Linear and Non-Linear Regression Models Assuming a Stable Distribution

Modelos de Regressión Lineal y No lineal Suponiendo una Distribución Estable

Jorge A. Achcar[a]* and Sílvia R.C. Lopes[b]

(a) Medical School - USP, Ribeirão Preto, SP, Brazil
(b) Mathematics Institute - UFRGS, Porto Alegre, RS, Brazil

March 23, 2015

Abstract

In this paper, we present some computational aspects for a Bayesian analysis involving stable distributions. It is well known that, in general, there is no closed form for the probability density function of a stable distribution. However, the use of a latent or auxiliary random variable facilitates us to obtain any posterior distribution when related to stable distributions. To show the usefulness of the computational aspects, the methodology is applied to linear and non-linear regression models. Posterior summaries of interest are obtained using the OpenBUGS software.

Key words: Stable laws, Bayesian analysis, MCMC methods, OpenBUGS software.

Resumen

En este trabajo, presentamos algunos aspectos computacionales de análisis bayesiano con distribuciones estables. Es bien sabido que, en general, no hay forma cerrada para la función de densidad de probabilidad de distribuciones estables. Sin embargo, el uso de una variable aleatoria latente facilita obtener la distribución à posteriori. La metodologia se aplica a regresión lineal y non lineal utilizando el software OpenBUGS.

Palabras clave: Leyes Estable, análisis bayesiano, métodos MCMC, software Open-BUGS.

1 Introduction

A wide class of distributions that encompasses the Gaussian one is given by the class of stable distributions. This large class defines location-scale families that are closed under

---

*Corresponding author: achcar@fmrp.usp.br

convolution. This distribution family (see for instance Samorodnitsky and Taqqu, 1994) is described by four parameters α, β, δ and γ. The $\alpha \in (0, 2]$ parameter defines the "fatness of the tails", and when $\alpha = 2$ this class reduces to Gaussian distributions. The $\beta \in [-1, 1]$ is the skewness parameter and for $\beta = 0$ one has symmetric distributions. The location and scale parameters are, respectively, $\delta \in (-\infty, \infty)$ and $\gamma \in (0, \infty)$ (see Lévy, 1924).

Stable distributions are usually denoted by $S_\alpha(\beta, \delta, \gamma)$. If a random variable $X \sim S_\alpha(\beta, \delta, \gamma)$, then $Z = \frac{X-\delta}{\gamma} \sim S_\alpha(\beta, 0, 1)$, whenever $\alpha \neq 1$ (see Lukacs, 1970 and Nolan, 2015).

The difficulty associated to stable distributions $S_\alpha(\beta, \delta, \gamma)$ is that in general there is no simple closed form for their probability density functions. However, it is known the probability density functions of stable distributions are continuous (Gnedenko and Kolmogorov, 1968; Skorohod, 1961) and unimodal (Ibragimov and Černin, 1959; Kanter, 1976). Also the support of all stable distributions is given in $(-\infty, \infty)$, except for $\alpha < 1$ and $|\beta| = 1$ when the support is $[\delta, \infty)$ for $\beta = 1$ and $(-\infty, \delta]$ for $\beta = -1$ (see page 12 in Nolan, 2015). The characteristic function $\varphi(\cdot)$ of a stable distribution is given by

$$\log(\varphi(t)) = \begin{cases} i\delta t - \gamma^\alpha |t|^\alpha [1 + i\beta \, sign(t) \tan(\frac{\pi\alpha}{2})(|\gamma \, t|^{1-\alpha} - 1)], & \text{for } \alpha \neq 1 \\ i\delta t - \gamma |t|[1 + i\beta \frac{2}{\pi} sign(t) \log(\gamma |t|)], & \text{for } \alpha = 1, \end{cases} \tag{1}$$

where $i = \sqrt{-1}$ and the $sign(\cdot)$ function is given by

$$sign(x) = \begin{cases} -1, & \text{if } x < 0 \\ 0, & \text{if } x = 0 \\ 1, & \text{if } x > 0. \end{cases} \tag{2}$$

Although this is a good class for data modeling in different areas, one has difficulties to obtain parameter estimates under a classical inference approach due to the lack of closed form expressions for the probability density functions. An alternative is the use of Bayesian methods. However, the computational cost can be further exacerbated in assessing posterior summaries of interest.

A Bayesian analysis of stable distributions is introduced by Buckle (1995) using Markov Chain Monte Carlo (MCMC) methods (see also Achcar et al., 2013a). The use of Bayesian methods with MCMC simulation can have great flexibility by considering latent variables (see, for instance, Damien et al., 1999 or Tanner and Wong, 1987), where samples of latent variables are simulated in each step of the Gibbs or Metropolis-Hastings algorithms.

Considering a latent or an auxiliary variable, Buckle (1995) proved a theorem that is useful to simulate samples from the joint posterior distribution of the parameters α, β, δ and γ. This theorem establishes that a stable distribution for a random variable Z defined in $\mathbb{R} - \{0\}$ is obtained as the marginal of a bivariate distribution for the random variable Z itself and an auxiliary random variable Y. This variable Y is defined in the interval $(-0.5, a_{\alpha,\beta})$, when $Z \in (-\infty, 0)$, and in $(a_{\alpha,\beta}, 0.5)$, when $Z \in (0, \infty)$. The quantity $a_{\alpha,\beta}$ is given by

$$a_{\alpha,\beta} = -\frac{b_{\alpha,\beta}}{\alpha \pi} \tag{3}$$

where $b_{\alpha,\beta} = \beta \frac{\pi}{2} \min\{\alpha, 2 - \alpha\}$.

The joint probability density function for random variables Z and Y is given by

$$f(z, y | \alpha, \beta) = \frac{\alpha}{|\alpha - 1|} \exp \left\{ - \left| \frac{z}{t_{\alpha,\beta}(y)} \right|^{\theta} \right\} \left| \frac{z}{t_{\alpha,\beta}(y)} \right|^{\theta} \frac{1}{|z|}, \tag{4}$$

where $\theta = \frac{\alpha}{\alpha - 1}$,

$$t_{\alpha,\beta}(y) = \left(\frac{\sin(\pi \alpha y + b_{\alpha,\beta})}{\cos(\pi y)} \right) \left(\frac{\cos(\pi y)}{\cos(\pi(\alpha - 1)y + b_{\alpha,\beta})} \right)^{\frac{1}{\theta}} \tag{5}$$

and $Z = \frac{X - \delta}{\gamma}$, for $\gamma \neq 0$ and $\alpha \neq 1$.

From the bivariate density (4), Buckle (1995) shows that the marginal distribution for the random variable Z is $S_{\alpha}(\beta, 0, 1)$. Usually, the computational costs to obtain posterior summaries of interest using MCMC methods is high for this class of models, which could give some limitations for practical applications. Another problem can be the simulation algorithm convergence. In this paper, we propose the use of a popular free available software to obtain the posterior summaries of interest: the OpenBUGS software.

The paper is organized as follows: in Section 2, we present a Bayesian analysis for stable distributions; in Section 3 we introduce the use of stable distributions for linear regression models while non-linear regression models are considered in Section 4; in Section 5, we introduce some numerical illustrations; finally, in Section 6, we present some concluding remarks.

