Brazilian Journal of Probability and Statistics (2003), 17, pp. 1–16.

Bayesian inference of a linear segmented regression model

Carlos Alberto R. Diniz[1], **Luis A. Milan**[1]
and
Josmar Mazucheli[2]
[1] *Universidade Federal de São Carlos*
[2] *Universidade Estadual de Maringá*

Abstract: Robust Bayesian Inference in a linear-linear segmented regression model, assuming non-homogeneous error variance, is explored for cases in which the errors follow Student's t distributions. Metropolis-within-Gibbs algorithms are used in order to estimate the posterior distribution of the change point and the model parameters. The methodology is illustrated by the analysis of two simulated data sets and a real data set from a clinical study designed to determine the anaerobic threshold of a healthy male during dynamic exercise.

Key words: Bayesian analysis; metropolis-within-Gibbs algorithm; segmented regression models.

1 Introduction

A Bayesian approach based on Gibbs sampling with Metropolis-Hasting algorithm is considered for the problem of making inferences about the change point in a segmented non-homogeneous error variance regression model supposing the error terms with multivariate t distributions with small number of degrees of freedom.

A heteroscedastic linear-linear model in the case that a sequence of observations (X_i, Y_i) for $i = 1, 2, \ldots, T$ is considered, may be written as

$$Y_i = \begin{cases} \alpha_1 + \beta_1 X_i + e_{1i}, & i = 1, 2, \ldots, \lambda; \\ \alpha_2 + \beta_2 X_i + e_{2i}, & i = \lambda + 1, \lambda + 2, \ldots, T. \end{cases}$$

where Y is assumed random and X is fixed, e_{1i} and e_{2i} are the error terms, $\alpha_1, \beta_1, \alpha_2, \beta_2$ are the model parameters and λ is the unknown change point. It will be convenient to rewrite the above model in matrix form, i.e., for $\mathbf{Y}_1(\lambda) = (Y_1, \ldots, Y_\lambda)'$, $\mathbf{Y}_2(\lambda) = (Y_{\lambda+1}, \ldots, Y_T)'$, $\mathbf{e}_1 = (e_1, \ldots, e_\lambda)'$ and $\mathbf{e}_2 = (e_{\lambda+1}, \ldots, e_T)'$,

$$\begin{pmatrix} \mathbf{Y}_1(\lambda) \\ \mathbf{Y}_2(\lambda) \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1(\lambda) & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2(\lambda) \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{pmatrix}, \tag{1.1}$$

where $\theta_1 = (\alpha_1, \beta_1)'$, $\theta_2 = (\alpha_2, \beta_2)'$, $\mathbf{0}$ is the zero matrix, $\mathbf{X}_1(\lambda)$ is the regression matrix for the first λ observations and $\mathbf{X}_2(\lambda)$ for the last $T - \lambda$ observations. The error vectors $\mathbf{e}_j$ $(j = 1, 2)$ are assumed to be mutually independent and each have a joint multivariate Student-t distribution with zero location vector, scale matrix $\sigma_j^2 I_{d_j}$, and v_j degrees of freedom, where $d_1 = \lambda$ and $d_2 = T - \lambda$. The density function for $\mathbf{e}_j$ is of the form

$$f(\mathbf{e}_j|v_{j,}\sigma_j^2) = \frac{v_j^{d_j/2}\Gamma\left[(v_j + d_j)/2\right]}{(\sigma_j^2)^{d_j/2}\pi^{d_j/2}\Gamma(v_j/2)}(v_j + \mathbf{e}_j'\mathbf{e}_j/\sigma_j^2)^{-(d_j+v_j)/2}, \qquad (1.2)$$

where $\sigma_j > 0$, $j = 1, 2$, $-\infty < e_{1i},\ e_{2i} < \infty$, $i = 1, 2, \ldots, T$.

The two-phase regression model described in (1.1) is an extension of the usual linear model under the assumption that the error vector has a multivariate t-distribution (Zellner, 1976) to allow the inclusion of a change-point. The segmentation of the model, useful when it is suspected that the model parameters change after an unknown time, makes the inference problem complicated. The difficulties of deriving the maximum likelihood estimators of the change-point and for the parameters in the model are apparent.

Bayesian methods in segmented regression models have received some attention in a series of papers. Ferreira (1975), assuming three different prior distributions of the change point parameter, provides marginal posterior distributions for all parameters of the simple linear-linear segmented regression model. Holbert and Broemeling (1977), assigning improper uniform density to all regression parameters, present marginal posterior density for the change point. Chin Choy and Broemeling (1980), working with multiple linear-linear segmented regression model, provide posterior distributions of the change point, variance and regression parameter vectors. Bayesian methods in segmented time series regression models are discussed in Gelfand and Sfiridis (1996), Kim and Nelson (1999), and Perreault *et al.* (1999).

Many of these works are based on normal error distribution and constant error variance assumptions. It is well known that there are circumstances in which models based on normal error distribution are not adequate. For instance, problems where the parent distribution differs from the normal due the presence of heavy tails, or in the case where there are unusual or outlying values, or the data are too dispersed (see Box and Tiao, 1973; Hsu, 1982a; Gelman *et al.*, 1995).

Metropolis-within-Gibbs algorithm is discussed in Gelfand and Smith (1990), Smith and Roberts (1993), and Chib and Greenberg (1995). Discussion of Student-t error terms models can be found in Zellner (1976), who analyzes the usual linear multiple regression model under the assumption that the error vector has a multivariate t-distribution; Sutranghar and Ali (1986), extend Zellner's regression model to the case of several variables; Lange *et al.* (1989), suggest models with multivariate t errors for many different problems; and Tsurumi (1977), presents a Bayesian test for a change point in two regressions assuming errors terms with multivariate Student-t distributions.

