

## Research Article

# Robust Mean Change-Point Detecting through Laplace Linear Regression Using EM Algorithm

Fengkai Yang<sup>1,2</sup>

<sup>1</sup> School of Mathematics, Shandong University, Jinan 250100, China

<sup>2</sup> School of Mathematics and Statistics, Shandong University, Weihai 264209, China

Correspondence should be addressed to Fengkai Yang; [yfk@sdu.edu.cn](mailto:yfk@sdu.edu.cn)

Received 20 July 2014; Revised 5 September 2014; Accepted 5 September 2014; Published 4 November 2014

Academic Editor: Zihua Zhang

Copyright © 2014 Fengkai Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We proposed a robust mean change-point estimation algorithm in linear regression with the assumption that the errors follow the Laplace distribution. By representing the Laplace distribution as an appropriate scale mixture of normal distribution, we developed the expectation maximization (EM) algorithm to estimate the position of mean change-point. We investigated the performance of the algorithm through different simulations, finding that our methods is robust to the distributions of errors and is effective to estimate the position of mean change-point. Finally, we applied our method to the classical Holbert data and detected a change-point.

## 1. Introduction

Change-point analysis has been an active research area since the early 1950s. During the following period of sixty-some years, numerous articles have been published in various journals and proceedings. Chen and Gupta [1] summed up the main methods and applications on change-point detection and estimation. Ever since the change-point hypothesis was introduced into statistical analyses, the study of switching regression models has taken place in regression analysis. This made some previously poorly fitted regression models better fitted to some datasets after the change-point has been located in the regression models.

The Schwarz information criterion (SIC) proposed by Schwarz [2] has been applied to change-point analysis for different underlying models by many authors in the literature. Chen [3] was the first to use the SIC model selection method to study the mean change-point problem in normal linear regression model, and later Chen and Gupta [4] used the same method to detect both mean and variance change-point in normal model. Chen and Wang [5] developed a statistical change-point model approach for the detection of DNA copy number variations in array CGH data using the SIC method and assuming the error follows the normal distribution.

However, in practice, we do not know the real distribution of the data, and it is difficult to determine the real distribution especially when some change-point is present in the data. So, the normal assumption is not always suitable, for a lot of real data usually shows heavy tail and skewness. In such cases, some robust change-point detecting model with heavy-tailed distribution might be better than the normal model. Osorio and Galea [6] developed a mean change-point linear regression model with independent errors distributed according to the Student  $t$ -distribution and located the change-point using the SIC method. Lin et al. [7] studied the variance change-points in the Student  $t$  regression model under Bayesian framework and analyzed the U.S. stock market data.

The symmetric Laplace distribution, also known as the double exponential distribution or the first law of Laplace, is another heavy-tailed error distribution besides the Student  $t$ -distribution. It is less sensitive to the outlier and more robust than the normal distribution. Kotz et al. [8] gave a systematic overview of the Laplace distribution, in their book, and the authors were devoted to presenting properties, generalizations of the Laplace distribution, and their applications in communications, economics, engineering, and finance.

In recent years, statistical models based on Laplace distribution have developed rapidly both in theory and application.

Purdum and Holmes [9] found that the error distribution for gene expression data from microarray experiments can be better fitted by Laplace distribution than normal distribution. Pop [10] identified a Laplace distribution in the change in daily sunspot number, and later Noble and Wheatland [11] showed the physical origin of Laplace distribution and its use in daily sunspot numbers. van Sanden and Burzykowski [12] considered the analysis of microarray data by using ANOVA models under the assumption of Laplace-distributed error terms. Phillips [13] developed the expectation maximization (EM) algorithm for the least absolute deviation regression, which is also known as the Laplace regression or median regression. Park and Casella [14] interpreted the Lasso estimate for linear regression parameters as a Bayesian posterior mode estimate when the regression parameters have independent Laplace priors. Song et al. [15] proposed a robust estimation procedure for mixture linear regression models assuming that the error terms follow the Laplace distribution.

In this paper, we study the single mean change-point problem in linear regression model assuming that the error follows the Laplace distribution via EM algorithm and use the SIC model selection method to estimate the position of the mean change-point. Then, we investigate the robustness of the algorithm through simulations under different error distributions. Finally, we apply our method to some stock market data set.