2   A Bayesian Analysis for General Stable Distributions

Let us assume that x_i, for $i = 1, \cdots, n$, is a random sample of size n, where $X_i \sim S_{\alpha}(\beta, \delta, \gamma)$, that is, $Z_i = \frac{X_i - \delta}{\gamma} \sim S_{\alpha}(\beta, 0, 1)$. Assuming a joint prior distribution for α, β, δ and γ, given by $\pi_0(\alpha, \beta, \delta, \gamma)$, Buckle (1995) shows that the joint posterior distribution for the parameters α, β, δ and γ is given by

$$\begin{aligned} \pi(\alpha, \beta, \delta, \gamma | \mathbf{x}) &\propto \int \left(\frac{\alpha}{|\alpha - 1| \gamma} \right)^n \times \exp \left\{ - \sum_{i=1}^{n} \left| \frac{z_i}{t_{\alpha,\beta}(y_i)} \right|^{\theta} \right\} \prod_{i=1}^{n} \left| \frac{z_i}{t_{\alpha,\beta}(y_i)} \right|^{\theta} \frac{1}{|z_i|} \\ &\times \pi_0(\alpha, \beta, \delta, \gamma) \, \mathbf{dy}, \end{aligned} \tag{6}$$

where $\theta = \frac{\alpha}{\alpha - 1}$, $z_i = \frac{x_i - \delta}{\gamma}$, for $i = 1, \cdots, n$, $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\delta \in (-\infty, \infty)$ and $\gamma \in (0, \infty)$; $\mathbf{x} = (x_1, x_2, \cdots, x_n)'$ and $\mathbf{y} = (y_1, y_2, \cdots, y_n)'$ are respectively, the observed and non-observed data vectors. Notice that the multivariate distribution in expression (6) is given in terms of x_i and the latent variables y_i, and not in terms of z_i and y_i (there is the Jacobian γ^{-1} multiplied by the right-hand-side of expression (4)).

Observe that when $\alpha = 2$ we have $\theta = 2$ and $b_{\alpha,\beta} = 0$. In this case, we have a Gaussian distribution with mean δ and variance $2\gamma^2$.

For a Bayesian analysis of the proposed model, we assume uniform $\mathcal{U}(a, b)$ independent priors for α, β, δ and γ, where the hyperparameters a and b are assumed to be known in each application following the restrictions $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\delta \in (-\infty, \infty)$ and $\gamma \in (0, \infty)$.

In the simulation algorithm to obtain a Gibbs sample for the random quantities α, β, δ and γ having the joint posterior distribution (6), we assume a uniform $\mathcal{U}(-0.5, 0.5)$ prior

distribution for the latent random quantities Y_i, for $i = 1, \cdots, n$. With this choice of priors, one has the possibility to use standard software package like OpenBUGS (see Spiegelhalter et al., 2003) with great simplification to obtain the simulated Gibbs samples for the joint posterior distribution. In this way, one has the following algorithm:

(i) Start with the initial values $\alpha^{(0)}, \beta^{(0)}, \delta^{(0)}, \gamma^{(0)}$;

(ii) Simulate a sample $\mathbf{y} = (y_1, \cdots, y_n)'$ from the conditional distributions $\pi(y_i | \alpha^{(0)}, \beta^{(0)}, \delta^{(0)}, \gamma^{(0)}, \mathbf{x})$, for $i = 1, \cdots, n$; that is, we simulate these latent random variables $\mathbf{y}$ similarly as we simulate the parameters;

(iii) Update $\alpha^{(0)}, \beta^{(0)}, \delta^{(0)}, \gamma^{(0)}$ by $\alpha^{(1)}, \beta^{(1)}, \delta^{(1)}, \gamma^{(1)}$ from the conditional distributions $\pi(\alpha | \beta^{(0)}, \delta^{(0)}, \gamma^{(0)}, \mathbf{x}, \mathbf{y})$, $\pi(\beta \mid \alpha^{(0)}, \delta^{(0)}, \gamma^{(0)}, \mathbf{x}, \mathbf{y})$, $\pi(\delta | \alpha^{(0)}, \beta^{(0)}, \gamma^{(0)}, \mathbf{x}, \mathbf{y})$ and $\pi(\gamma | \alpha^{(0)}, \beta^{(0)}, \delta^{(0)}, \mathbf{x}, \mathbf{y})$;

(iv) Repeat steps (i), (ii) and (iii) until convergence.

From expression (6), the joint posterior probability distribution for $\alpha, \beta, \delta, \gamma$ and $\mathbf{y} = (y_1, y_2, \cdots, y_n)'$ is given by

$$
\begin{aligned}
\pi(\alpha, \beta, \delta, \gamma, \mathbf{y} | \mathbf{x}) \quad \propto \quad & \left(\frac{\alpha}{|\alpha - 1| \gamma} \right)^n \exp \left\{ -\sum_{i=1}^n \left| \frac{z_i}{t_{\alpha,\beta}(y_i)} \right| \right\} \\
\times \quad & \prod_{i=1}^n \left| \frac{z_i}{t_{\alpha,\beta}(y_i)} \right|^\theta \frac{1}{|z_i|} \times \prod_{i=1}^n h(y_i) \pi_0(\alpha, \beta, \delta, \gamma), \quad (7)
\end{aligned}
$$

where θ and $t_{\alpha,\beta}(\cdot)$ are respectively defined in (4) and (5) and $h(\mathbf{y})$ is a $\mathcal{U}(-0.5, 0.5)$ density function.

Since we are using the OpenBUGS software to simulate samples from the joint posterior distribution we do not present here all full conditional distributions needed for the Gibbs sampling algorithm. This software only requires the data distribution and prior distributions for the interested random quantities. This gives great computational simplification for determining posterior summaries of interest as shown in the applications in Section 6.

3 Linear regression models assuming stable distributions

Consider a random variable X related to a controlled variable V given by the linear relationship

$$
x_i = d_0 + d_1 v_i + \varepsilon_i, \quad \text{for} \quad i = 1, 2, \cdots, n, \quad (8)
$$

where

- the random variable X_i represents the response for the i-th unit associated with an experimental value of the independent or explanatory variable v assumed as a fixed value (a common regression model assumption). In this way, x_i is an observation of X_i;

- the variables $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ are considered as components of unknown errors and are unobserved random variables. Assume that these random variables ε_i, for $i = 1, 2, \cdots, n$, are independent and identically distributed with normal distribution $\mathcal{N}(0, \sigma_\varepsilon^2)$;

- the parameters d_0 and d_1 are unknown.

From the above assumptions, we have normality for the responses, that is,

$$X_i \sim \mathcal{N}(d_0 + d_1 v_i; \sigma_\varepsilon^2). \tag{9}$$

In this way X_i has a normal distribution with mean $d_0 + d_1 v_i$ and common variance σ_ε^2. Usually we get estimators for the regression parameters using least squares approach or standard maximum likelihood methods (see, for instance, Draper and Smith, 1981 or Seber and Lee, 2003).

Standard generalization for the linear model (8) is given in the presence of k independent or explanatory variables, that is, a multiple linear regression model given by

$$x_i = d_0 + d_1 v_{i1} + d_2 v_{i2} + \cdots + d_k v_{ik} + \varepsilon_i. \tag{10}$$

From the normality assumption for the error ε_i in (10), the random variable X_i has normal distribution with mean $d_0 + d_1 v_{i1} + d_2 v_{i2} + \cdots + d_k v_{ik}$ and variance σ_ε^2.

In practical applications, we need to check if the above assumptions are satisfied. In this way, we consider graphical approaches to verify if the model residuals satisfy the above assumptions.