The paper is organized as follows. In Section 2 is proposed a Bayesian version of the heteroscedastic linear-linear model; in Section 3 is described the Metropolis-within-Gibbs algorithm. Two simulated data sets are analyzed in Section 4 and a real data set from a clinical study is analyzed in Section 5. A discussion is also presented in Section 5. General comments in Section 6 concludes the paper.

2 The Bayesian framework

A classical statistical method for fitting heteroscedastic linear-linear model with multivariate t distribution errors is not available. The reasons could be the difficulties of deriving the maximum likelihood estimators of the parameters. Considering such fact and the possibility of introducing in the model prior information on the parameters, a Bayesian inference is adopted in this study.

Using (1.1) and (1.2), the likelihood function for the regression model, given λ, θ_1, θ_2, σ_1^2, σ_2^2, is

$$p(\mathbf{Y}_1, \mathbf{Y}_2|\lambda, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2) =$$

$$\prod_{j=1}^{2} C_j[v_j + (\mathbf{Y}_j(\lambda) - \mathbf{X}_j(\lambda)\theta_j)'(\mathbf{Y}_j(\lambda) - \mathbf{X}_j(\lambda)\theta_j)/\sigma_j^2]^{-(d_j+v_j)/2}, \qquad (2.1)$$

where

$$C_j = \frac{v_j^{d_j/2}\Gamma((v_j + d_j)/2)}{(\sigma_j^2)^{d_j/2}\pi^{d_j/2}\Gamma(v_j/2)}, j = 1, 2.$$

Assuming independence between (θ_1,σ_1^2) and (θ_2,σ_2^2) and noninformative priors for the model parameters, the joint prior distribution for $\theta_1, \theta_2, \sigma_1^2$, and σ_2^2 is

$$\pi(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2) = \pi(\theta_1,\sigma_1^2)\pi(\theta_2,\sigma_2^2) \propto \frac{1}{\sigma_1^2}\frac{1}{\sigma_2^2}. \qquad (2.2)$$

Combining the likelihood function (2.1) with the prior density previously mentioned, the joint posterior density is

$$\begin{aligned} pi(\lambda, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2|\mathbf{Y}) &\propto \pi(\lambda)\frac{1}{\sigma_1^2}\frac{1}{\sigma_2^2} \\ &\quad \times \prod_{j=1}^{2} C_j \left[v_j + (\mathbf{Y}_j(\lambda) - \mathbf{X}_j(\lambda)\theta_j)'(\mathbf{Y}_j(\lambda) - \mathbf{X}_j(\lambda)\theta_j)/\sigma_j^2\right]^{-(d_j+v_j)/2} \\ &\propto \quad \pi(\lambda)\frac{1}{\sigma_1^2}\frac{1}{\sigma_2^2} \times \prod_{j=1}^{2} C_j \left[v_j + \left(\mathbf{Y}_j(\lambda) - \widehat{\mathbf{Y}}_j(\lambda)\right)' \left(\mathbf{Y}_j(\lambda) - \widehat{\mathbf{Y}}_j(\lambda)\right)/\sigma_j^2\right. \\ &\quad \left. + (\theta_j - \widehat{\theta}_j)'\mathbf{X}_j'(\lambda)\mathbf{X}_j(\lambda)(\theta_j - \widehat{\theta}_j)/\sigma_j^2\right]^{-(d_j+v_j)/2}, \end{aligned}$$

where $\pi(\lambda)$ is the prior distribution of the change point parameter λ, $\widehat{\mathbf{Y}}_j(\lambda) = \mathbf{X}_j(\lambda)\widehat{\theta}_j$, and $\widehat{\theta}_j = \left(\mathbf{X}_j'(\lambda)\mathbf{X}_j(\lambda)\right)^{-1}\mathbf{X}_j'(\lambda)\mathbf{Y}_j(\lambda)$, $j = 1, 2$. The most common prior distribution of the change point parameter λ is the uniform distribution (Ferreira, 1975; Holbert and Broemeling, 1977).

3 The Metropolis-within-Gibbs algorithm

The marginal posterior densities of the parameters in equation (2.3) are not easily obtained. An alternative to handle this situation is the use of the Metropolis-within-Gibbs algorithm.

The Metropolis-within-Gibbs algorithm allows simulation of observations from a complex joint distribution. Here the posterior simulation is performed in part using the Gibbs sampler (Casella and George, 1992), for the cases where the conditional posterior distributions of the parameters are available for sampling, and in part using the Metropolis-Hasting algorithm (Chib and Greenberg, 1995), for the cases when the conditional distributions are not of a common form.

3.1 The conditional distributions

The conditional posterior density for θ_1 given λ, θ_2, σ_1^2, σ_2^2 and $\mathbf{Y}$ is

