normal
t(2)
Logi
PE(1.5)
Density function
0.0
0.1
0.2
0.3
0.4
0.5
-6
-4
-2
0
2
4
6
x


Normal
t(2)
Logi
PE(1.5)
Density function
0.0
0.1
0.2
0.3
0.4
0.5
-6
-4
-2
0
2
4
6
x



Density function
0.0
0.1
0.2
0.3
0.4
0.5
-6
-4
-2
0
2
4
6
x
Normal
t(2)
Logi
PE(1.5)

Elliptical distributions: useful property

Let $\boldsymbol{A}$ be a $(r \times d)$ matrix of rank r and $\boldsymbol{a}$ be a r-dimensional vector.

Theorem: If $\boldsymbol{Y} \sim \mathcal{E}_d\big(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g\big)$, then

$$\boldsymbol{W} = \boldsymbol{A}\boldsymbol{Y} + \boldsymbol{a} \sim \mathcal{E}_r\big(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{a}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top, g\big).$$

That is, the elliptical class is **closed under affine transformations**.

The multivariate regression model with general parametrization

The multivariate regression model with general parametrization

Let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ be random q_i-vectors for $i = 1, \ldots, n$, and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be known covariates. The regression model is defined by

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta}) + \boldsymbol{e}_i, \quad i = 1, \ldots, n,$$

where $\boldsymbol{e}_i \stackrel{ind}{\sim} \mathcal{E}_{q_i}\big(\boldsymbol{0}, \boldsymbol{\Sigma}_i(\boldsymbol{\theta})\big)$,

- $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ is the vector of parameters.
- $\boldsymbol{\mu}_i(\boldsymbol{\theta}) := \boldsymbol{\mu}_i(\boldsymbol{\theta}, \boldsymbol{x}_i)$ is a (smooth) vector-valued function,
- $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) := \boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \boldsymbol{x}_i)$ is a (smooth) positive definite matrix function.

The multivariate regression model with general parametrization

Then, the response random vectors can be written as

$$\boldsymbol{Y}_i \overset{ind}{\sim} \mathcal{E}_{q_i}\big(\boldsymbol{\mu}_i(\boldsymbol{\theta}), \boldsymbol{\Sigma}_i(\boldsymbol{\theta})\big), \quad i = 1, \ldots, n. \tag{1}$$

- This model extends the one proposed by (Patriota and Lemonte, 2009).
- This model unifies the previous models in one single model (under the assumption that the error terms are jointly elliptically distributed).

Remark: The previous assumption imposes for the error terms the same generator density function.

Linear Models

For the **homoscedastic linear** model:

$$Y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + e_i, \quad \text{with} \quad e_i \overset{iid}{\sim} \mathcal{E}_1\big(0, \sigma^2\big)$$

we have

$$q_i = 1,$$
$$\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top,$$
$$\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \boldsymbol{x}_i^\top \boldsymbol{\beta},$$
$$\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \sigma^2.$$

Non-Linear Models

For the **heteroscedastic non-linear linear** model:

$$Y_i = f(\boldsymbol{x}_{1i}, \boldsymbol{\beta}) + e_i$$

with $e_i \overset{ind}{\sim} \mathcal{E}_1\big(0, h(\boldsymbol{x}_{2i}, \boldsymbol{\gamma})\big)$ we have

$$q_i = 1,$$

$$\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top,$$

$$\boldsymbol{\mu}_i(\boldsymbol{\theta}) = f(\boldsymbol{x}_{1i}, \boldsymbol{\beta}),$$

$$\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = h(\boldsymbol{x}_{2i}, \boldsymbol{\gamma}).$$

