Objective Bayesian Analysis for Heteroscedastic Regression

Helio Migon & Esther Salazar

Universidade Federal do Rio de Janeiro

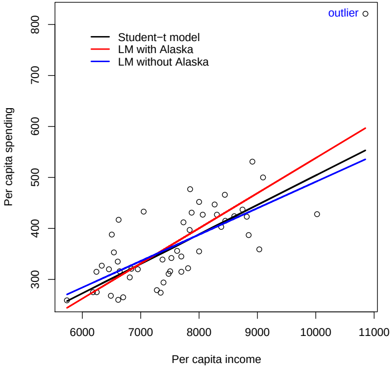
Colóquio Inter-institucional: Modelos Estocásticos e Aplicações
2009

Collaborators: Marco Ferreira and Thais Fonseca

Summary

Introduction: Example 1

Data: Per capita income and per capita spending in public schools by state in the United States in 1979.

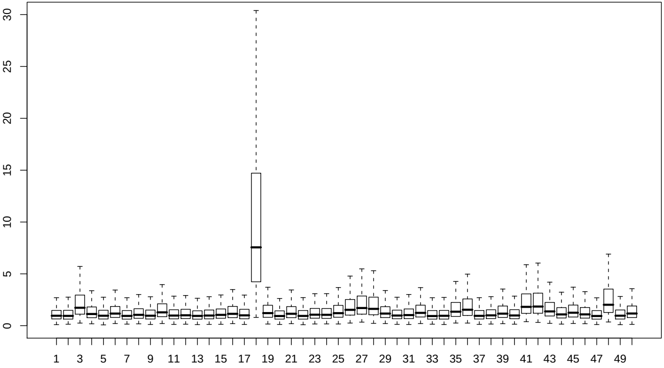


Linear and quadratic fits under Gaussian errors, dashed, and Student-t errors, solid.

Introduction: Example 1

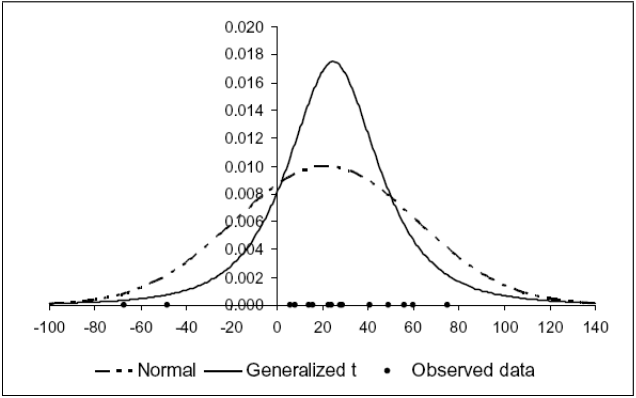
Model: Student-t **linear regression model**

λ_i^{-1}: Large values are associated with possible outliers



Introduction: Example 2

Darwin's data: Differences in heights of 15 pairs of self- and cross-fertilized plants.

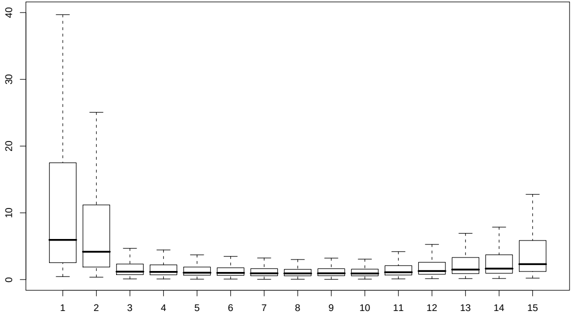


Dot plot of Darwin's data and the sampling distributions of the data.