In the presence of outliers or discordant observations, we could have large impact on the obtained estimators for the regression model given by (10) which could invalidate the obtained inferences. In this situation, we could use non-parametric regression models or to assume more robust probability distributions for the data. One possibility is to assume that the random variable X in (10) or (8) has a stable distribution $S_\alpha(\beta, \delta, \gamma)$.

In this way, we assume the response x_i in the linear regression model (10), for $i = 1, \cdots, n$, has a stable distribution $X_i \sim S_\alpha(\beta, \delta_i, \gamma)$, that is, $Z_i = \frac{X - \delta_i}{\gamma} \sim S_\alpha(\beta, 0, 1)$ and where the location parameter δ_i of the stable distribution is related to the explanatory variables by a linear relation given by,

$$\delta_i = d_0 + d_1 v_{i1} + d_2 v_{i2} + \cdots + d_k v_{ik}. \tag{11}$$

Assuming a joint prior distribution for $\alpha, \beta, \mathbf{d}$ and γ, where $\mathbf{d} = (d_0, d_1, d_2, \cdots, d_k)'$ given by $\pi_0(\alpha, \beta, \mathbf{d}, \gamma)$, Buckle (1995) or Achcar et al. (2013b), show that the joint posterior distribution for parameters α, β, $\mathbf{d}$ and γ, is given by,

$$\pi(\alpha, \beta, \mathbf{d}, \gamma | \mathbf{x}) \propto \int \left(\frac{\alpha}{|\alpha - 1|\gamma} \right)^n \exp\left\{ \sum_{i=1}^n \left| \frac{z_i}{t_{\alpha,\beta}(y_i)} \right|^\theta \right\} \prod_{i=1}^n \left| \frac{z_i}{t_{\alpha,\beta}(y_i)} \right|^\theta \frac{1}{|z_i|} \pi_0(\alpha, \beta, \mathbf{d}, \gamma) \mathbf{dy}, \tag{12}$$

where $\theta = \frac{\alpha}{\alpha-1}$, $z_i = \frac{x_i - y_i}{\gamma}$, for $i = 1, \cdots, n$, $\alpha \in (0, 2]$, $\beta \in [-1, 1]$ and $\gamma \in (0, \infty)$; $\mathbf{x} = (x_1, x_2, \cdots, x_n)'$ and $\mathbf{y} = (y_1, y_2, \cdots, y_n)'$ are respectively, the observed and non-observed data vectors.

4 Non-linear growth regression models assuming stable distributions

Growth curves are included in a class of nonlinear models widely used in biology to model different problems as the population size or biomass (in population ecology and demography, for population growth analysis) or the individual body height or biomass (in physiology, for growth analysis of individuals). A growth curve is an empirical model of the evolution of a quantity over time. Growth curves are employed in many disciplines besides biology, particularly in statistics, where there is a large literature on this subject related to nonlinear models. Under a more probabilistic and mathematical statistics approach, growth curves are often modeled as being a continuous stochastic process.

Standard classical inference methods to obtain point or interval estimates for the parameters of growth curves are presented within the nonlinear modeling methodology.

Nonlinear regression methodology is similar to the linear regression methodology, that is, a modeling approach to relate a response X to a vector of covariates, $\mathbf{v} = (v_1, \cdots, v_k)'$, where $\mathbf{v}'$ denotes the transpose of the vector $\mathbf{v}$. Different from linear models, nonlinear regression is characterized by the fact that the prediction equation depends nonlinearly on one or more unknown parameters. Different from the linear regression methodology often used for building a purely empirical model, nonlinear regression methodology usually arises when there is some physical reason which implies that the relationship between the response and the predictors follows a particular functional form.

A nonlinear regression model has the general form,

$$X_i = f(\mathbf{v_i}, \boldsymbol{\theta}) + \varepsilon_i, \tag{13}$$

where X_i are the responses, for $i = 1, \cdots, n$; f is a known function of the covariate vector; $\mathbf{v_i} = (v_{i1}, \cdots, v_{ik})'$ is a vector of k covariates or independent variables; $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)'$ is the vector of p parameters and ε_i are random errors. The errors ε_i are usually assumed to be uncorrelated and normally distributed with mean zero and constant variance.

The most popular criterion to estimate the p parameter vector $\boldsymbol{\theta}$ in the nonlinear model (13) is to find estimates for the parameters (via nonlinear least squares) which minimize the sum of squared errors, given by,

$$\sum_{i=1}^{n} (x_i - f(\mathbf{v_i}, \boldsymbol{\theta}))^2. \tag{14}$$

Remark 1. *If the errors ε_i, for $i \in \{1, 2, \cdots, n\}$, follows a normal distribution, then the least squares estimator for $\boldsymbol{\theta}$ is also the maximum likelihood estimator.*

Usually, nonlinear regression estimates must be computed by iterative procedures using optimization methods to minimize the sum of squared errors given by the expression (14). It is important to point out that the definition of nonlinearity is related to the unknown parameters and not to the relationship between the covariates and the response. As an example, $X = \beta_0 + \beta_1 v + \beta_2 v^2 + \varepsilon$ is considered as a linear model (see, for instance, Bates and Watts, 1988; Ratkowsky, 1983; Seber and Wild, 1989).

A popular iterative technique to find the least squares estimator of nonlinear models is the Gauss-Newton algorithm. The Gauss-Newton algorithm increments the working estimate $\widehat{\boldsymbol{\theta}}$ at each iteration by an amount equal to the coefficients from the linear regression

of the current residuals u_i, defined as $u_i = x_i - f(\mathbf{v_i}, \widehat{\boldsymbol{\theta}})$, on the current gradient matrix X.

If the function $f(\cdot)$ in (13) is continuously differentiable in $\boldsymbol{\theta}$, then it can be linearized locally as

$$f(\mathbf{v}, \boldsymbol{\theta}) = f(\mathbf{v}, \boldsymbol{\theta_0}) + \mathbf{V_0}(\boldsymbol{\theta} - \boldsymbol{\theta_0}), \tag{15}$$

where $\mathbf{V_0}$ is the $n \times p$ gradient matrix with elements $\partial f(\mathbf{v_i}, \boldsymbol{\theta_0})/\partial \theta_j$ and $\boldsymbol{\theta_0}$ is a vector of initial values for the iterative procedure. This leads to the Gauss-Newton algorithm for estimating $\boldsymbol{\theta}$,

$$\boldsymbol{\theta_1} = \boldsymbol{\theta_0} + (\mathbf{V_0}'\mathbf{V_0})^{-1}\mathbf{V_0}'\mathbf{e}, \tag{16}$$

where $\mathbf{e}$ is the vector of working residuals $e_i = x_i - f(\mathbf{v_i}, \boldsymbol{\theta_0})$, for $i = 1, 2, \cdots, n$.

If the errors ε_i are independent and normally distributed $\mathcal{N}(0, \sigma_\varepsilon^2)$, then the Gauss-Newton algorithm is an application of Fisher's method of scoring. This algorithm is implemented in many existing statistical softwares as, for instance, R, Minitab (version 16) or SAS.

If $\mathbf{X}$ is of full column rank in a neighborhood of the least squares solution, then it can be shown that the Gauss-Newton algorithm converges to the solution from a sufficiently good starting value. In practical applications, there is no guarantee, though, that the algorithm will converge from values further from the solution. Some improvement of the Gauss-Newton algorithm are given in the literature as the Levenberg-Marquart damping algorithm (see, for instance, Seber and Wild, 1989).