$$\begin{aligned}\pi(\theta_1|\lambda,\theta_2,\sigma_1^2,\sigma_2^2,\mathbf{Y}) &\propto C_1\Big[v_1+\left(\mathbf{Y}_1(\lambda)-\widehat{\mathbf{Y}}_1(\lambda)\right)'\left(\mathbf{Y}_1(\lambda)-\widehat{\mathbf{Y}}_1(\lambda)\right)/\sigma_1^2\\ &\quad +(\theta_1-\widehat{\theta}_1)'\mathbf{X}_1'(\lambda)\mathbf{X}_1(\lambda)(\theta_1-\widehat{\theta}_1)/\sigma_1^2\Big]^{-(\lambda+v_1)/2}\\ &\propto \big[(\sigma_1^2 v_1+(\lambda-2)S_1)/\sigma_1^2+ \\ &\quad (\theta_1-\widehat{\theta}_1)'\mathbf{X}_1'(\lambda)\mathbf{X}_1(\lambda)(\theta_1-\widehat{\theta}_1)/\sigma_1^2\big]^{-(\lambda+v_1)/2} \qquad (3.1)\\ &\propto \left[v+\frac{v(\theta_1-\widehat{\theta}_1)'\mathbf{X}_1'(\lambda)\mathbf{X}_1(\lambda)(\theta_1-\widehat{\theta}_1)}{(\sigma_1^2 v_1+(\lambda-2)S_1)}\right]^{-(v+2)/2},\end{aligned}$$

where $(\lambda-2)S_1=\left(\mathbf{Y}_1(\lambda)-\widehat{\mathbf{Y}}_1(\lambda)\right)'\left(\mathbf{Y}_1(\lambda)-\widehat{\mathbf{Y}}_1(\lambda)\right)$ and $v=\lambda-2+v_1$.

The expression in the last line of (3.1) is readily recognized to be the kernel of a multivariate Student-t distribution with location vector $\widehat{\theta}_1$, scale matrix

$$\mathbf{COV}(\theta_1)=\left(\mathbf{X}_1'(\lambda)\mathbf{X}_1(\lambda)\right)^{-1}(\sigma_1^2 v_1+(\lambda-2)S_1)/(v-2),$$

and v degrees of freedom, $v>2$. Thus

$$\theta_1|\lambda,\theta_2,\sigma_1^2,\sigma_2^2,\mathbf{Y} \sim t_2\left(\widehat{\theta}_1,\left(\mathbf{X}_1'(\lambda)\mathbf{X}_1(\lambda)\right)^{-1}(\sigma_1^2 v_1+(\lambda-2)S_1)/(v-2),v\right). \qquad (3.2)$$

Similarly, the conditional posterior density for θ_2 given λ, θ_1, σ_1^2, σ_2^2 and $\mathbf{Y}$ is in the form of a multivariate Student-t distribution with location vector $\widehat{\theta}_2$, scale matrix

$$\mathbf{COV}(\theta_2)=\left(\mathbf{X}_2'(\lambda)\mathbf{X}_2(\lambda)\right)^{-1}(\sigma_2^2 v_2+(T-\lambda-2)S_2)/(\omega-2),$$

and ω degrees of freedom, $\omega>2$, where

$$(T-\lambda-2)S_2=\left(\mathbf{Y}_2(\lambda)-\widehat{\mathbf{Y}}_2(\lambda)\right)'\left(\mathbf{Y}_2(\lambda)-\widehat{\mathbf{Y}}_2(\lambda)\right)$$

and $\omega = T - \lambda - 2 + v_2$. Thus

$$\theta_2|\lambda, \theta_1, \sigma_1^2, \sigma_2^2, \mathbf{Y} \sim t_2\left(\widehat{\theta}_2, (\mathbf{X}_2'(\lambda)\mathbf{X}_2(\lambda))^{-1}(\sigma_2^2 v_2 + (T-\lambda-2)S_2)/(\omega-2), \omega\right). \tag{3.3}$$

The conditional posterior density for λ given θ_1, θ_2, σ_1^2, σ_2^2 and $\mathbf{Y}$ is

$$\begin{aligned}
\pi(\lambda|\theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \mathbf{Y}) \quad \propto \quad & \pi(\lambda)\frac{v_1^{\lambda/2}\Gamma((v_1+\lambda)/2)}{\pi^{\lambda/2}}\left(\sigma_1^2 v_1 + (\lambda-2)S_1\right)^{-(\lambda+v_1)/2} \\
& \times \left[1 + \frac{(\theta_1 - \widehat{\theta}_1)'\mathbf{X}_1'(\lambda)\mathbf{X}_1(\lambda)(\theta_1 - \widehat{\theta}_1)}{\sigma_1^2 v_1 + (\lambda-2)S_1}\right]^{-(\lambda+v_1)/2} \\
\times \frac{v_2^{(T-\lambda)/2}\Gamma((v_2+T-\lambda)/2)}{\pi^{(T-\lambda)/2}} & \left(\sigma_2^2 v_2 + (T-\lambda-2)S_2\right)^{-(T-\lambda+v_2)/2} \\
& \times \left[1 + \frac{(\theta_2 - \widehat{\theta}_2)'\mathbf{X}_2'(\lambda)\mathbf{X}_2(\lambda)(\theta_2 - \widehat{\theta}_2)}{\sigma_2^2 v_2 + (T-\lambda-2)S_2}\right]^{-(T-\lambda+v_2)/2}, \\
for \quad \lambda \quad = \quad & 2, \ldots, T-2.
\end{aligned}$$

Despite the choice of the prior $\pi(\lambda)$, it is impracticable to simulate from this distribution. In the numerical illustration presented in next section, $\pi(\lambda)$ is the uniform distribution.