## 2. Laplace Linear Regression Model with Mean Change-Point

*2.1. Laplace Distribution as Scale Mixture of the Normal Distribution.* The symmetric Laplace distribution is commonly denoted by  $L(\mu, \sigma)$ , where  $\mu \in (-\infty, \infty)$  is the location parameter and  $\sigma > 0$  is the scale parameter. The density function is given by

$$f(x) = \frac{1}{2\sigma} \exp\left(-\frac{|x-\mu|}{\sigma}\right), \quad x \in \mathbb{R}, \quad (1)$$

with mean  $\mu$  and variance  $2\sigma^2$ . The Laplace distribution is more peaked in the center and more heavy-tailed compared to the normal distribution, and, for a given samples  $X_1, X_2, \dots, X_n$  from  $L(\mu, \sigma)$ , the maximum likelihood estimate of  $\mu$  is the median of the samples, which is robust to the outliers.

Andrews and Mallows [16] presented necessary and sufficient conditions under which a random variable  $X$  may be generated as the ratio  $Z/V$ , where  $Z$  and  $V$  are independent and  $Z$  has a standard normal distribution. Random variable  $X$  is referred to as a normal variance mixture distribution or a scale mixture of Gaussian distribution. It was established that when  $V^2/2$  is exponential,  $X$  is double exponential, so Laplace is some kind of scale mixture of Gaussian distribution, where the mixing distribution is exponential distribution.

The statement above can be described by the following proposition.

**Proposition 1** (representation of Laplace distribution). *Suppose random variable  $\epsilon \sim L(\mu, \sigma)$ , and then there exists a random variable  $\omega \sim \Gamma(1, 1/2)$ , such that  $\epsilon \mid \omega \sim N(\mu, \omega\sigma^2)$ , where  $\Gamma(\alpha, \beta)$  denotes the Gamma distribution with parameters  $\alpha$  and  $\beta$ .*

*2.2. Laplace Regression Model with Mean Change-Point.* Let  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$  be a sequence of observations obtained in a practical situation, where  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ ,  $i = 1, 2, \dots, n$ , is a nonstochastic  $p$  vector variable, and the model we are going to discuss is

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad i = 1, 2, \dots, n, \quad (2)$$

where  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  is a  $p$  unknown parameter vector and  $\epsilon_i$ ,  $i = 1, \dots, n$  are random errors, which are independent and identically distributed as  $L(0, \sigma)$ , with  $\sigma$  unknown. We can see that  $y_i$ ,  $i = 1, 2, \dots, n$ , are independently distributed and, for  $i = 1, 2, \dots, n$ ,  $y_i \sim L(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma)$ . To develop EM algorithm for the Laplace linear regression model, we utilize a mixture representation of the Laplace regression, which is given by Proposition 2, derived from Proposition 1.

**Proposition 2** (representation of Laplace linear regression). *Suppose random variables  $y_i \sim L(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma)$ ,  $i = 1, \dots, n$ , and then there exist random latent variables  $\omega_i$ ,  $i = 1, \dots, n$ , where  $\omega_i \sim \Gamma(1, 1/2)$ , such that  $y_i \mid \omega_i \sim N(\mathbf{x}_i^T \boldsymbol{\beta}, \omega_i \sigma^2)$ , for  $i = 1, 2, \dots, n$ .*

Proposition 2 represents the Laplace linear regression model as a normal mixture model with mixture variables following the Gamma distribution. This kind of representation is crucial to perform the EM algorithm in the following mean change-point model.

The single mean (regression coefficients) change-point problem in a Laplacian linear regression model can be formulated as to test the following null hypothesis:

$$H_0 : y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad i = 1, 2, \dots, n, \quad (3)$$

versus the alternative

$$H_1 : y_i = \mathbf{x}_i^T \boldsymbol{\beta}_1 + \epsilon_i, \quad i = 1, 2, \dots, k, \quad (4)$$

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}_2 + \epsilon_i, \quad i = k + 1, \dots, n,$$

where

$$\boldsymbol{\beta}_1 = (\beta_0, \beta_1, \dots, \beta_{p-1})^T, \quad \boldsymbol{\beta}_2 = (\beta_0^*, \beta_1^*, \dots, \beta_{p-1}^*)^T, \quad (5)$$

$$\epsilon_i \sim L(0, \sigma), \quad i = 1, 2, \dots, n.$$

That is, a change exists (in the regression coefficients) in an unknown position  $k$ , denominated mean change-point. In

the following section, we perform EM algorithm to estimate the unknown parameters and use the SIC model selection method to detect the position of change-point.