Mixed models

For the mixed model with non-linear fixed effects:

$$\boldsymbol{Y}_i = \boldsymbol{f}(\boldsymbol{X}_i, \boldsymbol{\beta}) + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{e}_i$$

with

$$\begin{pmatrix} \boldsymbol{b}_i \\ \boldsymbol{e}_i \end{pmatrix} \overset{ind}{\sim} \mathcal{E}_{m+q_i}\left(\boldsymbol{0}, \begin{pmatrix} \boldsymbol{\Gamma}(\gamma) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}(\boldsymbol{\sigma}) \end{pmatrix}\right)$$

we have

$$\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \gamma^\top, \boldsymbol{\sigma}^\top)^\top,$$

$$\boldsymbol{x}_i = (\mathrm{vec}(\boldsymbol{X}_i)^\top, \mathrm{vec}(\boldsymbol{Z}_i)^\top)^\top,$$

$$\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \boldsymbol{f}(\boldsymbol{X}_i, \boldsymbol{\beta}),$$

$$\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \boldsymbol{Z}_i \boldsymbol{\Gamma}(\gamma) \boldsymbol{Z}_i^\top + \boldsymbol{\Lambda}(\boldsymbol{\sigma}).$$

Structural Errors-in-variables model

$$z_i = \beta_0 + \beta_1 w_i + q_i, \quad i = 1, \ldots, n,$$

$$\boldsymbol{Y}_i = \begin{pmatrix} Z_i \\ W_i \end{pmatrix} = \begin{pmatrix} z_i + e_i \\ w_i + u_i \end{pmatrix} \quad \begin{pmatrix} q_i \\ e_i \\ u_i \\ w_i - \mu_w \end{pmatrix} \overset{ind}{\sim} \mathcal{E}_4(\boldsymbol{0}, \boldsymbol{Q}_i),$$

$$\text{where} \quad \boldsymbol{Q}_i = \begin{pmatrix} \sigma_q^2 & 0 & 0 & 0 \\ 0 & \sigma_{e_i}^2 & 0 & 0 \\ 0 & 0 & \sigma_{u_i}^2 & 0 \\ 0 & 0 & 0 & \sigma_w^2 \end{pmatrix}.$$

we have

$$\boldsymbol{\theta} = (\beta_0, \beta_1, \mu_w, \sigma_w^2, \sigma_q^2)^\top,$$

$$\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \begin{pmatrix} \beta_0 + \beta_1 \mu_w \\ \mu_w \end{pmatrix} \text{ and } \boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \begin{pmatrix} \beta_1^2 \sigma_w^2 + \sigma_q^2 + \sigma_{e_i}^2 & \beta_1 \sigma_w^2 \\ \beta_1 \sigma_w^2 & \sigma_w^2 + \sigma_{u_i}^2 \end{pmatrix},$$

Maximum likelihood estimation

Maximum likelihood estimation

Let $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \boldsymbol{x}_i)$, $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \boldsymbol{w}_i)$, $\boldsymbol{z}_i = \boldsymbol{Y}_i - \boldsymbol{\mu}_i$ and $u_i = \boldsymbol{z}_i^\top \boldsymbol{\Sigma}_i^{-1} \boldsymbol{z}_i$.

The log-likelihood function, except for a constant term, is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \qquad (2)$$

where $\ell_i(\boldsymbol{\theta}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| + \log g(u_i)$.

The Score function and the Fisher information

The score function and the expected Fisher information are given by

$$\boldsymbol{U}_{\boldsymbol{\theta}} = \sum_{i=1}^{n} \boldsymbol{F}_i^\top \boldsymbol{H}_i \boldsymbol{s}_i \quad \text{and} \quad \boldsymbol{K}_{\boldsymbol{\theta}} = \sum_{i=1}^{n} \boldsymbol{F}_i^\top \boldsymbol{H}_i \boldsymbol{M}_i \boldsymbol{H}_i \boldsymbol{F}_i,$$

with

$$\boldsymbol{F}_i = \begin{pmatrix} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\theta}^\top} \\ \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_i)}{\partial \boldsymbol{\theta}^\top} \end{pmatrix}, \ \boldsymbol{H}_i = \begin{bmatrix} \boldsymbol{\Sigma}_i & \boldsymbol{0} \\ \boldsymbol{0} & 2\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i \end{bmatrix}^{-1}, \ \boldsymbol{s}_i = \begin{bmatrix} v_i \boldsymbol{z}_i \\ -\mathrm{vec}(\boldsymbol{\Sigma}_i - v_i \boldsymbol{z}_i \boldsymbol{z}_i^\top) \end{bmatrix},$$