Introduction: Example 2

Model: Student-t distribution

λ_i^{-1} Large values are associated with possible outliers



Boxplots of λ_i^{-1}, $i = 1, \ldots, 15$

Location-scale model

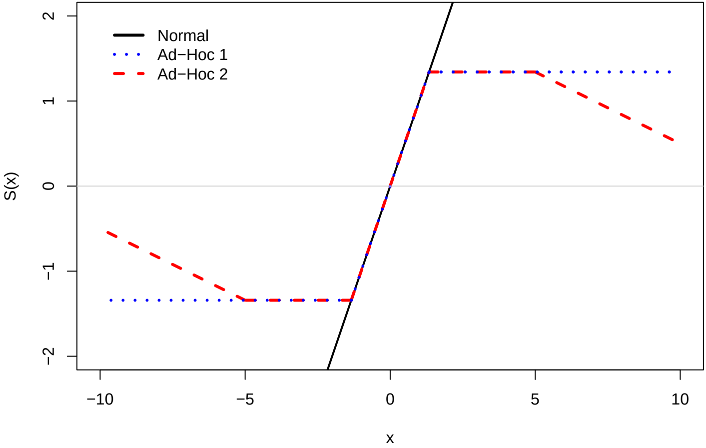
Location-scale model

Models based on the normal distribution are not robust to outliers

Alternative? Location-scale models with heavy-tailed prior distributions

- Student t-distribution
- exponential power distribution
- between others

Influence function: Ad-Hoc



Influence function

Helio Migon
UFRJ

$$s(x) = -\frac{\partial}{\partial x} \log p(x)$$

where $x = y - \mu$

Examples:

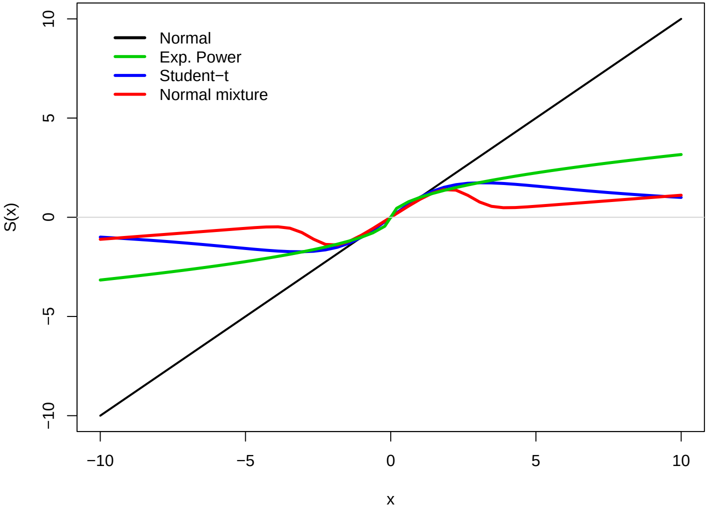
- Normal $(0, 1)$: $s(x) = x$
- Exponential Power $(0, 1, p)$:

$$s(x) = -(-x)^{p-1} 1_{(x \leq 0)} + x^{p-1} 1_{(x > 0)}$$

- Student-t $(0, 1, \nu)$: $s(x) = \frac{(\nu+1)x}{\nu+x^2}$, ν: d.f.
- Mixture of normals $\pi N(0, 1) + (1 - \pi) N(0, \sigma^2)$:

$$s(x) = \frac{\pi x \exp(-x^2/2) + (1-\pi) x \sigma^{-3} \exp(-x^2/(2\sigma^2))}{\pi \exp(-x^2/2) + (1-\pi)\sigma^{-1} \exp(-x^2/(2\sigma^2))}$$

Influence function



Normal scale mixture distributions

Let X be a continuous random variable with location μ and scale σ. The pdf of X has the scale mixture of normal representation if

$$f_X(x|\mu,\sigma) = \int_0^\infty N(x|\mu,\kappa(\lambda)\sigma^2)\pi(\lambda)d\lambda$$

where $\kappa(\lambda)$ is a positive function and $\pi(\cdot)$ is a density function on $\Re^+$.

Normal scale mixture distributions

Let X be a continuous random variable with location μ and scale σ. The pdf of X has the scale mixture of normal representation if

$$f_X(x|\mu,\sigma) = \int_0^\infty N(x|\mu,\kappa(\lambda)\sigma^2)\pi(\lambda)d\lambda$$

where $\kappa(\lambda)$ is a positive function and $\pi(\cdot)$ is a density function on $\Re^+$.