Standard inferences for the parameters of nonlinear models are obtained from the asymptotical normality of the least squares estimators $\widehat{\boldsymbol{\theta}}$ with mean $\boldsymbol{\theta}$ and variance-covariance matrix $\sigma_\varepsilon^2(\mathbf{V}'\mathbf{V})^{-1}$, where the variance σ_ε^2 is usually estimated by,

$$s^2 = \frac{1}{n-p} \sum_{i=1}^{n} [x_i - f(\mathbf{v_i}, \widehat{\boldsymbol{\theta}})]^2. \tag{17}$$

It is important to point out that since most asymptotic inference for nonlinear regression models is based on the analogy to linear models, and since this inference is only approximated as the actual model differs from a linear model, various measures of nonlinearity have been proposed in the literature to verify how good linear approximations are likely to be in each case. One class of measures focuses on curvature (intrinsic curvatures) of the function $f(\cdot)$ and it is based on the size of the second derivatives of f (see, for instance, Bates and Watts, 1980). Another class of measures is the intrinsic curvature defined by the residuals $u_i = x_i - f(\mathbf{v_i}, \widehat{\boldsymbol{\theta}})$ or parameters effect curvatures (see, for instance, Bates and Watts, 1988).

In many applications, the systematic part of the response is known to be monotonic increasing in v, where v might represent time or dosage. Nonlinear regression models with this property are called growth models. The simplest growth model is the exponential growth model defined as

$$f(v, \boldsymbol{\theta}) = \theta_1 \exp(-\theta_2 v), \tag{18}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)'$.

There are several growth models introduced in the literature. In this paper we consider the following special cases

1. $f(v, \boldsymbol{\theta}) = \theta_1 + \theta_2 v^{\theta_3}$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$.

2. $f(v, \boldsymbol{\theta}) = \theta_1 + \theta_2 \exp(\theta_3 v)$ (*exponential growth model*), where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$.

3. $f(v, \boldsymbol{\theta}) = \theta_1 \exp(-\theta_2 v) + \theta_3 \exp(-\theta_4 v)$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)'$.

4. $f(v, \boldsymbol{\theta}) = \theta_1 v / (\theta_2 + v)$ (*Michaelis-Menten*), where $\boldsymbol{\theta} = (\theta_1, \theta_2)'$.

5. $f(v, \boldsymbol{\theta}) = \theta_1 \exp(-\exp(\theta_2 - \theta_3 v))$ (*Gompertz*), where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$.

6. $f(v, \boldsymbol{\theta}) = \theta_1 + (\theta_2 - \theta_1)/[1 + \exp((v - \theta_3)/\theta_4)]$ (*logistic*), where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)'$.

7. $f(v, \boldsymbol{\theta}) = \theta_1 + (\theta_2 - \theta_1) \exp(\theta_3 v^{\theta_4})$ (*Weibull*), where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)'$.

8. $f(v, \boldsymbol{\theta}) = 1 - \exp[-\exp(\theta_1 - \theta_2 v)]$ (*sigmoid*), where $\boldsymbol{\theta} = (\theta_1, \theta_2)'$.

Observe that these different models could be considered to model growth curves. Very often we have physical interpretations for the use of a particular model. In others we try different growth models to decide which one best fits the data.

In many situations the usual normality assumption for the errors in (13) will not be appropriate. For instance, this can be the case when we have discordant observations which could have an impact on the obtained inferences. In this way, we could assume more robust distributions for the data, as the stable distribution.

We point out that the use of asymptotical inference results could not be accurate depending on the sample sizes and on the intrinsic curvature of the function $f(\cdot)$ in (13). In those cases, we propose the use of stable distributions for the response X_i in (13).

Let us assume that the response x_i, for $i = 1, \cdots, n$, in the nonlinear regression model (13) has a stable distribution $X_i \sim S_\alpha(\beta, \delta_i, \gamma)$, that is, $Z_i = \frac{X - \delta_i}{\gamma} \sim S_\alpha(\beta, 0, 1)$ and where the location parameter δ_i of the stable distribution is related to the explanatory variables by a nonlinear relation given by,

$$\delta_i = f(\mathbf{v_i}, \mathbf{d}). \tag{19}$$

Assuming a joint prior distribution for α, β, $\mathbf{d}$ and γ, where $\mathbf{d} = (d_0, d_1, d_2, \cdots, d_k)'$ given by $\pi_0(\alpha, \beta, \mathbf{d}, \gamma)$, Buckle (1995) shows that the joint posterior distribution for parameters α, β, $\mathbf{d}$ and γ, is given by the general form (12).

5   Applications

5.1   An example of a simple linear regression model

In Table 5.1, we have a data set related to an industrial experiment, where x denotes the response and v denotes an explanatory variable associated to each response, for $n = 15$ observations (see Johnson and Bhattacharyya, 1980).

Table 5.1: An industrial experiment data set.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x	1.2	1.5	1.5	3.3	2.5	2.1	2.5	3.2	2.8	1.5	2.2	2.2	1.9	1.8	2.8
v	19	15	35	52	35	33	30	57	49	26	45	39	25	40	40

From a preliminary data analysis, we see that a linear regression model (8) is suitable for this data set. The estimated regression straight line obtained by minimum squares estimates and using the software MINITAB (version 16) is given by $\widehat{x}_i = 0.603 + 0.0444\,v_i$, where the regression parameter d_1 is statistically different from zero (p-value < 0.05). From standard residuals graphs we verify that the required assumptions (residuals normality and homoscedastic variance are satisfied).

Under a Bayesian approach, we have in Table 5.2, the posterior summaries of interest assuming the linear regression model defined by (8) with a stable distribution for the response. Using the OpenBugs software we assume the following prior distributions: $\alpha \sim \mathcal{U}(0,2)$, $\beta \sim \mathcal{U}(-1,0)$, $d_0 \sim \mathcal{N}(0,1)$, $d_1 \sim \mathcal{N}(0,1)$ and $\gamma \sim \mathcal{U}(0,3)$. Observe that we are assuming approximately non-informative priors for the parameters of the model. This procedure will be used in the other Bayesian analysis. We further assume independence among the random quantities. We also assume a uniform $\mathcal{U}(-0.5, 0.5)$ distribution for the latent variable Y_i, for $i = 1, 2, \cdots, 15$. We simulated 800,000 Gibbs samples, with a "burn-in-sample" of 300,000 samples discarded to eliminate the effects of the initial values in the iterative simulation procedure. We take a final sample of size 1,000 (every 500th sample chosen from the 500,000 samples). Gibbs sampling algorithm convergence was monitored from standard trace plots of the simulated samples.

In Table 5.2, we also have the sum of absolute values (SAV) for the differences between the observed and fitted mean values, given by,

$$\text{SAV} = \sum_{i=1}^{n} \text{absolute}[\text{observed}(i) - \text{fitted mean}(i)]. \tag{20}$$

In Table 5.2, we also have the posterior summaries of the regression model (8) assuming a normal $\mathcal{N}(0, \sigma_\varepsilon^2)$ distribution for the error and the following priors for the parameters of the model: $d_0 \sim \mathcal{N}(0,1)$, $d_1 \sim \mathcal{N}(0,1)$ and $\zeta = 1/\sigma_\varepsilon^2 \sim \mathcal{U}(0,3)$. In this case, we simulated 55,000 Gibbs samples taking a "burn-in-sample" of size 5,000 using the OpenBUGS software. We take a final sample of size 1,000 (every 50th sample chosen from the 50,000 samples). From the results of Table 5.2, we observe similar results assuming both normality or stable distribution for the data. In this case, we conclude that we do not need to assume a stable distribution for the data, since the results are very similar to the results obtained from the normality assumption for the errors. Besides, the computational cost using stable distributions is very high (see Achcar et al., 2013b).