### 3. EM Algorithm and Schwarz Information Criterion

3.1. *EM Algorithm under  $H_0$ .* Denote  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ ,  $\mathbf{y} = (y_1, \dots, y_n)^T$  and random latent vector  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ . The likelihood function of  $\mathbf{y}$  under  $H_0$  is

$$L(\boldsymbol{\beta}, \sigma | \mathbf{y}) = \prod_{i=1}^n \frac{1}{2\sigma} \exp \left\{ -\frac{|y_i - \mathbf{x}_i^T \boldsymbol{\beta}|}{\sigma} \right\}. \quad (6)$$

From the stochastic representation in Section 2.2, the regression model under  $H_0$  becomes

$$y_i | \omega_i \sim N(\mathbf{x}_i^T \boldsymbol{\beta}, \omega_i \sigma^2); \quad \omega_i \sim \Gamma\left(1, \frac{1}{2}\right), \quad (7)$$

$$i = 1, 2, \dots, n.$$

So, the likelihood function of the complete data  $(\mathbf{y}, \boldsymbol{\omega})$  has the following form:

$$L(\boldsymbol{\beta}, \sigma | \mathbf{y}, \boldsymbol{\omega}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\omega_i}\sigma} \exp \left\{ -\frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{2\sigma^2\omega_i} \right\} \times \frac{1}{2} \exp \left\{ -\frac{1}{2}\omega_i \right\}. \quad (8)$$

Consequently, the complete log-likelihood function is

$$l(\boldsymbol{\beta}, \sigma | \mathbf{y}, \boldsymbol{\omega}) = -\frac{n}{2} \log \pi - n \log 2 - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\omega_i} - \frac{1}{2} \sum_{i=1}^n \omega_i - \frac{1}{2} \sum_{i=1}^n \log \omega_i. \quad (9)$$

Given the initial values  $\boldsymbol{\beta}^{(0)}$ ,  $\sigma^{(0)}$ , we can obtain the maximum likelihood estimate of  $\boldsymbol{\beta}$ ,  $\sigma$  based on EM algorithm via the following two steps.

(i) *E Step.* Given the  $(t-1)$ th iteration values of  $(\boldsymbol{\beta}^{(t-1)}, \sigma^{(t-1)})$  and omitting the terms having no relation to  $(\boldsymbol{\beta}, \sigma)$ , the Q function of the  $(t)$ th iteration is

$$Q(\boldsymbol{\beta}, \sigma | \mathbf{y}, \boldsymbol{\beta}^{(t-1)}, \sigma^{(t-1)}) = E[l(\boldsymbol{\beta}, \sigma | \mathbf{y}, \boldsymbol{\omega}) | \mathbf{y}, \boldsymbol{\beta}^{(t-1)}, \sigma^{(t-1)}] = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n \delta_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2, \quad (10)$$

where  $\delta_i = E((1/\omega_i) | \mathbf{y}, \boldsymbol{\beta}^{(t-1)}, \sigma^{(t-1)})$ .

In order to obtain  $\delta_i$ , we firstly compute the conditional probability density function (pdf) of  $f(\omega_i | y_i, \boldsymbol{\beta}, \sigma)$ . Note that the joint pdf of  $\mathbf{y}$  and  $\boldsymbol{\omega}$  is

$$f(\mathbf{y}, \boldsymbol{\omega} | \boldsymbol{\beta}, \sigma) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\omega_i}\sigma} \exp \left\{ -\frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{2\sigma^2\omega_i} \right\} \times \frac{1}{2} \exp \left\{ -\frac{1}{2}\omega_i \right\}. \quad (11)$$