Example: The Student t-distribution

$$t_\nu(x|\mu,\sigma) = \int_0^\infty N(x|\mu,\frac{\sigma^2}{\lambda})Ga(\lambda|\frac{\nu}{2},\frac{\nu}{2})d\lambda$$

that is, $X \sim t_\nu(\mu,\sigma)$ follows the hierarchical form:

$$X|\mu,\sigma,\nu,\lambda \sim N(\mu,\frac{\sigma^2}{\lambda}) \quad \text{and} \quad \lambda|\nu \sim Ga(\frac{\nu}{2},\frac{\nu}{2})$$

Uniform scale mixture distributions

Let X be a continuous random variable with location μ and scale σ. The pdf of X has the scale mixture of uniform representation if

$$f_X(x|\mu,\sigma) = \int_0^\infty U(x|\mu-\kappa(\lambda)\sigma^2, \mu+\kappa(\lambda)\sigma^2)\pi(\lambda)d\lambda$$

where $\kappa(\cdot)$ is a positive function and $\pi(\cdot)$ is a density function on $\Re^+$

Uniform scale mixture distributions

Let X be a continuous random variable with location μ and scale σ. The pdf of X has the scale mixture of uniform representation if

$$f_X(x|\mu,\sigma) = \int_0^\infty U(x|\mu - \kappa(\lambda)\sigma^2, \mu + \kappa(\lambda)\sigma^2)\pi(\lambda)d\lambda$$

where $\kappa(\cdot)$ is a positive function and $\pi(\cdot)$ is a density function on $\Re^+$

Example: The exponential power (EP) distribution

$$f_X(x|\mu,\sigma,\beta) = \frac{c_1}{\sigma}\exp\left(-\left|\frac{c_0^{1/2}(x-\mu)}{\sigma}\right|^{2/\beta}\right)$$

where $\beta \in (0,2]$ controls the kurtosis. The EP distribution follows the hierarchical form:

$$\begin{aligned} X|\mu,\sigma,\beta,\lambda &\sim U\left(\mu - \frac{\sigma}{\sqrt{2c_0}}\lambda^{\beta/2}, \mu + \frac{\sigma}{\sqrt{2c_0}}\lambda^{\beta/2}\right) \\ \lambda|\beta &\sim Ga(1+\frac{\beta}{2}, 2^{-1/\beta}) \end{aligned}$$

Data analysis

Application: School spending

Student-t linear model

- Dependent variable (y_i): per capita spending
- Regressor variables (x_i): per capita income and its square

Student-t model

$$y_i|\theta,\sigma,\nu,\lambda_i \sim N(x_i'\theta, \frac{\sigma^2}{\lambda_i})$$
$$\lambda_i|\nu \sim Ga(\frac{\nu}{2}, \frac{\nu}{2})$$

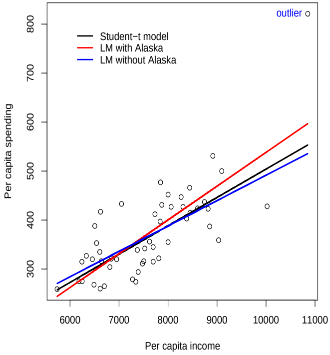
Exponential power model

$$y_i|\theta,\sigma,\beta,\lambda_i \sim U\left(x_i'\theta - \frac{\sigma}{\sqrt{2c_0}}\lambda_i^{\beta/2}, x_i'\theta + \frac{\sigma}{\sqrt{2c_0}}\lambda_i^{\beta/2}\right)$$
$$\lambda_i|\beta \sim Ga(1+\frac{\beta}{2}, 2^{-1/\beta})$$

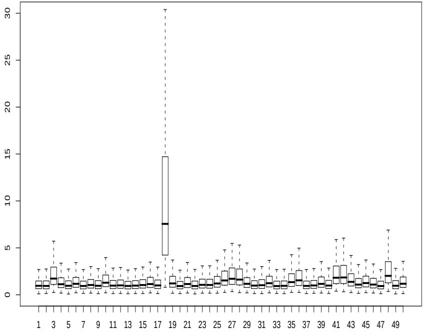
Application: School spending (Student-t model)

Linear model

Linear fit



Boxplots of λ_i^{-1}, $i = 1, \ldots, 50$



Application: School spending (Student-t model)

Linear model

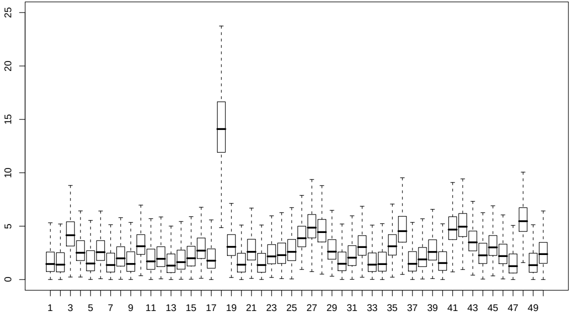
Posterior summaries based on the independent Jeffreys prior (acceptance rate of ν: 0.42) and coefficients of linear model with and without Alaska.