The conditional posterior density for σ_1^2 given λ, θ_1, θ_2, σ_2^2 and $\mathbf{Y}$ is

$$\begin{aligned}
\pi(\sigma_1^2|\lambda, \theta_1, \boldsymbol{\theta_2}, \sigma_2^2, \mathbf{Y}) & \propto (\sigma_1^2)^{(-1+v_1/2)}\left(\sigma_1^2 v_1 + (\lambda-2)S_1\right)^{-(\lambda+v_1)/2} \\
& \times \left[1 + \frac{(\theta_1 - \widehat{\theta}_1)'\mathbf{X}_1'(\lambda)\mathbf{X}_1(\lambda)(\theta_1 - \widehat{\theta}_1)}{\sigma_1^2 v_1 + (\lambda-2)S_1}\right]^{-(\lambda+v_1)/2}.
\end{aligned}$$

The conditional posterior density for σ_2^2 given λ, θ_1, θ_2, σ_1^2 and $\mathbf{Y}$ is

$$\begin{aligned}
\pi(\sigma_2^2|\lambda, \theta_1, \boldsymbol{\theta_2}, \sigma_1^2, \mathbf{Y}) & \propto (\sigma_2^2)^{(-1+v_2/2)}\left(\sigma_2^2 v_2 + (T-\lambda-2)S_2\right)^{-(T-\lambda+v_2)/2} \\
& \times \left[1 + \frac{(\theta_2 - \widehat{\theta}_2)'\mathbf{X}_2'(\lambda)\mathbf{X}_2(\lambda)(\theta_2 - \widehat{\theta}_2)}{(\sigma_2^2 v_2 + (T-\lambda-2)S_2}\right]^{-(T-\lambda+v_2)/2}.
\end{aligned}$$

3.2 The generations schemes

Given the starting values $\theta_1^{(0)} = (\alpha_1^{(0)}, \beta_1^{(0)})'$, $\theta_2^{(0)} = (\alpha_2^{(0)}, \beta_2^{(0)})'$, $\lambda^{(0)}$, $\sigma_1^{2^{(0)}}$, $\sigma_2^{2^{(0)}}$, generate $\theta_1^{(1)}$ from the conditional multivariate Student-t distribution (3.2) given $\lambda = \lambda^{(0)}$, $\theta_2 = \theta_2^{(0)}$, $\sigma_1^2 = \sigma_1^{2^{(0)}}$, and $\sigma_2^2 = \sigma_2^{2^{(0)}}$, then generate $\theta_2^{(1)}$ from the conditional multivariate Student-t distribution (3.3) given $\lambda = \lambda^{(0)}$, $\theta_1 = \theta_1^{(1)}$, $\sigma_1^2 = \sigma_1^{2^{(0)}}$, and $\sigma_2^2 = \sigma_2^{2^{(0)}}$, in both cases using Gibbs sampling scheme. The procedure described by Johnson e Kotz (1972), page 144, is used to generate samples from the multivariate Student-t distribution.

Generate $\sigma_1^{2^{(1)}}$ using Metropolis sampling scheme, generating a candidate σ_{c1}^2 from a distribution $Gamma(v_1/2, 2S_1/v_1)$, where v_1 corresponds to the degrees of freedom of the multivariate t distribution of the error vector $\mathbf{e}_1$ associated to the first subset and S_1 is the sample variance of the first $\lambda^{(0)}$ observations. Submit the candidate generated to the Metropolis acceptance/rejection test as described in Chib and Greenberg (1995), generating a value U from $U(0,1)$, if $U \leq \psi(\sigma_1^{2^{(0)}}, \sigma_{c1}^2)$, set $\sigma_1^{2^{(1)}} = \sigma_{c1}^2$, else, set $\sigma_1^{2^{(1)}} = \sigma_1^{2^{(0)}}$, where

$$\varphi(\sigma_1^{2^{(0)}}, \sigma_{c1}^2) = \min\left[\frac{\pi(\sigma_{c1}^2|\lambda^{(0)}, \theta_1^{(1)}, \theta_2^{(1)}, \sigma_2^{2^{(0)}}, \mathbf{Y})}{\pi(\sigma_1^{2^{(0)}}|\lambda^{(0)}, \theta_1^{(1)}, \theta_2^{(1)}, \sigma_2^{2^{(0)}}, \mathbf{Y})}, 1\right].$$

A similar procedure is used to generate $\sigma_2^{2^{(1)}}$, with a candidate σ_{c2}^2 generated from a distribution $Gamma(v_2/2, 2S_2/v_2)$.

The discrete parameter $\lambda^{(1)}$ is obtained by generating U from a distribution $U(0,1)$, applying the search-in-table procedure, i.e., $\lambda^{(1)}$ must satisfy

$$\sum_{k=2}^{\lambda^{(1)}} \pi(k|\theta_1^{(1)}, \theta_2^{(1)}, \sigma_1^{2^{(1)}}, \sigma_2^{2^{(1)}}, \mathbf{Y}) < U \leq \sum_{k=2}^{\lambda^{(1)}+1} \pi(k|\theta_1^{(1)}, \theta_2^{(1)}, \sigma_1^{2^{(1)}}, \sigma_2^{2^{(1)}}, \mathbf{Y}).$$

Repeat this process using ($\theta_1^{(1)}$, $\theta_2^{(1)}$, $\lambda^{(1)}$, $\sigma_1^{2^{(1)}}$, $\sigma_2^{2^{(1)}}$) as starting values and so on.

4 Simulation studies

The methodology proposed is illustrated by two simulated sets of $T = 60$ observations, simulated from the model (1.1) with the parameters as presented in Tables 1 and 2. The scatterplots of the pairs (x, y) for the two datasets are showed in Figure 1. The first dataset presents a ramp-ramp situation with change-point at position $\lambda = 20$. It means that the first 20 observations were taken from a multivariate t distribution with location vector $\mathbf{X}_1\theta_1$, and scale matrix $\sigma_1^2\mathbf{I}_{20}$ and the remaining 40 observations from a multivariate t distribution with location vector $\mathbf{X}_2\theta_2$ and scale matrix $\sigma_2^2\mathbf{I}_{40}$. The second dataset also presents a ramp-ramp situation but with change-point at position $\lambda = 30$, at the center of the series. In both cases, the parameter vectors are $\theta_1^{'} = (\alpha_1, \beta_1)^{'} = (2, -8)^{'}$, $\theta_2^{'} = (\alpha_2, \beta_2)^{'} = (-4, 8)^{'}$, and variances given by $\sigma_1^2 = 1$ and $\sigma_2^2 = 4$.