$$\boldsymbol{M}_i = \begin{bmatrix} \frac{4d_{gi}}{q_i}\boldsymbol{\Sigma}_i & \boldsymbol{0} \\ \boldsymbol{0} & \frac{8f_{gi}}{q_i(q_i+2)}\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \left(\frac{4f_{gi}}{q_i(q_i+2)} - 1\right)\mathrm{vec}(\boldsymbol{\Sigma}_i)\mathrm{vec}(\boldsymbol{\Sigma}_i)^\top \end{bmatrix},$$

where v_i, d_{gi} and f_{gi} are quantities related with the elliptical distribution.

We assume also that

$$\boldsymbol{F} = (\boldsymbol{F}_1^\top, \ldots, \boldsymbol{F}_n^\top)$$

has rank p

and the functions $g(\cdot)$, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ must be defined in such way that $\ell(\boldsymbol{\theta})$ be a regular function with respect to $\boldsymbol{\theta}$ (Cox and Hinkley, 1974, Ch. 9).

The Fisher scoring method:

$$(\boldsymbol{F}^{(m)\top}\boldsymbol{W}^{(m)}\boldsymbol{F}^{(m)})\boldsymbol{\theta}^{(m+1)} = \boldsymbol{F}^{(m)\top}\boldsymbol{W}^{(m)}\boldsymbol{s}^{*(m)}, \ m = 0, 1, \ldots$$

where

$$\boldsymbol{W}^{(m)} = \boldsymbol{H}^{(m)}\boldsymbol{M}^{(m)}\boldsymbol{H}^{(m)}, \qquad \boldsymbol{F}^{(m)} = (\boldsymbol{F}_1^{(m)\top}, \boldsymbol{F}_2^{(m)\top}, \ldots, \boldsymbol{F}_n^{(m)\top})^\top,$$

$$\boldsymbol{H}^{(m)} = \text{block–diag}\{\boldsymbol{H}_1^{(m)}, \boldsymbol{H}_2^{(m)}, \ldots, \boldsymbol{H}_n^{(m)}\},$$

$$\boldsymbol{M}^{(m)} = \text{block–diag}\{\boldsymbol{M}_1^{(m)\top}, \boldsymbol{M}_2^{(m)\top}, \ldots, \boldsymbol{M}_n^{(m)\top}\},$$

$$\boldsymbol{s}^{*(m)} = \boldsymbol{F}^{(m)}\boldsymbol{\theta}^{(m)} + \boldsymbol{H}^{-1(m)}\boldsymbol{M}^{-1(m)}\boldsymbol{s}^{(m)}$$

$$\boldsymbol{s}^{(m)} = (\boldsymbol{s}_1^{(m)\top}, \boldsymbol{s}_2^{(m)\top}, \ldots, \boldsymbol{s}_n^{(m)\top})^\top$$

and m is the iteration counter.

The second-order bias of the MLEs

Bias Correction of the Maximum Likelihood Estimators

- Under regular conditions, the MLEs are **consistent** and asymptotically **normally distributed**.

- For finite samples and non-linear models, the MLEs can be strongly **biased** (producing misleading diagnostic analysis). Their biases are typically of order $\mathcal{O}(n^{-1})$.

- Cox and Snell (1968) provide the second-order biases of the MLEs
  - By considering **higher order terms** in the score function expansion.
  - The corrected MLEs are, in general, **lesser biased** than the non-corrected ones for small samples.

The second-order bias vector $B_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{\theta})$ under the general model is given by

$$B_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{\theta}) = (\boldsymbol{F}^\top \boldsymbol{W} \boldsymbol{F})^{-1} \boldsymbol{F}^\top \boldsymbol{W} \boldsymbol{\xi}, \qquad (3)$$

where $\boldsymbol{\xi}$ is a given in Melo et al. (2017a).

The bias-corrected estimator is defined as

$$\tilde{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}} - \widehat{B_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{\theta})}$$

Remark: This result extends the one attained by Patriota and Lemonte (2009) under normally distributed errors.