Parameter	Median	95% C.I.	LM with Alaska	LM without Alaska
θ_1	-74.26	(-205.89, 54.57)	-151.27	-26.80
θ_2	578.25	(407.44, 752.98)	689.39	518.31
σ	46.68	(33.44, 62.88)	61.41	49.90
ν	4.36	(1.81, 16.67)	-	

ν: degrees of freedom

Linear model

Boxplots of λ_i, $i = 1, \ldots, 50$



Posterior summaries:

Parameter	Median	95% C.I.
θ_1	-103.0	(-258.0, 13.5)
θ_2	618.0	(464.0, 830.0)
σ	51.9	(38.9, 71.4)
β	1.32	(1.03, 2.19)

Application: School spending (quadratic model)

Student-t model vs EP model

Student-t model:

Parameter	Median	95% C.I.
θ_1	891.58	(-80.69, 1591.13)
θ_2	-2051.82	(-3842.39, 612.12)
θ_3	1771.90	(-38.06, 2915.73)
σ	46.90	(32.12, 63.54)
ν	5.25	(1.83, 43.14)

ν: degrees of freedom

Exponential power model:

Parameter	Median	95% C.I.
θ_1	1188.83	(923.71, 1561.89)
θ_2	-2796.18	(-3760.45, -2140.38)
θ_3	2225.43	(1801.55, 2863.88)
σ	49.87	(36.88, 73.33)
β	1.35	(1.03, 3.46)

Objective Bayesian analysis

Student-t regression model: objective Bayesian analysis

Fonseca, Ferreira and Migon (2008)

Consider the linear regression model

$$y = X\beta + \epsilon$$

- $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ is the error vector
- ϵ_i's are i.i.d according to the Student-t distribution $(0, \sigma, \nu)$
- $X = (x_1, \ldots, x_n)'$ is the $n \times p$ matrix of explanatory variables and of full-rank p
- Model parameters: $\theta = (\beta, \sigma, \nu) \in \Re^p \times (0, \infty)^2$

Jeffreys priors: Student-t regression model

Class of improper prior distributions

$$\pi(\Theta) \propto \frac{\pi(p)}{\sigma^a}$$

The independence Jeffreys prior and the Jeffreys-rule prior for $\Theta = (\beta, \sigma, \nu)$, denoted by $\pi^I(\Theta)$ and $\pi^R(\Theta)$ are given by

$$\pi^I(\Theta): \; a = 1,$$

$$\pi^I(\nu) \propto \left(\frac{\nu}{\nu+3}\right)^{1/2} \left\{ \Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right\}^{1/2}$$

$$\pi^R(\Theta): \; a = p + 1,$$

$$\pi^R(\nu) \propto \pi^I(\nu) \left(\frac{\nu+1}{\nu+3}\right)^{p/2}$$

Jeffreys priors: Student-t regression model

Corollary

The marginal independence Jeffreys prior for ν given by $\pi^I(\theta)$ is a continuous function in $[0, \infty)$ and is such that $\pi^I(\nu) = O(\nu^{-1/2})$ as $\nu \to 0$ and $\pi^I(\nu) = O(\nu^{-2})$ as $\nu \to \infty$.

Corollary

Provided that $n \geq p + 1$, (i) the independence Jeffreys prior $\pi^I(\theta)$ and the Jeffreys-rule prior $\pi^R(\theta)$ yield proper posterior densities, and (ii) the marginal posteriors $\pi^I(\nu|y, x)$ and $\pi^R(\nu|y, x)$ do not have any positive integer moments.