In Figure 5.1(a) we have the plots of observed and fitted means considering both models versus samples. From the graphs of Figure 5.1(a), we observe similar fit for both models (linear regression model assuming both normality and stable distributions). Observe that we have SAV $= 4.496$ when we assume a stable distribution while SAV $= 4.504$ when we assume a normal distribution. Hence, we have very close results.

Now let us consider the presence of an outlier or discordant response (considered as a measurement error) replacing the 15th response (equal to 2.8) in Table 5.1 by the value 8.0. In Table 5.3, we have the obtained posterior summaries assuming the same priors and the same simulation procedure as in Table 5.2. In Figure 5.1(b), we have the plots of observed and fitted means considering both models versus samples in the presence of one outlier. From this figure, we observe that the model with a stable distribution is very robust to the presence of the outlier given similar inference results as obtained without its presence (see results in Table 5.2). Table 5.3 also reports the estimated regression

Table 5.2: Posterior summaries (linear regression).

Stable distribution (SAV= 4.496)			
Parameter	Mean	Standard Deviation	95% Credible Interval
α	1.7160	0.2123	$(1.2380, 1.9890)$
β	-0.6183	0.2391	$(-0.9881, -0.0719)$
d_0	0.5431	0.3559	$(-0.1981, 1.2270)$
d_1	0.0460	0.0093	$(0.0283, 0.0652)$
γ	0.2931	0.0638	$(0.1960, 0.446)$
Normal distribution (SAV= 4.504)			
Parameter	Mean	Standard Deviation	95% Credible Interval
d_0	0.4802	0.4927	$(-0.56587, 1.4230)$
d_1	0.0475	0.0131	$(0.0213, 0.0737)$
ζ	0.4949	0.1709	$(0.2119, 0.8834)$

parameters with normal error and we observe how strongly the results are affected by the presence of this outlier. Observe that we have SAV = 4.623 assuming a stable distribution (a value very close to the SAV values given in Table 5.2, without the presence of an outlier) and SAV = 6.394 assuming a normal distribution.
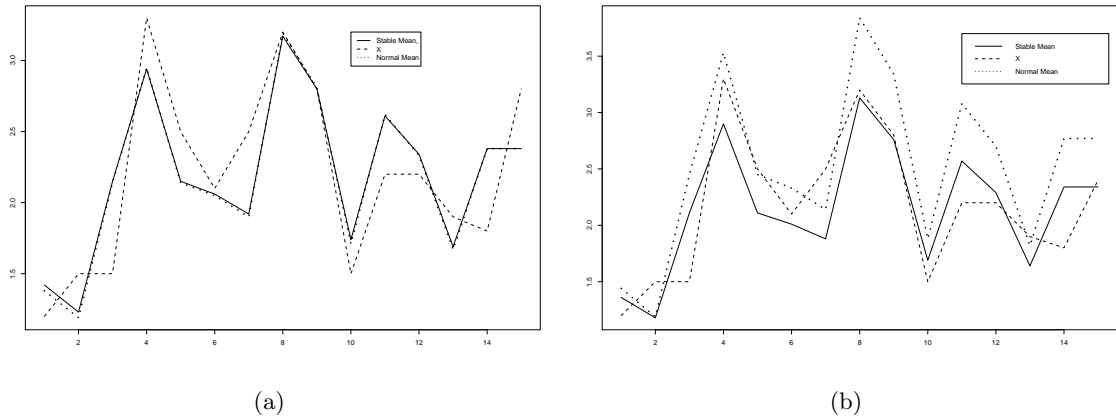

Figure 5.1: Observed and fitted means values, considering both models: (a) samples; (b) presence of one outlier.

5.2   An example with a growth non-linear model

In Table 5.4, we have a data set, reported by Bache et al. (1972), related to the concentration of polychlorinated biphenyl (PCB) residues in a series of lake trout from Cayuga Lake, NY, USA. The ages of the fishes were accurately known, since they were annually stocked as yearlings and distinctly marked as to the year class. Every fish was mechanically chopped, ground, thoroughly mixed, and 5-gram samples were taken. The samples were treated and PCB residues in parts per million (ppm) were estimated using column chromatography.

Table 5.3: Posterior summaries (presence of one outlier).

Stable distribution (SAV = 4.623)			
Parameter	Mean	Standard Deviation	95% Credible Interval
α	1.3730	0.2195	$(1.0580, 1.8430)$
β	-0.5726	0.2703	$(-0.9821, -0.0426)$
d_0	0.4797	0.3461	$(-0.3483, 1.0400)$
d_1	0.0465	0.0096	$(0.0292, 0.0701)$
γ	0.3221	0.0802	$(0.2004, 0.4856)$
Normal distribution (SAV = 6.394)			
Parameter	Mean	Standard Deviation	95% Credible Interval
d_0	0.2408	0.7866	$(-1.3100, 1.8200)$
d_1	0.0632	0.0225	$(0.0201, 0.1061)$
ζ	0.4949	0.1709	$(0.2119, 0.8834)$

Table 5.4: Lake trout data set.

Row	1	2	3	4	5	6	7	8	9	10
Age	1	1	1	1	2	2	2	3	3	3
PCB	0.6	1.6	0.5	1.2	2.0	1.3	2.5	2.2	2.4	1.2
log(PCB)	-0.5118	0.4700	-0.6932	0.1823	0.6932	0.2624	0.9163	0.7885	0.8755	0.1823
Row	11	12	13	14	15	16	16	18	19	20
Age	4	4	4	5	6	6	6	7	7	7
PCB	3.5	4.1	5.1	5.7	3.4	9.7	8.6	4.0	5.5	10.7
log(PCB)	1.2528	1.4110	1.6229	1.7405	1.2238	2.2721	2.1518	1.3863	1.7048	2.3514
Row	21	22	23	24	25	26	27	28		
Age	8	8	8	9	11	12	12	12		
PCB	17.5	13.4	4.5	30.4	12.4	13.4	26.2	7.4		
log(PCB)	2.8622	2.5953	1.5041	3.4144	2.5177	2.5953	3.2658	2.0015		

From a preliminary data analysis, we see that a nonlinear regression model (13) is suitable for the data set both in the original and transformed scales (see Figure 5.2).