Simulations

Consider $q_i = 1$, $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^\top$, $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \sigma^2$ and

$$\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \alpha_1 + \frac{\alpha_2}{1 + \alpha_3 x_i^{\alpha_4}}, \quad i = 1, \ldots, n. \tag{4}$$

- The values of x_i were obtained as random draws from the uniform distribution $U(0, 100)$.
- The sample sizes considered are $n = 10, 20, 30, 40$ and 50.
- The parameter values are $\alpha_1 = 50$, $\alpha_2 = 500$, $\alpha_3 = 0.50$, $\alpha_4 = 2$ and $\sigma_i^2 = 200$.
- Distributions: normal and Student t ($\nu = 4$).

We compute the ML estimator $\widehat{\theta}$ and its bias-corrected version

$$\widetilde{\theta} = \widehat{\theta} - \widehat{B_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{\theta})}$$

Results: Normal distribution

n	$\boldsymbol{\theta}$	MLE Bias	MLE $\sqrt{MSE}$	Bias-corrected MLE Bias	Bias-corrected MLE $\sqrt{MSE}$
10	α_1	-0.29	6.69	-0.13	6.67
	α_2	2.16	20.07	0.70	19.40
	α_3	0.01	0.13	0.00	0.12
	α_4	0.03	0.30	0.01	0.29
	σ^2	-80.05	106.44	-32.06	103.32
20	α_1	-0.08	4.07	-0.01	4.07
	α_2	0.66	17.94	-0.08	17.84
	α_3	0.00	0.09	0.00	0.09
	α_4	0.02	0.21	0.01	0.20
	σ^2	-40.07	69.73	-8.09	68.95
30	α_1	-0.10	3.11	-0.04	3.10
	α_2	0.71	17.24	-0.05	17.15
	α_3	0.00	0.09	-0.00	0.09
	α_4	0.02	0.20	0.00	0.19
	σ^2	-26.41	55.26	-3.26	55.11
40	α_1	-0.08	2.69	-0.02	2.69
	α_2	0.83	16.80	0.09	16.70
	α_3	0.00	0.09	0.00	0.09
	α_4	0.02	0.19	0.00	0.18
	σ^2	-20.04	47.26	-2.04	47.13
50	α_1	-0.08	2.39	-0.03	2.38
	α_2	1.07	14.25	0.30	14.12
	α_3	0.00	0.08	0.00	0.08
	α_4	0.01	0.19	0.00	0.18
	σ^2	-15.93	41.41	-1.21	41.30

	n = 10	20	30	40	50
$\frac{B(\widehat{\sigma}^2)}{B(\widetilde{\sigma}^2)} =$	2.5	5	8.1	9.8	13.2

Results: Student-t distribution $\nu = 4$

n	$\boldsymbol{\theta}$	MLE Bias	MLE $\sqrt{MSE}$	Bias-corrected MLE Bias	Bias-corrected MLE $\sqrt{MSE}$
10	α_1	-0.51	8.66	-0.31	8.63
	α_2	3.34	28.47	1.39	27.34
	α_3	0.01	0.17	0.00	0.16
	α_4	0.06	0.42	0.03	0.39
	σ^2	-93.18	127.60	-54.24	130.73
20	α_1	-0.17	5.03	-0.07	5.02
	α_2	2.01	25.64	0.91	25.11
	α_3	0.01	0.14	0.01	0.14
	α_4	0.04	0.29	0.01	0.28
	σ^2	-41.24	85.51	-12.30	89.41
30	α_1	-0.10	3.81	-0.01	3.80
	α_2	2.25	25.75	1.13	25.34
	α_3	0.01	0.14	0.01	0.14
	α_4	0.04	0.29	0.01	0.27
	σ^2	-27.15	70.02	-6.15	72.64
40	α_1	-0.10	3.27	-0.02	3.26
	α_2	1.82	24.94	0.75	24.67
	α_3	0.01	0.12	0.00	0.12
	α_4	0.03	0.26	0.01	0.25
	σ^2	-20.38	60.43	-4.01	62.21
50	α_1	-0.13	2.86	-0.05	2.85
	α_2	1.48	18.86	0.38	18.59
	α_3	0.01	0.11	0.00	0.11
	α_4	0.02	0.24	0.00	0.23
	σ^2	-15.40	53.99	-1.94	55.56

	n				
	10	20	30	40	50
$\frac{B(\widehat{\sigma}^2)}{B(\widetilde{\sigma}^2)} =$	1.7	3.4	4.4	5.1	7.9