Exponential power model: objective Bayesian analysis

Density: $\text{EP}(\mu, \sigma_p, p)$

$$p(y|\mu, \sigma_p, p) = \left[2p^{1/p}\sigma_p\Gamma(1 + 1/p)\right]^{-1} \exp\left[-(p\sigma_p^p)^{-1}|x - \mu|^p\right]$$

with $-\infty < y < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$ and $p > 0$

- $\mu = E(y)$, the location parameter
- $\sigma_p = [E(|y - \mu|^p)]^{1/p}$, the scale parameter
- p the shape parameter

Exponential power model: objective Bayesian analysis

Density: EP(μ, σ_p, p)

$$p(y|\mu,\sigma_p,p) = \left[2p^{1/p}\sigma_p\Gamma(1+1/p)\right]^{-1} \exp\left[-(p\sigma_p^p)^{-1}|x-\mu|^p\right]$$

with $-\infty < y < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$ and $p > 0$

- $\mu = E(y)$, the location parameter
- $\sigma_p = [E(|y-\mu|^p)]^{1/p}$, the scale parameter
- p the shape parameter

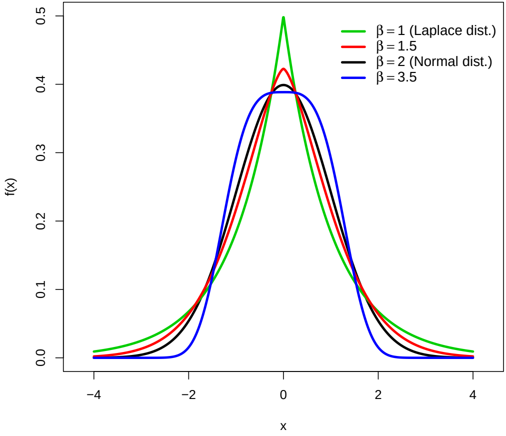
Reparametrization, similar to Zhu & Zinde-Walsh (2009):

$$p(y|\mu,\sigma,p) = \frac{1}{2\sigma} \exp\left[-\left(\frac{\Gamma(1+1/p)|y-\mu|}{\sigma}\right)^p\right]$$

where $\sigma = p^{1/p}\sigma_p\Gamma(1+1/p)$

Characteristics

- $p = 1 \Rightarrow$ Laplace distribution
- $p = 2 \Rightarrow$ Normal model
- $p \to \infty \Rightarrow$ Uniform distribution
- $0 < p < 2 \Rightarrow$ leptokurtic distributions
- $p > 2 \Rightarrow$ platikurtic distributions



Kurtosis

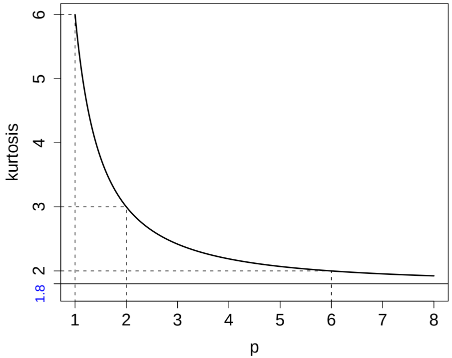


Figure: Kurtosis function for values of p between 1 and 8. Dashed lines represent especial cases: Laplace distribution $(p = 1, \kappa = 6)$, normal distribution $(p = 2, \kappa = 3)$. Horizontal line represents the kurtosis value for uniform distribution $(p \to \infty, \kappa = 1.8)$.

EP regression model

The data consist of n observations $y = (y_1, \ldots, y_n)'$ satisfying

$$y = x\beta + \epsilon$$

where

- $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ is the error vector
- ϵ_i's are i.i.d according to the EP$(0, \sigma, p)$
- x is the known $n \times k$ matrix of explanatory variables assumed to have full rank and
- $\beta = (\beta_1, \ldots, \beta_k)' \in \mathbb{R}^k$ are unknown regression parameters

Jeffreys prior: EP model

The Fisher information matrix on $\Theta = (\theta, \sigma, p)$ is given by

$$I(\theta) = \begin{bmatrix} \frac{1}{\sigma^2}\Gamma(\frac{1}{p})\Gamma(2-\frac{1}{p})\sum_{i=1}^{n} x_i x_i' & 0 & 0 \\ 0 & \frac{np}{\sigma^2} & -\frac{n}{\sigma p} \\ 0 & -\frac{n}{\sigma p} & \frac{n}{p^3}(1+\frac{1}{p})\Psi'(1+\frac{1}{p}) \end{bmatrix}$$

$\Psi(\cdot)$ and $\Psi'(\cdot)$ are the digamma and trigamma functions, respectively. Note that:

- Jeffreys-rule prior is that associated with the single group $\{(\beta, \sigma, p)\}$ and
- Independence Jeffreys prior is that associated with $\{\beta, (\sigma, p)\}$