Then, we have

$$f(\boldsymbol{\omega} | \mathbf{y}, \boldsymbol{\beta}, \sigma) \propto f(\mathbf{y}, \boldsymbol{\omega} | \boldsymbol{\beta}, \sigma) \propto \prod_{i=1}^n \omega_i^{-1/2} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\omega_i} \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\sigma^2} + \omega_i \right] \right\}. \quad (12)$$

Due to the conditionally independent of  $\omega_1, \omega_2, \dots, \omega_n$ , given  $\mathbf{y}$ , we obtain the marginal conditional pdf of  $\omega_i$  given  $y_i$  as

$$f(\omega_i | y_i, \boldsymbol{\beta}, \sigma) \propto \omega_i^{(1/2)-1} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\omega_i} \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\sigma^2} + \omega_i \right] \right\} \sim \text{GIG} \left( \frac{1}{2}, \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\sigma^2}, 1 \right), \quad (13)$$

where  $\text{GIG}(\lambda, a, b)$  is the generalized inverse Gaussian distribution and its pdf is given by

$$f(x | \lambda, a, b) = \frac{(b/a)^{\lambda/2}}{2K_\lambda(\sqrt{ab})} x^{\lambda-1} \exp \left\{ -\frac{1}{2} (ax^{-1} + bx) \right\}, \quad (14)$$

where  $K_\lambda(\cdot)$  is the modified Bessel function of the third kind in Barndorff-Nielsen and Shephard [17].

Therefore, we can obtain

$$\delta_i = E \left( \frac{1}{\omega_i} | \mathbf{y}, \boldsymbol{\beta}^{(t-1)}, \sigma^{(t-1)} \right) = \frac{\sigma^{(t-1)}}{|y_i - \mathbf{x}_i^T \boldsymbol{\beta}^{(t-1)}|}. \quad (15)$$

(ii) *M Step.* we maximize the Q function in E step with respect to  $\boldsymbol{\beta}$ ,  $\sigma$ .

Denote  $\boldsymbol{\Sigma} = \text{diag}(\delta_1, \dots, \delta_n)$ , and we get

$$\boldsymbol{\beta}^{(t)} = (\mathbf{X}^T \boldsymbol{\Sigma} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma} \mathbf{y}, \quad (16)$$

$$\sigma^{(t)} = \left( \frac{1}{n} \sum_{i=1}^n \delta_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta}^{(t)})^2 \right)^{-1/2}. \quad (17)$$

It can be seen from (15) and (16) that the updating formulae of  $\boldsymbol{\beta}$  is independent of the estimated value of  $\sigma$ ; in other words, we need not to update  $\sigma$  at each EM iteration, and, after the final estimate  $\tilde{\boldsymbol{\beta}}$  of  $\boldsymbol{\beta}$  is obtained, the estimate of  $\sigma$  can then be gotten by maximizing the original log-likelihood function which produces

$$\tilde{\sigma} = \frac{1}{n} \sum_{i=1}^n |y_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}|. \quad (18)$$

Finally, the maximum of the log-likelihood function under  $H_0$ , is obtained as

$$\begin{aligned} l_0(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}) &= -n \log 2 - n \log \tilde{\sigma} - n \\ &= -n \log \frac{2}{n} - n \log S(\tilde{\boldsymbol{\beta}}) - n \end{aligned} \quad (19)$$

with

$$S(\tilde{\boldsymbol{\beta}}) = \sum_{i=1}^n |y_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}|. \quad (20)$$

**3.2. EM Algorithm under  $H_1$ .** Denote  $\mathbf{X}_1 = (\mathbf{x}_1, \dots, \mathbf{x}_k)^T$ ,  $\mathbf{y}_1 = (y_1, \dots, y_k)^T$ ,  $\mathbf{X}_2 = (\mathbf{x}_{k+1}, \dots, \mathbf{x}_n)^T$ ,  $\mathbf{y}_2 = (y_{k+1}, \dots, y_n)^T$ , and random latent vectors  $\boldsymbol{\omega}_1 = (\omega_1, \dots, \omega_k)^T$ ,  $\boldsymbol{\omega}_2 = (\omega_{k+1}, \dots, \omega_n)^T$ .