Lange *et al.* (1989) suggest, based on their applications, that, for small samples, the degrees of freedom of the proposed multivariate t-distribution model being set as $v = 4$. The numerical examples presented in this section assume that the degrees of freedom used for the simulation of the dataset are $v_1 = v_2 = 3$.

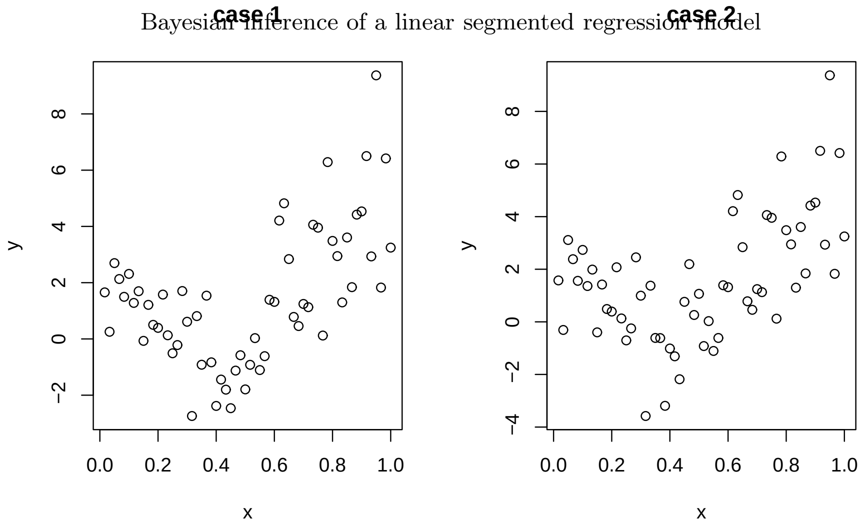

Figure 1 *Plots of two data sets with $T = 60$ observations. Case 1 presents a ramp-ramp shape with change-point at position $\lambda = 20$. Case 2 presents a ramp-ramp shape with change-point at the center position $\lambda = 30$.*

Table 1 *True Parameter and Posterior Estimates - Case 1*

Parameters	True values	Posterior Median	95% Credible Interval
λ	20	18.00	[13.00, 34.00]
σ_1^2	1	1.30	[0.10, 6.61]
σ_2^2	4	2.92	[0.27, 13.92]
α_1	2	2.05	[0.72, 3.44]
β_1	-8	-6.23	[-13.42, 1.77]
α_2	-4	-4.38	[[-7.12, -0.94]]
β_2	8	9.48	[5.03, 13.24]

Table 2 *True Parameter and Posterior Estimates - Case 2*

Parameters	True values	Posterior Median	95% Credible Interval
λ	30	33.00	[18.00, 34.00]
σ_1^2	1	0.88	[0.07, 3.85]
σ_2^2	4	3.29	[0.30, 16.23]
α_1	2	2.00	[1.19, 2.88]
β_1	-8	-6.77	[-10.01, -3.94]
α_2	-4	-4.14	[-8.90, 1.76]
β_2	8	9.19	[1.89, 15.28]