We assume standard classical approach for nonlinear models considering the eigth growth models introduced in Section 4. The responses are in the logarithmic scale log(PCB), and we use the software MINITAB (version 16). Table 5.5 presents the estimation results for each assumed growth model and we observe that for some growth models we need a large number of iterations for the Gauss-Newton algorithm convergence.
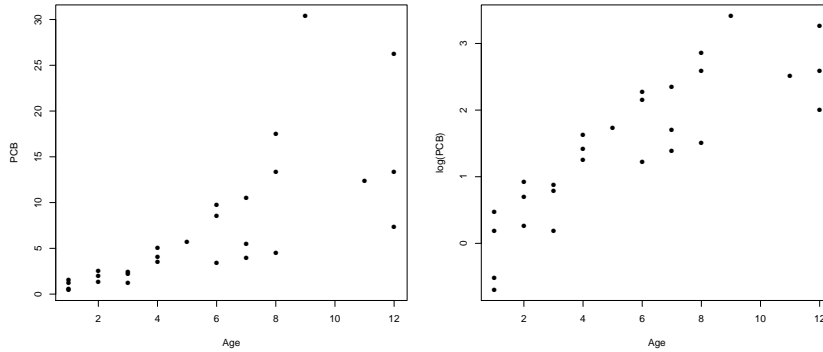
Figure 5.2: Observed values against age for PCB and log(PCB).

In Figure 5.3, we have the plots of observed log(PCB) and fitted models versus observations. From these plots, we observe that the most of the assumed models give reasonable fit for the data (models 1, 2, 4, 5, 6 and 7 given in Table 5.5). Only models 3 and 8 are not suitable for the data. It is important to observe that the sum of absolute values for the differences (observed - estimated mean) for models 1, 2, 4, 5, 6 and 7 are given, respectively, by 11.4412, 11.1411, 12.0196, 11.0047, 11.0209 and 11.0310. That is, those values are very similar to each other, showing an indication of similar fit. Considering models 3 and 8, the values for these differences are given, respectively by, 15.5568 and 23.4592 (see also Figure 5.3).

Observe that the obtained classical estimates given in Table 5.5, could be very unstable heavily depending on the initial values used in the iterative algorithm. Usually it is difficult to get empirical good initial values to be used in the iterative procedure. We also point out that the standard errors (given in parenthesis in the right panel of Table 5.5) of the obtained estimates could be very large as it is observed in this table.

Under a Bayesian approach, we have in Table 5.6, the posterior summaries of interest considering the eight growth models 1-8 introduced in Section 4 with a normal distribution for the error (see expression (13)) and the OpenBUGS software assuming the following prior distributions: $\theta_j \sim \mathcal{U}(0,1)$, $j = 1, 2, 3, 4$ and $\zeta = 1/\sigma^2 \sim \mathcal{G}(1,1)$, where $\mathcal{G}(a,b)$ denotes a gamma distribution with mean a/b and variance a/b^2. We simulated 10,000 Gibbs samples, with a "burn-in-sample" of 1,000 samples discarded to eliminate the effects of the initial values in the iterative simulation procedure and taking a final sample of size 900 (every 10th sample choosen from the 9,000 samples). Gibbs sampling algorithm convergence was monitored from standard trace plots of the simulated samples.

From the results of Table 5.6, we observe that the posterior standard deviation for each estimated parameter of the growth model is very small. It is important to point

Table 5.5: Classical estimates for the growth models.

Fitted model	Iteration	Estimate (Standard Error)			
		θ_1	θ_2	θ_3	θ_4
(1) $\log(\text{PCB}) = \theta_1 + \theta_2 \text{Age}^{\theta_3}$	11	-4.8664 (8.4290)	4.7033 (8.2768)	0.1968 (0.2739)	-
(2) $\log(\text{PCB}) = \theta_1 + \theta_2 \exp[\theta_3 \text{Age}]$	17	3.1293 (0.5785)	-3.9292 (0.4513)	-0.1903 (0.0744)	-
(3) $\log(\text{PCB}) = \theta_1 \exp[-\theta_2 \text{Age}] + \theta_3 \exp[-\theta_4 \text{Age}]$	200	12.9631 (10164.1)	0.0304 (7.1)	-12.7809 (10163.5)	0.0495 (8.1)
(4) $\log(\text{PCB}) = \theta_1 \text{Age}/[\theta_2 + \text{Age}]$	12	8.6431 (5.3952)	23.9661 (20.0487)	-	-
(5) $\log(\text{PCB}) = \theta_1 \times \exp[-\exp(\theta_2 - \theta_3 \text{Age})]$	14	2.7062 (0.3003)	1.5332 (0.4300)	0.4354 (0.1358)	-
(6) $\log(\text{PCB}) = \theta_1 + (\theta_2 - \theta_1)/[1 + \exp((\text{Age} - \theta_3)/\theta_4)]$	13	2.9168 (0.7308)	-5.1460 (22.1369)	-0.8508 (16.8113)	3.5608 (4.2535)
(7) $\log(\text{PCB}) = \theta_1 + (\theta_2 - \theta_1) \exp[\theta_3 \text{Age}^{\theta_4}]$	18	2.92618 (0.8067)	-0.5612 (0.8808)	0.1445 (0.1734)	1.1874 (0.7762)
(8) $\log(\text{PCB}) = 1 - \exp[-\exp(\theta_1 - \theta_2 \text{Age})]$	21	-3.4462 (5.8041)	-1.2850 (2.2605)	-	-
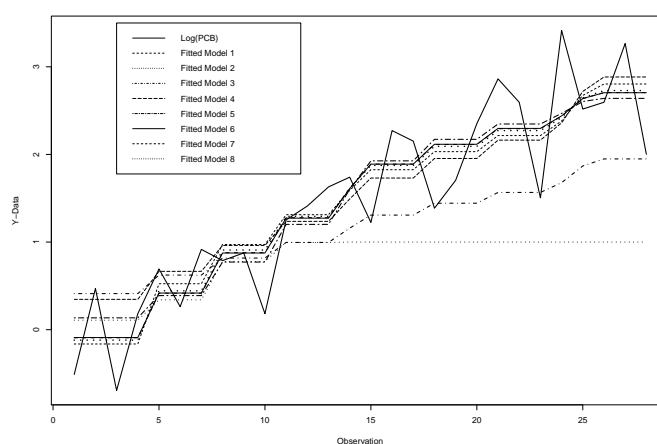
Figure 5.3: Observed log(PCB) values against estimated means considering the eight fitted models.

out that these posterior summaries obtained using MCMC methods are very accurate not depending on approximations as when considering the classical approach. From Figure 5.4 we observe that we have a good fit of some models for the data.

To decide on the best statistical model we also could use the selection Bayesian Deviance Information Criterion (DIC) introduced by Spiegelhalter et al. (2002). This criterion is especially useful in problems where samples of the posterior distribution for the parameters of the model have been simulated using Markov Chain Monte Carlo (MCMC) methods.

Define the deviation as

$$D(\boldsymbol{\theta}) = -2 \log L(\boldsymbol{\theta}) + C, \tag{21}$$

Table 5.6: Bayesian estimates for the growth models assuming a normal error.