The likelihood function of  $y$  under  $H_1$  is

$$\begin{aligned} L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma | \mathbf{y}) &= \prod_{i=1}^k \frac{1}{2\sigma} \exp \left\{ -\frac{|y_i - \mathbf{x}_i^T \boldsymbol{\beta}_1|}{\sigma} \right\} \\ &\times \prod_{i=k+1}^n \frac{1}{2\sigma} \exp \left\{ -\frac{|y_i - \mathbf{x}_i^T \boldsymbol{\beta}_2|}{\sigma} \right\}. \end{aligned} \quad (21)$$

From the stochastic representation in Section 2.2, the regression model under  $H_1$  becomes

$$\begin{aligned} y_i | \omega_i &\sim N(\mathbf{x}_i^T \boldsymbol{\beta}_1, \omega_i \sigma^2), \quad i = 1, \dots, k, \\ y_i | \omega_i &\sim N(\mathbf{x}_i^T \boldsymbol{\beta}_2, \omega_i \sigma^2), \quad i = k+1, \dots, n, \\ \omega_i &\sim \Gamma\left(1, \frac{1}{2}\right), \quad i = 1, 2, \dots, n, \end{aligned} \quad (22)$$

where  $\omega_i, i = 1, \dots, n$  are independent.

So, the likelihood function of the complete data  $(\mathbf{y}, \boldsymbol{\omega})$  under  $H_1$  has the following form:

$$\begin{aligned} L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma | \mathbf{y}, \boldsymbol{\omega}) \\ = \prod_{i=1}^k \frac{1}{\sqrt{2\pi\omega_i}\sigma} \exp \left\{ -\frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_1)^2}{2\sigma^2\omega_i} \right\} \end{aligned}$$

$$\begin{aligned} &\times \frac{1}{2} \exp \left\{ -\frac{1}{2}\omega_i \right\} \\ &\times \prod_{i=k+1}^n \frac{1}{\sqrt{2\pi\omega_i}\sigma} \exp \left\{ -\frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_1)^2}{2\sigma^2\omega_i} \right\} \\ &\times \frac{1}{2} \exp \left\{ -\frac{1}{2}\omega_i \right\}. \end{aligned} \quad (23)$$

Consequently, the complete log-likelihood function is

$$\begin{aligned} l(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma | \mathbf{y}, \boldsymbol{\omega}) &= -\frac{n}{2} \log 2\pi - n \log 2 \\ &- \frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum_{i=1}^n \omega_i - \frac{1}{2} \sum_{i=1}^n \log \omega_i \\ &- \frac{1}{2\sigma^2} \left[ \sum_{i=1}^k \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_1)^2}{\omega_i} \right. \\ &\quad \left. + \sum_{i=k+1}^n \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_2)^2}{\omega_i} \right]. \end{aligned} \quad (24)$$

Given initial values  $\boldsymbol{\beta}_1^{(0)}, \boldsymbol{\beta}_2^{(0)}$ , and  $\sigma^{(0)}$ , we can obtain the maximum likelihood estimate of  $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ , and  $\sigma$  based on EM algorithm via the following two steps.

(i) *E Step.* Given the  $(t-1)$ th iteration values of  $(\boldsymbol{\beta}^{(t-1)}, \sigma^{(t-1)})$  and omitting the terms having no relation to  $(\boldsymbol{\beta}, \sigma)$ , the Q function of the  $(t)$ th iteration is

$$\begin{aligned} Q(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma | y, \boldsymbol{\beta}_1^{(t-1)}, \boldsymbol{\beta}_2^{(t-1)}, \sigma^{(t-1)}) \\ = E[l(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma | \mathbf{y}, \boldsymbol{\omega}) | y, \boldsymbol{\beta}_1^{(t-1)}, \boldsymbol{\beta}_2^{(t-1)}, \sigma^{(t-1)}] \\ = -\frac{n}{2} \log \sigma^2 \end{aligned} \quad (25)$$

$$\begin{aligned} -\frac{1}{2\sigma^2} \left[ \sum_{i=1}^k \gamma_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_1)^2 \right. \\ \left. + \sum_{i=k+1}^n \gamma_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_2)^2 \right], \end{aligned}$$

where  $\gamma_i = E((1/\omega_i) | y, \boldsymbol{\beta}_1^{(t-1)}, \boldsymbol{\beta}_2^{(t-1)}, \sigma^{(t-1)})$ .