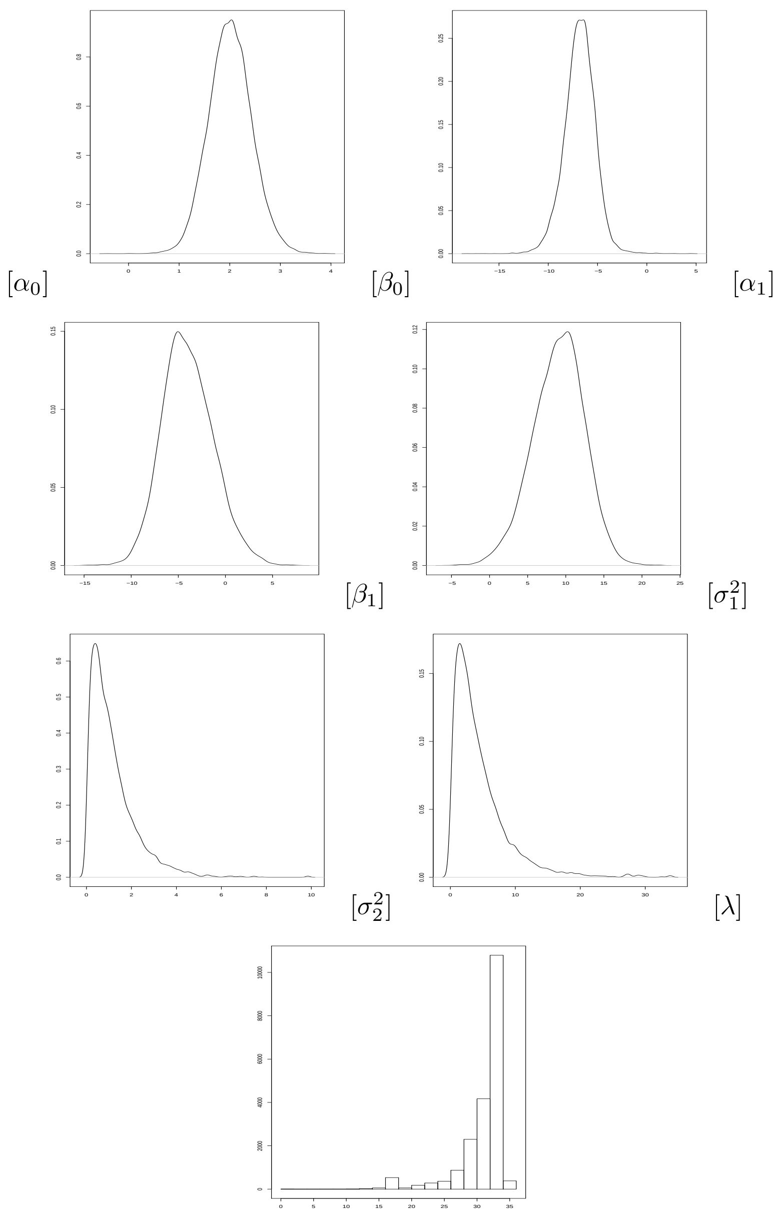

Figure 2 *Case 1 - Posterior densities of the model parameters and histogram of the change point.*

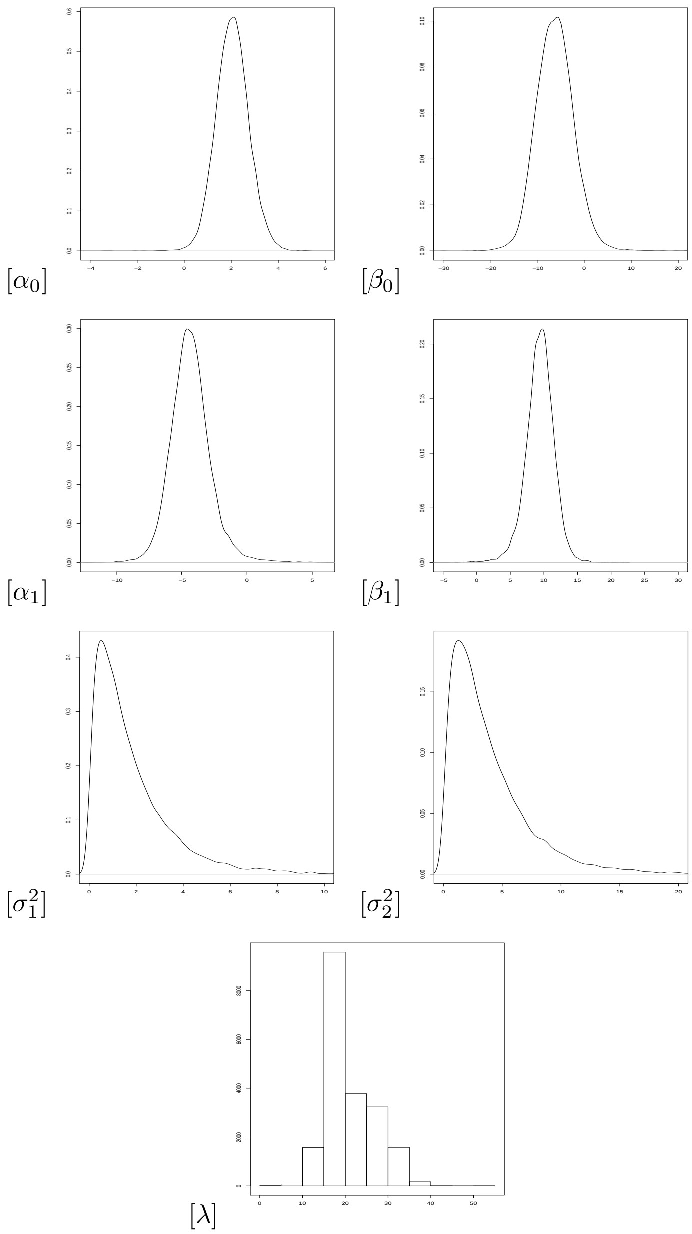

Figure 3 *Case 2 - Posterior densities of the model parameters and histogram of the change point.*

Characteristics of the posterior distributions of the parameters λ, σ_1^2, σ_2^2 α_1, β_1, α_2 and β_2 can be calculated from samples generated by the Metropolis-within-Gibbs technique. For the two cases, a Pentium II 333 MHZ and the statistical software R (Ihaka and Gentleman, 1996) are used to generate two chains of 10001 iterations each for all seven parameters. The first 1000 values are discarded (burn-in) and, since the autocorrelation within each chain for each parameter at lag 1 is non significan, the remaining sample is taken forward for analysis.

The posterior distributions of λ, α_1, β_1, α_2, β_2, σ_1^2, and σ_2^2, for the two cases, are also summarized in Tables 1 and 2. Posterior medians (the medians of the simulated values), and the 95% credibility intervals are presented. The Metropolis procedure has a good acceptance rate and do not present difficulties of implementation. The same happens with the Gibbs steps of the method. The initial values are not critical, provided that exotic values are not used.

Figure 2 and 3 present the approximate marginal posterior densities based on the combined chains for all the parameter models and the change point for both cases. The convergence of the chains is verified using CODA package (Best *et al.,* 1995), showing that the chains converged.

4.1 Working with outliers

A new data set of $T = 60$ observations contaminated with two outliers is simulated from the model (1.1). The scatterplot of the pairs (x, y) is showed in Figure 4. The two outliers are indicated by cross marks. The posterior distributions of λ, α_1, β_1, α_2, β_2, σ_1^2, and σ_2^2, is summarized in Table 3. The convergence of the chains is verified using CODA package (Best *et al.,* 1995), showing that the chains converged.