Fitted model	DIC	Posterior Mean (Standard Deviation)			
		θ_1	θ_2	θ_3	θ_4
(1) $\log(\text{PCB}) = \theta_1 + \theta_2 \text{Age}^{\theta_3}$	45.0	-0.7804 (0.4839)	0.9227 (0.3843)	0.5946 (0.1458)	-
(2) $\log(\text{PCB}) = \theta_1 + \theta_2 \exp[\theta_3 \text{Age}]$	61.71	-0.5457 (0.5339)	1.0960 (0.4373)	0.1065 (0.0279)	-
(3) $\log(\text{PCB}) = \theta_1 \exp[-\theta_2 \text{Age}] + \theta_3 \exp[-\theta_4 \text{Age}]$	52.56	-1.701 (0.6316)	0.4978 (0.2242)	1.0940 (0.2721)	-0.0852 (0.0241)
(4) $\log(\text{PCB}) = \theta_1 Age/[\theta_2 + \text{Age}]$	67.6	2.3140 (0.3152)	2.0540 (0.6620)	-	-
(5) $\log(\text{PCB}) = \theta_1 \times \exp[-\exp(\theta_2 - \theta_3 \text{Age})]$	49.47	2.5060 (0.2677)	1.5406 (0.4094)	0.5010 (0.1404)	-
(6) $\log(\text{PCB}) = \theta_1 + (\theta_2 - \theta_1)/[1 + \exp((\text{Age} - \theta_3)/\theta_4)]$	51.17	2.3190 (0.2294)	-1.4580 (0.6554)	1.8730 (0.6762)	1.7460 (0.4869)
(7) $\log(\text{PCB}) = \theta_1 + (\theta_2 - \theta_1) \exp[\theta_3 \text{Age}^{\theta_4}]$	-6.609	-1.318 (0.6242)	-0.4967 (0.4649)	-0.7765 (0.3716)	0.3727 (0.1372)
(8) $\log(\text{PCB}) = 1 - \exp[-\exp(\theta_1 - \theta_2 \text{Age})]$	87.31	-0.3885 (1.0480)	-0.7789 (0.5216)	-	-

where $\boldsymbol{\theta}$ is the vector of unknown parameters in the model, $L(\boldsymbol{\theta})$ is the likelihood function and C is a constant that does not need to be known in the models comparison. The DIC criterion is given by

$$DIC = D(\widehat{\boldsymbol{\theta}}) + 2n_D, \tag{22}$$

where $D(\widehat{\boldsymbol{\theta}})$ is the deviation evaluated at the posterior mean $\widehat{\boldsymbol{\theta}} = \mathbb{E}(\boldsymbol{\theta}|data)$ and n_D is the effective number of parameters of the model given by $n_D = \overline{D} - D(\widehat{\boldsymbol{\theta}})$, with $\overline{D} = \mathbb{E}(D(\boldsymbol{\theta})|data)$ is the posterior deviation measuring the quality of the data fit for the model. Smaller values of DIC indicate better models. Notice that these values can be negative.
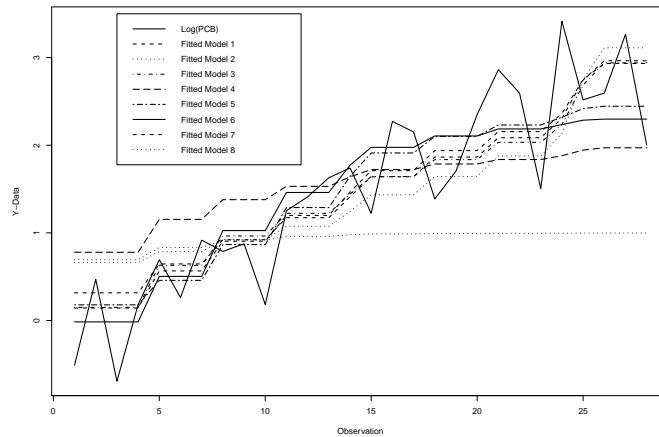

Figure 5.4: Observed log(PCB) values against posterior means considering eight fitted models.

Overall, model 5 (Gompertz) gives the smallest value for the sum of absolute values

Table 5.7: Bayesian estimates for the Gompertz growth model assuming a stable distribution for the responses.

Parameter	Mean	Standard Deviation	95% Credible Interval
α	0.8859	0.0759	(0.7069, 0.9738)
β	-0.7813	0.0918	(-0.9552, -0.5773)
d_1	2.6740	0.4199	(2.1720, 3.3060)
d_2	1.7300	0.5028	(0.8745, 2.6220)
d_3	0.4568	0.1405	(0.2735, 0.7191)
γ	0.4554	0.0805	(0.2670, 0.5993)

for the differences (observed - estimated mean) and small DIC (49.47) value, that is, we could consider this model as the best fitted model.

Now, we assume a stable distribution for the responses considering the Gompertz growth curve model (model 5) introduced in Section 4. Under a Bayesian approach, we have in Table 5.7, the posterior summaries of interest assuming the nonlinear regression model 5, that is, considering the regression model (13) for the location parameter of the stable distribution given by,

$$\delta_i = d_1 \exp[-\exp(d_2 - d_3 \text{Age}_i)]. \tag{23}$$

In the Bayesian analysis for this model, we use the OpenBugs software assuming the following prior distributions: $\alpha \sim \mathcal{U}(0,2)$, $\beta \sim \mathcal{U}(-1,0)$, $d_j \sim \mathcal{N}(0,1)$, $j = 1, 2, 3$ and $\gamma \sim \mathcal{U}(0,3)$. We also assume a uniform $\mathcal{U}(-0.5, 0.5)$ distribution for the latent variable Y_i, for $i = 1, 2, \cdots, 15$. We simulated 4,000,000 Gibbs samples, with a "burn-in-sample" of 1,000,000 samples discarded to eliminate the effects of the initial values in the iterative simulation procedure. We take a final sample of size 3,000 (every 1000th sample chosen from the 3,000,000 samples). Gibbs sampling algorithm convergence was monitored from standard trace plots of the simulated samples.

In Figure 5.5, we have the plot of the posterior means assuming Gompertz growth model (model 5) with a normal error against the observations together with the plot of the location posterior against observations assuming a stable distribution. We observe similar results. We also observe that the sum of absolute values for the differences (observed - estimated mean) for model 5 with a stable distribution is given by 12.1437, that is, similar to the obtained value assuming normal errors (11.0915).

Let us now consider the presence of one outlier or discordant response (considered as a measurement error) replacing the 5th PCB response (2.0) value in Table 5.4 by the value 200.0. This high value is necessary because otherwise the effect can not be observed. In Table 5.8, we have the obtained posterior summaries assuming the same priors and same simulation procedure as in Table 5.7. In Figure 5.6, we have the plots of observed and fitted means considering both models versus samples. From the graphs of Figure 5.6, we observe that the model with a stable distribution is very robust to the presence of the outlier given similar inference results as obtained without the presence of this outlier (see results in Table 5.7). We also observe in Table 5.8, that the estimated regression parameters with normal error are strongly affected by the presence of this outlier. Observe that we have sum of absolute values for the differences (observed - estimated mean) equals to
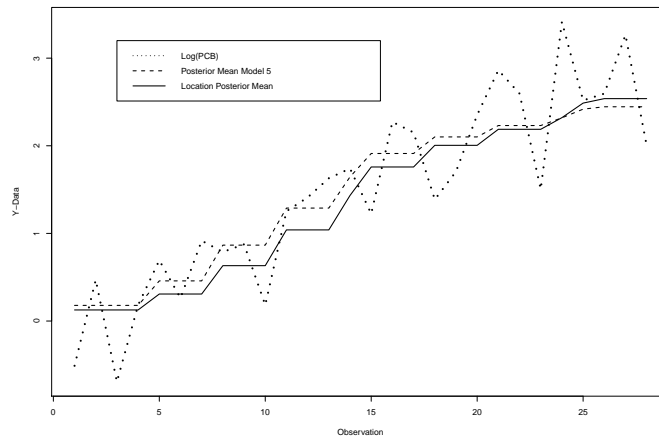
Figure 5.5: Observed log(PCB) values against posterior means considering the Gompertz growth model, with Normal error and assuming a stable distribution.
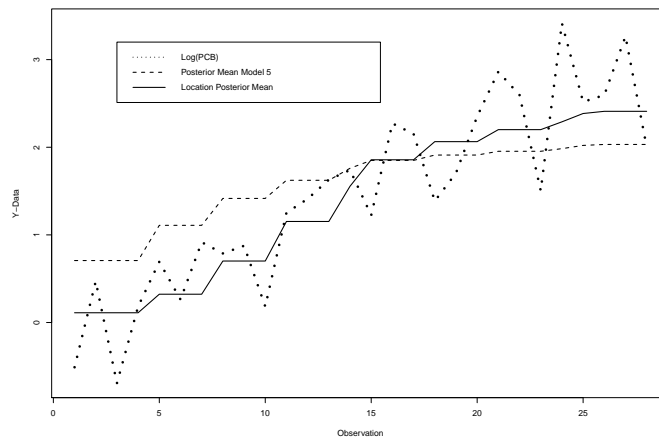

Figure 5.6: Observed log(PCB) values against posterior means considering the Gompertz growth model with Normal error, assuming a stable distribution (presence of one outlier).