Skovgaard adjustment for LR statistic

The likelihood-ratio statistic

Consider the null and alternative hypotheses are

$$H_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0 \quad \text{and} \quad H_1 : \boldsymbol{\psi} \neq \boldsymbol{\psi}_0,$$

where $\boldsymbol{\psi}_0$ is known and $\boldsymbol{\theta} = (\boldsymbol{\psi}^\top, \boldsymbol{\omega}^\top) \in \Theta \subseteq \mathbb{R}^p$.

Under regular conditions (Severine, 2000), the $(-2\log\text{-})$ Likelihood-ratio statistic

$$LR_n = 2\Big(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widetilde{\boldsymbol{\theta}}_0)\Big) \xrightarrow{D} \chi^2_q, \quad \text{under } H_0$$

where $\widehat{\boldsymbol{\theta}}$ is the MLE and $\widetilde{\boldsymbol{\theta}}_0$ is the restricted MLE under the null hypothesis.

Remark: For small samples, this approximation may not be good.

Skovgaard adjustment for the likelihood-ratio statistic

Skovgaard (2001)'s adjustment for the likelihood-ratio statistic

$$LR_n^{**} = LR_n - 2\log\rho_n \qquad (LR^{**} \xrightarrow{D} \chi^2_q), \quad \text{under } H_0$$

where ρ_n depends on $\widehat{\boldsymbol{\theta}}$, $\widetilde{\boldsymbol{\theta}}_0$, some derivatives of $\ell(\cdot)$ and an ancillary statistic $\boldsymbol{a}$ such that $(\widehat{\boldsymbol{\theta}}, \boldsymbol{a})$ be a sufficient statistic.

Problem: It is not easy to find an ancillary statistic $\boldsymbol{a}$. All the other quantities are achievable through integration and differentiation.

Solution: Melo et al. (2017b) used an approximate ancillary statistic for the general model, namely, $\boldsymbol{a} = (\boldsymbol{a}_1^\top, \ldots, \boldsymbol{a}_n^\top)^\top$, where

$$\boldsymbol{a}_i = \widehat{\boldsymbol{P}}_i(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_i),$$

where $\boldsymbol{P}_i$ is a lower triangular matrix such that $\boldsymbol{P}_i\boldsymbol{P}_i^\top = \boldsymbol{\Sigma}_i$.

Consider the mixed model:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \quad \text{where } (\boldsymbol{e}_i,\ \boldsymbol{b}_i) \overset{ind}{\sim} \mathcal{E}_{q_i+2}(\boldsymbol{0}, \boldsymbol{S}_i),$$

$$\boldsymbol{S}_i = \begin{pmatrix} \sigma^2 \boldsymbol{I}_{q_i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}(\boldsymbol{\gamma}) \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Lambda}(\boldsymbol{\gamma}) = \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & \gamma_3 \end{pmatrix},$$

with $q_i \in \{1, ..., 5\}$ chosen randomly and $n = 16$.

- Normal, Student-t ($\nu = 3$) and Power Exponential ($\lambda = 7$) distributions were considered.
- $\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{1} & \boldsymbol{x}_{i1} & \boldsymbol{x}_{i2} & \boldsymbol{x}_{i3} & \boldsymbol{x}_{i4} \end{pmatrix}$ and $\boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{1} & \boldsymbol{x}_{i1} \end{pmatrix}$, where $\boldsymbol{x}_{i1}$ is the first q_i components of $\{5, 10, 15, 30, 60\}$, $\boldsymbol{x}_{ij}$ are dummies variables, $j = 2, 3, 4$.
- $\beta_0 = 0.7$, $\beta_1 = 0.5$, $\underbrace{\beta_2 = \beta_3 = \beta_4 = 0}_{H_0 : \psi = 0}$, $\gamma_1 = 500$, $\gamma_2 = 2$, $\gamma_3 = 200$, $\sigma^2 = 5$.