As shown in the Table 3, in the case that the data set is contaminated with outliers, the methodology proposed in this article provides resistant result with respect to the estimation of the change point. However, the point estimators of the model parameters, associated with the region of the outliers (submodel 1), are influenced by the presence of extreme observations, and the credible intervals for those parameters and for the change point became wider than the credible intervals calculated from the same data set without outliers.

Table 3 *True Parameter and Posterior Estimates - Case with outliers*

Parameters	True values	Posterior Median	95% Credible Interval
λ	20	20.00	[6.00, 24.00]
σ_1^2	1	7.39	[0.60, 39.95]
σ_2^2	4	7.55	[0.65, 32.85]
α_1	4	6.01	[0.94, 9.32]
β_1	-8	-10.76	[-20.27, 22.10]
α_2	-2	-1.05	[-5.18, 3.43]
β_2	8	7.74	[4.65, 10.63]

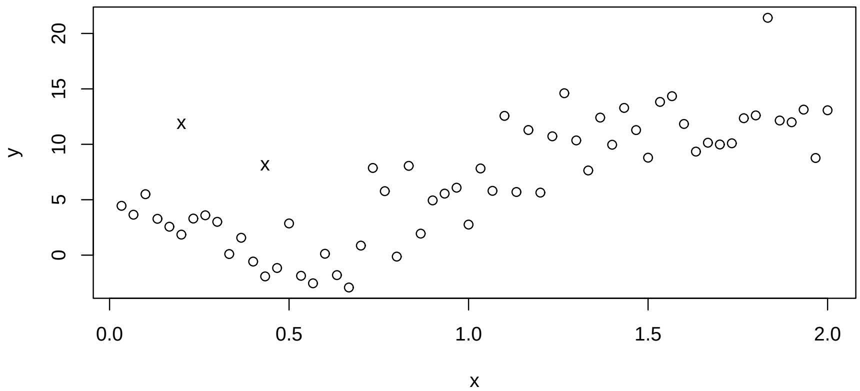

Figure 4 *Plots of a data set with $T = 60$ observations with/without outliers.*

5 Application

In this section the methodology developed in this article is illustrated considering a real data set. The data comes from a clinical study designed to determine the anaerobic threshold of a healthy male during dynamic exercise (Marães et al., 2000). Anaerobic threshold is defined as the level of power output or oxygen consumption just below that at which metabolic acidosis (blood lactate accumulation) and associated changes in gas exchange occur (Chwalbinska-Moneta et al., 1989), or it is also defined as the intensity of the exercise on which the production of energy for the aerobic metabolism is supplemented for the anaerobic metabolism (Beaver et al., 1986). There exist different techniques of determinating the Anaerobic threshold presented in the literature, among them is the one based on the heart rate variability (HRV)(Marães et al., 2000). This method deals with the fact that the anaerobic threshold during dynamic exercise signals a point of shift in the physiological state in such a way that at this point or close to it many ventilatory metabolic and cardiovascular variables suffer certain changes, including HRV.

The data set involves the heart rate (HR), recorded beat-to-beat, in real time, through software STEF (Silva et al., 1994), of a young volunteer during continuous dynamic exercise test at a ramp with initial load of 4 Watts (W) for 2 minutes, proceeded by an increment of 20 W every minute until the individual's exhaustion. The data set used for the analysis is displayed at Figure 4.

In this protocol of exercise test occurs a linear elevation of the HR in a segmented form as shown in the Figure 5. The point where happens the change of trend of the heart rate as a function of the time on exercise intensity is defined as the anaerobic threshold.

The biological cycle in HR, which is of difficult estimation, caused by a combination of factors, yields a small variability in the average of the responses along the time. This variability justify the use of a heavy tail distribution for the model. The segmented regression model based on Student's t distribution is used to estimate the relation between HR and time and, consequently, to identify the anaerobic threshold. In Table 3 it is presented the posterior summaries for the parameters, including the posterior medians, the associated posterior standard deviations and the 95% Credibility Interval. The heteroscedasticity assumption, indicated by the plot of the original data (Figure 4), is confirmed by the results showed in Table 3. Using λ and the coefficients of the segmented regression model equal to the posterior medians, the fitted regression equation have been superimposed on the plot of the observed HR versus time in Figure 5. It can be conclude that, for this case, the anaerobic threshold occur between the 336s and 356s, using a load of 64W.

Table 4 *The Posterior estimates of the heart rate data*

Parameters	Posterior Median	95% Credible Interval
λ	292	[278, 304]
Time	347	[336, 356]
σ_1^2	11.065	[0.792, 45.423]
σ_2^2	8.487	[0.687, 34.578]
α_1	74.245	[72.798, 75.830]
β_1	0.005	[-0.002, 0.011]
α_2	27.724	[26.316, 29.096]
β_2	0.141	[0.139, 0.144]

6 Conclusion

The use of segmented regression model based on the univariate or multivariate normal distribution has been considered by several authors. In this article an alternative to the normal model is discussed. The error terms are supposed to follow Student's t distributions with few degrees of freedom.

A Bayesian analysis of segmented regression models is implemented using Metropolis-within-Gibbs MCMC methods. The computer intensive method presented in this article can be easily implemented in R and gives accurate results.