The marginal conditional pdf of  $\omega_i$  given  $y_i$  is that, for  $i = 1, \dots, k$ ,

$$\begin{aligned} f(\omega_i | y_i, \boldsymbol{\beta}_1, \sigma) \\ \propto \omega_i^{(1/2)-1} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\omega_i} \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_1)^2}{\sigma^2} + \omega_i \right] \right\} \\ \sim \text{GIG} \left( \frac{1}{2}, \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_1)^2}{\sigma^2}, 1 \right), \end{aligned} \quad (26)$$

and, for  $i = k + 1, \dots, n$ ,

$$f(\omega_i | y_i, \boldsymbol{\beta}_2, \sigma) \propto \omega_i^{(1/2)-1} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\omega_i} \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_2)^2}{\sigma^2} + \omega_i \right] \right\} \quad (27)$$

$$\sim \text{GIG} \left( \frac{1}{2}, \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_2)^2}{\sigma^2}, 1 \right).$$

Therefore, for  $i = 1, \dots, k$ ,

$$\gamma_i = E \left( \frac{1}{\omega_i} \mid y, \boldsymbol{\beta}_1^{(t-1)}, \sigma^{(t-1)} \right) = \frac{\sigma^{(t-1)}}{|y_i - \mathbf{x}_i^T \boldsymbol{\beta}_1^{(t-1)}|}, \quad (28)$$

and, for  $i = k + 1, \dots, n$ ,

$$\gamma_i = E \left( \frac{1}{\omega_i} \mid y, \boldsymbol{\beta}_2^{(t-1)}, \sigma^{(t-1)} \right) = \frac{\sigma^{(t-1)}}{|y_i - \mathbf{x}_i^T \boldsymbol{\beta}_2^{(t-1)}|}. \quad (29)$$

(ii) *M Step.* We maximize the Q function in E step with respect to  $\boldsymbol{\beta}_1$ ,  $\boldsymbol{\beta}_2$ , and  $\sigma$ .

Denote  $\boldsymbol{\Sigma}_1 = \text{diag}(\gamma_1, \dots, \gamma_k)$ ,  $\boldsymbol{\Sigma}_2 = \text{diag}(\gamma_{k+1}, \dots, \gamma_n)$ , and we get

$$\boldsymbol{\beta}_1^{(t)} = (\mathbf{X}_1^T \boldsymbol{\Sigma}_1 \mathbf{X}_1)^{-1} \mathbf{X}_1^T \boldsymbol{\Sigma}_1 \mathbf{y}_1,$$

$$\boldsymbol{\beta}_2^{(t)} = (\mathbf{X}_2^T \boldsymbol{\Sigma}_2 \mathbf{X}_2)^{-1} \mathbf{X}_2^T \boldsymbol{\Sigma}_2 \mathbf{y}_2,$$

$$\sigma^{(t)} = \left( \frac{1}{n} \left[ \sum_{i=1}^k \gamma_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_1^{(t)})^2 + \sum_{i=k+1}^n \gamma_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_2^{(t)})^2 \right] \right)^{-1/2}. \quad (30)$$

It can be seen from (28)–(30) that the updating formulae of  $\boldsymbol{\beta}_1$ ,  $\boldsymbol{\beta}_2$  is independent of the estimated value of  $\sigma$ , so we need not to update  $\sigma$  at each EM iteration. After the final estimates  $\hat{\boldsymbol{\beta}}_1$  of  $\boldsymbol{\beta}_1$  and  $\hat{\boldsymbol{\beta}}_2$  of  $\boldsymbol{\beta}_2$  are obtained, the estimate of  $\sigma$  can then be gotten by maximizing the original log-likelihood function, which produces

$$\hat{\sigma} = \frac{1}{n} \left( \sum_{i=1}^k |y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_1| + \sum_{i=k+1}^n |y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_2| \right). \quad (31)$$