Jeffreys prior: EP model

The reference prior for $\{\beta, (\sigma, p)\}$ (independence Jeffreys prior) and for $\{(\beta, \sigma, p)\}$ (Jeffreys-rule prior) denoted by $\pi^I(\Theta)$ and $\pi^R(\Theta)$, belong to the class of improper prior distributions given by

$$\pi(\Theta) \propto \frac{\pi(p)}{\sigma^a}$$

where $a \in \mathbb{R}$ and $\pi(p)$ is the marginal prior of the shape parameter p

$$\pi^I(\Theta): \; a = 1, \qquad \pi^I(p) \propto \tfrac{1}{p}\left[\left(1+\tfrac{1}{p}\right)\Psi'\left(1+\tfrac{1}{p}\right)-1\right]^{1/2}$$

$$\pi^R(\Theta): \; a = k+1, \quad \pi^R(p) \propto \left[\Gamma\left(\tfrac{1}{p}\right)\Gamma\left(2-\tfrac{1}{p}\right)\right]^{k/2}\pi^I(p)$$

Another independence prior

$$\pi^{I_2}(\Theta): \; a = 1, \qquad \pi^{I_2}(p) \propto \tfrac{1}{p^{3/2}}\left(1+\tfrac{1}{p}\right)^{1/2}\left[\Psi'\left(1+\tfrac{1}{p}\right)\right]^{1/2}$$

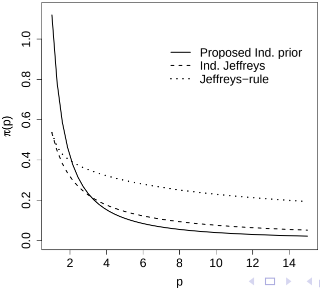
Jeffreys prior: EP model

For $p \to \infty$ $L(p;y) = O(1)$. Moreover:

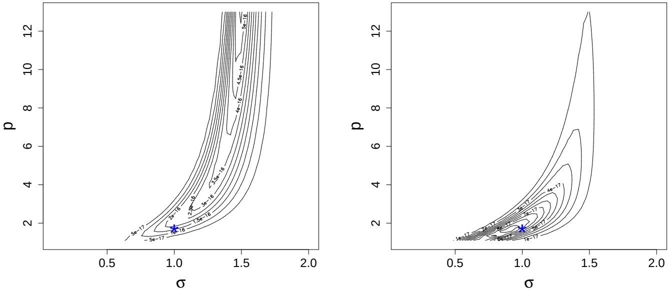
- $\pi^I(p) = O(p^{-1})$ for $p \to \infty$
- $\pi^R(p) = O(p^{k/2-1})$ for $p \to \infty$

  Due to the tail behavior of $\pi^I(p)$ and $\pi^R(p)$ both priors lead to improper posterior distributions
- $\pi^{I_2}(p) = O(p^{-3/2})$ for $p \to \infty$

  This prior leads to a proper posterior distribution



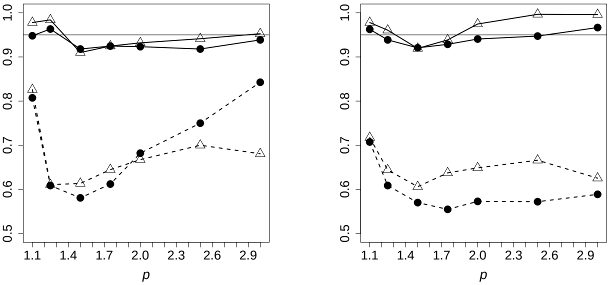
Proposed proper prior as a calibration tool



(a) (b)

Figure: (a) Contour plot of the likelihood function for (σ, p) considering a data set of size $n = 50$ with parameters $\beta = 0$ (fixed), $\sigma = 1$ and $p = 1.8$. (b) Contour plot of the joint posterior distribution of (σ, p) based on the proposed prior. The symbol $*$ indicates the position of the true values.