11.8064 assuming a stable distribution (a value very close to the respective one without the presence of an outlier) while the sum of absolute values for the differences (observed - estimated mean) is equal to 16.1205 assuming a normal distribution.

6   Some concluding remarks

The use of stable distributions could be a good alternative, especially using Bayesian methods with MCMC simulation technique. This class of distributions is very robust to the presence of discordant observations. The presence of outliers or discordant observations, many times due to measurement errors, is very common in practical applications of linear or nonlinear regression analysis. In the presence of these discordant observations, the usual obtained classical inferences on the regression parameters or in the predictions un-

Table 5.8: Bayesian estimates for the Gompertz growth model assuming a stable distribution for the responses (presence of one outlier).

Stable distribution (SAV = 11.8064)			
Parameter	Mean	Standard Deviation	95% Credible Interval
α	0.8682	0.0806	(0.6694, 0.9674)
β	-0.6747	0.2485	(-0.9476, -0.0344)
d_1	2.4580	0.2719	(2.0980; 3.1040)
d_2	1.8230	0.4832	(0.9447; 2.6920)
d_3	0.5305	0.1201	(0.3114; 0.7528)
γ	0.4202	0.1141	(0.1708, 0.6201)
Normal distribution (SAV = 16.1205)			
Parameter	Mean	Standard Deviation	95% Credible Interval
θ_1	2.0730	0.3714	(1.4060, 2.8870)
θ_2	0.7145	0.6787	(-0.7258, 2.0210)
θ_3	0.6920	0.3985	(0.1847, 1.6290)

der the usual assumption of errors normality and homoscedastic variance could be greatly affected, which could imply in wrong inference results. In this way, the use of stable distributions could be a good alternative, since this distribution has a great flexibility to fit for the data. With the use of Bayesian methods and MCMC simulation algorithms it is possible to obtain inferences for the model despite the nonexistence of an analytical form for the density function as it was showed in this work. It is important to point out that the computational effort in the sample simulations for the joint posterior distribution of interest can be largely simplified using standard free available softwares like the OpenBUGS software.

In the illustrative examples introduced here, we observed that the use of data augmentation techniques (see, for instance, Damien et al., 1999) is the key to obtain a good performance for the MCMC simulation method for applications using stable distributions.

We emphasize that the use of OpenBUGS software does not require large computational time to get posterior summaries of interest, even when the simulation of a large number of Gibbs samples are needed to achieve the algorithm convergence. These results could be of great interest for researchers and practitioners, when dealing with non Gaussian data, as in the applications presented here.

Acknowledgments

References

Abramowitz, M. and Stegun, I.A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications.

Achcar, J.A., Lopes, S.R.C., Mazucheli, J. and Linhares, R.R. (2013a). "A Bayesian approach for stable distributions: some computational aspects". *Open Journal of Statistics*, Vol. **3**, 268-277.

Achcar, J.A., Achcar, A. and Martinez, E.Z. (2013b). "Robust linear regression models: use of a stable distribution for the response data". *Open Journal of Statistics*, Vol. **3**, 409-416.

Bache, C.A., Serum, J.W., Youngs, W.D. and Lisk, D.J. (1972). "Polychlorinated bibhenyl residues: accumulation in Cayuga Lake trout with age". *Science*, Vol. **117**, 1192-1193.

Bates, D.M. and Watts, D.G. (1980). "Relative curvature measures of nonlinearity (with discussion)". *Journal of the Royal Statistical Society*, Series B, Vol. **42**, 1-25.

Bates, D.M. and Watts, D.G. (1988). *Nonlinear regression analysis and its applications*. New York: Wiley.

Buckle, D.J. (1995). "Bayesian inference for stable distributions". *Journal of the American Statistical Association*, Vol. **90**, 605-613.

Chib, S. and Greenberger, E. (1995). "Understanding the Metropolis-Hastings algorithm". *The American Statistician*, Vol. **49**, 327-335.

Damien, P., Wakefield, J. and Walker, S. (1999). "Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables". *Journal of the Royal Statistical Society*, Series B, Vol. **61**, 331-344.

Draper, N.R. and Smith, H. (1981). *Applied regression analysis*. New York: Wiley.

Gnedenko, B.V. and Kolmogorov, A.N. (1968). *Limit distributions for sums of independent random variables*. Massachussetts: Addison-Wesley.

Ibragimov, I.A. and Černin, K.E. (1959). "On the unimodality of stable laws". *Teor. Veroyatnost.i Primenen.*, Vol. **4**, 453-456.

Johnson, R.A. and G.K. Bhattacharyya (1980). *Statistical Principles and Methods*. 1st Edition. New York: John Wiley.

Kanter, M. (1976). "On the unimodality of stable densities". *Annals of Probability*, Vol. **4**, 1006-1008.

Lévy, P. (1924). "Théorie des erreurs la loi de Gauss et les lois exceptionelles", *Bulletin Society Mathematical*, Vol. **52**, 49-85.

Lukacs, E. (1970). *Characteristic Functions*. New York: Hafner Publishing.

Nolan, J.P. (2015). "Stable Distributions - Models for Heavy Tailed Data". Boston: Birkhäuser.

Ratkowsky, D.A. (1983). *Nonlinear regression modelling: a unified practical approach*. New York: Marcel Dekker.

Robert, C.P. and Casella, G. (2004). *Monte Carlo statistical methods*. 2nd ed.. New York: Springer-Verlag.

Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes*. New York: Chapman & Hall.

Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear regression*. New York: Wiley.

Seber, G.A.F. and Lee, A.J. (2003). *Linear regression analysis*. 2nd ed.. New York: Wiley.

Skorohod, A.V. (1961). "On a theorem concerning stable distributions". In: Select. Transl. Math. Statist. and Probability, Vol. **1**, Inst. Math. Statist. and Amer. Math. Soc., Providence, R.I., 169-170.

Spiegelhalter, D.J., Thomas, A., Best, N.G. and Lunn, D. (2003). *WinBUGS User's Manual*. Cambridge: MRC Biostatistics Unit.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002). "Bayesian measures of model complexity and fit". *Journal of the Royal Statistical Society*, Series B, Vol. **64**(4), 583-639.

Tanner, M.A. and Wong, W.H. (1987). "The calculation of posterior distributions by data augmentation". *Journal of American Statistical Association*, Vol. **82**, 528-550.