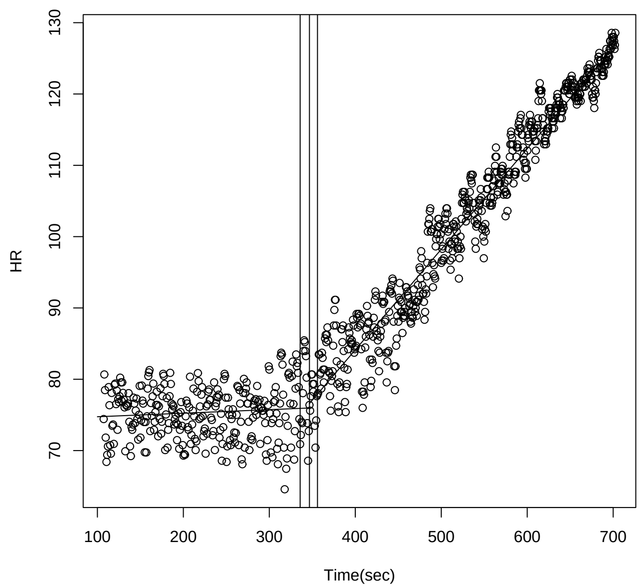

Figure 5 *The heart rate, recorded beat-to-beat, as function of time on exercise intensity of a young volunteer during continuous dynamic exercise test at a ramp, together with the fitted regression equation superimposed (solid lines), the 95% HPD and the posterior median for the change point (vertical lines).*

Acknowledgements

The authors are grateful to Dra. Ester da Silva of the Departamento de Fisioterapia - UFSCar, SP, Brazil, for providing the data set and for fruitful discussions on anaerobic threshold.

(Received September, 2001. Accepted February, 2003.)

References

Beaver, W. I., Wasserman, K. and Whipp, B. J. (1986). A new method for detecting the anaerobic threshold by gas exchange. *J. Appl. Physiol.,* **60**, 2020-2027.

Best, N. G., Cowles, M. K., and Vines, S. K. (1995). *CODA manual version 0.30.* MRC Biostatistics Unit, Cambridge, UK, **46**, 2020-2027.

Box, G. E. P., and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Reading, Mass., USA: Addison-Wesley.

Broemeling, L., and Chin Choy, J. H. (1981). Detecting structural change in linear models. *Commun. Statist.-Theor. Meth. A,* **10**, 2551-2561.

Chwalbinska-Moneta, J., Robergs, R. A., Costill, D. L., Eink, W. J. (1989). Threshold for muscle lactate accumulation during progressive exercise. *American Physiological Society*, 2710-2715.

Casella, G., and George, E. (1992). Explaining the Gibbs sampler. *The American Statistician*, **46**, 167-174.

Chib, S. e Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49**, 327-355.

Ferreira, P. E. (1975). A Bayesian analysis of a switching regression model: known number of regimes. *J. Amer. Statist. Assoc.*, **70**, 370-74.

Gelfand, A. E., and Sfiridis, J. M. (1996). Bayesian analysis of financial event studies data. *Advances in Econometrics*, **11,** 25-62, Part A.

Gelfand, A. E., and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.*, **85**, 398-409.

Gelman, A., Carlin, J. B., and Stern, H. S. and Rubin, D. B. (1995). *Bayesian Data Analysis.* London, UK: Chapman and Hall.

Holbert, D., and Broemeling, L. (1977). Bayesian inferences related to shifting sequences and two-phase regression. *Commun. Statist.-Theor. Meth. A,* **6**, 265-275.

Hsu, D. A. (1982a). A Bayesian robust detection of shift in the risk structure of stock-market returns. *J. Amer. Statist. Assoc.*, **77**, 29-39.

Hsu, D. A. (1982b). Robust inference for structural shift in regression models. *Journal of Econometrics*, **19**, 89-107.

Ihaka, R., and Gentleman, R. (1996). R: a language for data analysis and graphics. *J. Comput. Graphical Statist.*, **5**, 299-341.

Kim, C. J., and Nelson, C. R. (1999). Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. *Review of Economics and Statistics,* **81**, 608-616.

Lange, K. L., Little, R. J. A., and Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *J. Amer. Statist. Assoc.*, **84**, 881-896.

Marães, V. R. F. S., Silva, E., Catai, A. M., Ribeiro, T. F., Moura, M. A., Trevelin, L. C., Oliveira, L., and Gallo Jr., L. (2000). The heart rate variability in dynamic exercise. Its possible role to signal anaerobic threshold (Abstract). *The Physiologist.*, **43**, 339-339.

Perreault, L., Hache, M., Slivitzky, M., and Bobee, B. (1999). Detection of changes in precipitation and runoff over eastern Canada and US using a Bayesian approach. *Stochastic Environmental Research and Risk Assessment*, **13**, 201-216.

Silva, E., Catai, A. M., Trevelin, L. C., Guimaraes, J. O., Silva Jr, L. P., Oliveira, L., Milan, L. A., Martins, L. E. B., and Gallo Jr., L. (1994). Design of computerized system to evaluate the cardiac function during dynamic exercise. In *World Congress on Physics Medical and Biomedical Engineering.*, **33a**, 409-409.

Smith, A. F. M., and Roberts, G. O. (1993) Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society B*, **55**, 3-24.

Sutranghar, B. C., and Ali, M. M. (1986). Estimation of the parameters of a regression model with a multivariate T error variable. *Commun. Statist.-Theor. Meth.,* **15**, 429-450.

Tsurumi, H. (1977). A Bayesian test of a parameter shift and an application. *Journal of Econometrics*, **6**, 371-380.

Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms. *J. Amer. Statist. Assoc.*, **71**, 400-405.

Carlos Alberto R. Diniz
and **Luis A. Milan**
Departamento de Estatística,
Universidade Federal de São Carlos
São Carlos, SP, Brazil
dcad@power.ufscar.br
dlam@power.ufscar.br

Josmar Mazucheli
Departamento de Estatística,
Universidade Estadual de Maringá
Maringá, PR, Brazil
jmazucheli@uem.br