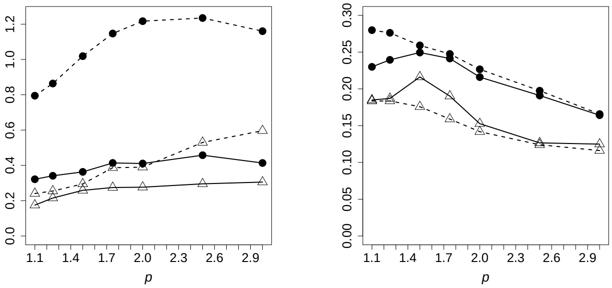
Frequentist properties of Bayesian estimators



(a) Coverage of 95%: p (b) Coverage of 95%: σ

Figure: Frequentist coverage probability of 95% HPD credible intervals (solid line) and 95% confidence interval (dashed line) for p and σ based on proposed prior for $n = 50$ (circle) and $n = 100$ (triangle). Horizontal line indicates the 95% nominal level.

Frequentist properties of Bayesian estimators



(a) $\sqrt{MSE}/p$ (b) $\sqrt{MSE}/\sigma$

Figure: Square root of the relative mean square error of estimators of p (left panel) and σ (right panel) based on the independence Jeffreys prior π^{I_2} (solid line) and maximum likelihood estimation (dashed line) for $n = 50$ (circle) and $n = 100$ (triangle).

Data analysis

Application 1: Excess returns for Martin Marietta company

- 60 monthly observations (January,1982 to December, 1986.
- Variables: The excess rate of return for the Martin Marietta company 20 (y) and the index for the excess rate returns (x) for the New York stock exchange (CRSP).

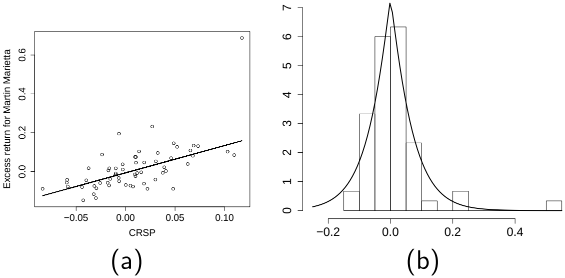


Figure: (a) Scatterplot of the data and fitted EP regression model. (b) Histogram of the residuals from the fitted EP regression model and fitted density (solid line).

Application 1: Excess returns for Martin Marietta company

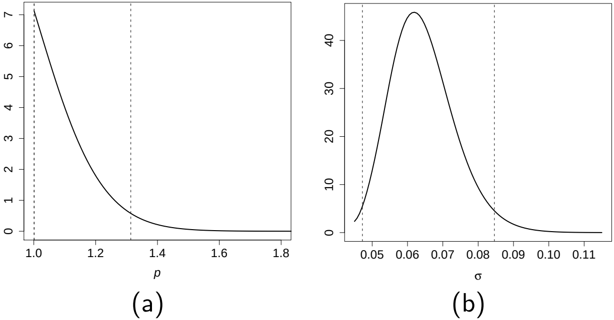


Figure: Marginal posterior densities for (a) p and (b) σ. Vertical dashed lines are the posterior 95% HPD credible intervals.

Application 1: Excess returns for Martin Marietta company

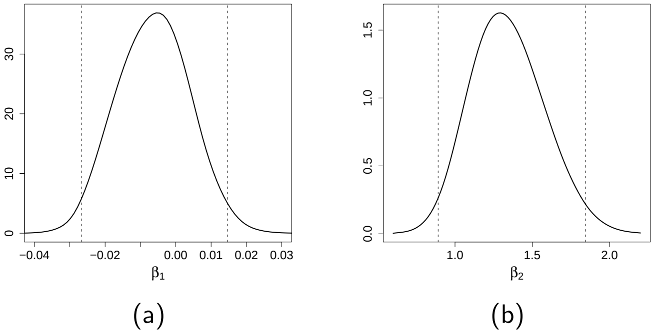


Figure: Marginal posterior densities for (a) β_1 and (b) β_2. Vertical dashed lines are the posterior 95% HPD credible intervals.

Application 1: Excess returns for Martin Marietta company

Table: Posterior summaries: median, mode and 95% HPD credible interval, based on the independence Jeffreys prior π^{I_2}.

Parameter	Median	Mode	95% C.I.
β_1	-0.006	-0.006	(-0.027, 0.014)
β_2	1.327	1.295	(0.891, 1.844)
σ	0.064	0.062	(0.047, 0.085)
p	1.092	1.000	(1.000, 1.314)

Thank you

migon@im.ufrj.br