Relative p-value discrepancy

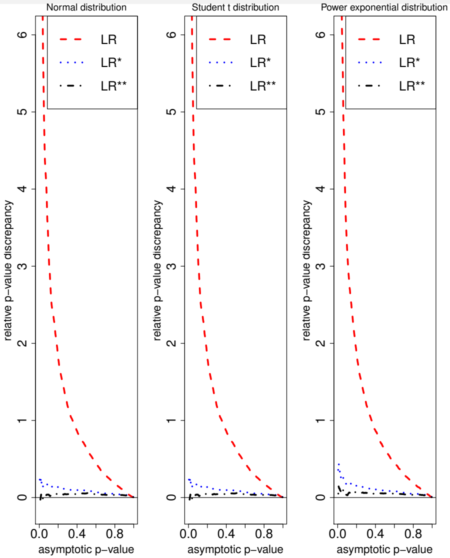
The relative p-value discrepancy is defined by the difference between the **exact** and the **asymptotic** p-values divided by the asymptotic p-value.

- The *exact* p-values are based on the LR statistics and their distributions obtained through Monte Carlo's simulations.

- The asymptotic p-values are based on the LR statistics and their asymptotic distributions.

$$\text{“relative } p\text{-value discrepancy”} = \frac{\text{“exact } p\text{-value”} - \text{“asymptotic } p\text{-value”}}{\text{“asymptotic } p\text{-value”}}$$

P-value discrepancies



Normal distribution

- Bias correction for the MLEs and influence diagnostics were developed in 2009 and 2010, respectively.

1. Patriota, AG, Lemonte, AJ. (2009). Bias correction in a multivariate normal regression model with general parameterization, *Statistics & Probability Letters*, **79**, 1655–1662.
2. Patriota, AG, Lemonte, AJ, de Castro, M. (2010b). Influence diagnostics in a multivariate normal regression model with general parameterization, *Statistical Methodology*, **7**, 644–654.

Elliptical distributions

- Influence diagnostics, bias correction for the MLEs and Skovgaard adjustments for the LR statistic were developed in 2011, 2017 and 2017, respectively

1. Lemonte, AJ, Patriota, AG. (2011). Multivariate elliptical models with general parameterization, *Statistical Methodology*, **8**, 389–400.
2. Melo, TFN, Ferrari, SLP, Patriota, AG. (2017a). Improved estimation in a general multivariate elliptical model, *Brazilian Journal of Probability and Statistics*. In Press.
3. Melo, TFN, Ferrari, SLP, Patriota, AG. (2017b). Improved hypothesis testing in a general multivariate elliptical model, *Journal of Statistical Computation and Simulation*, **87**, 1416–1428.

References

References:

Barndorff-Nielsen, O.E. (1986). Inference on full or partial parameters, based on the standardized signed log likelihood ratio. *Biometrika*, 73, 307–322.

Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society B*. **48**:133–169.

Cox, D.R., Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Cox, D.R., Snell, E. (1968). A general definition of residuals (with discussion). *Journal of the Royal Statistical Society B* 30:248–275.

Crepeau, H., Koziol, J., Reid, N. and Yuh, Y.S. (1985). Analysis of incomplete multivariate data from repeated measurement experiments. *Biometrics*, 41, 505–514.

Lucas, A. (1997). Robustness of the student t based M-estimator. *Communications in Statistics, Theory and Methods*. **26**:1165–1182.

References:

Severini TA. Likelihood methods in statistics. New York: Oxford University Press, 2000.

Skovgaard, I.M. (2001). Likelihood asymptotics. *Scandinavian Journal of Statistics*, 28, 3–32.

Tiede, J.J. and Pagano, M. (1979). Application of robust calibration to radioimmunoassay. *Biometrics*, **35** 567–574.

Thank you