Finally, the maximum of the log-likelihood function under  $H_1$  is obtained as

$$l_k(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\sigma}) = -n \log 2 - n \log \hat{\sigma} - n$$

$$= -n \log \frac{2}{n} - n \log [S(\hat{\boldsymbol{\beta}}_1) + S(\hat{\boldsymbol{\beta}}_2)] - n \quad (32)$$

with

$$S(\hat{\boldsymbol{\beta}}_1) = \sum_{i=1}^k |y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_1|, \quad S(\hat{\boldsymbol{\beta}}_2) = \sum_{i=k+1}^n |y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_2|. \quad (33)$$

**3.3. SIC Algorithm for Laplacian Regression Model.** The work by Chen [3] proposes transforming the process of hypothesis testing in a procedure of model selection using the Schwarz Information Criterion (SIC) defined by

$$\text{SIC} = -2l(\hat{\theta}) + s \log n, \quad (34)$$

where  $l(\hat{\theta})$  corresponds to the log-likelihood function evaluated on the maximum likelihood estimate of the parameters and  $s$  is the number of model parameters and  $n$  is the sample size. Note that maximizing the log-likelihood function is equivalent to minimizing the Schwarz information criterion.

In the Laplacian linear regression model, the Schwarz information criterion under  $H_0$ , denoted by  $\text{SIC}(n)$ , is given by

$$\text{SIC}(n) = -2l_0(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}) + (p+1) \log n$$

$$= 2n \log S(\tilde{\boldsymbol{\beta}}) + 2n \log \left( \frac{2}{n} \right) + 2n + (p+1) \log n, \quad (35)$$

where  $l_0(\tilde{\boldsymbol{\beta}}, \tilde{\sigma})$  and  $S(\tilde{\boldsymbol{\beta}})$  are given by (19) and (20). The Schwarz information criterion under  $H_1$ , denoted by  $\text{SIC}(k)$ , is given by

$$\text{SIC}(k) = -2l_k(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\sigma}) + (2p+1) \log n$$

$$= 2n \log [S(\hat{\boldsymbol{\beta}}_1) + S(\hat{\boldsymbol{\beta}}_2)] + 2n \log \left( \frac{2}{n} \right) + 2n + (2p+1) \log n \quad (36)$$

with  $l_k(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\sigma})$ ,  $S(\hat{\boldsymbol{\beta}}_1)$ , and  $S(\hat{\boldsymbol{\beta}}_2)$  given by (32) and (33).

The selection criteria are to choose a model with a change-point in the  $k$  position, if, for some  $k$ ,

$$\text{SIC}(n) > \text{SIC}(k). \quad (37)$$

When the null hypothesis is rejected, the maximum likelihood estimate of the change-point in the regression coefficients, denoted by  $\hat{k}$ , must satisfy

$$\text{SIC}(\hat{k}) = \min \{ \text{SIC}(k) : p \leq k \leq n-p \}$$

$$= \max \{ l_k(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\sigma}) : p \leq k \leq n-p \} \quad (38)$$

$$= \min \{ S(\hat{\boldsymbol{\beta}}_1) + S(\hat{\boldsymbol{\beta}}_2) \}.$$

## 4. Simulation Studies

In this section, we investigate the performance of the proposed approach to detect mean change-point through simulations, and we compare our procedure with the change-point detecting procedure proposed by Chen [3], who assumed

that the errors follow the normal distribution. A data set of  $n = 200$  observations is generated from the model

$$\begin{aligned} y_i &= \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i, & i = 1, \dots, k; \\ y_i &= \beta_1^* x_{1i} + \beta_2^* x_{2i} + \beta_3^* x_{3i} + \epsilon_i, & i = k + 1, \dots, n, \end{aligned} \quad (39)$$

where  $\beta_1 = (\beta_1, \beta_2, \beta_3)^T = (1, 1, 1)^T$ ,  $\beta_2 = (\beta_1^*, \beta_2^*, \beta_3^*)^T = (2, 3, 4)^T$ , and  $x_{ji} \sim U(-1, 1)$  for  $j = 1, \dots, 3$  and  $i = 1, \dots, n$ . For demonstrating the performances of the algorithm in different cases, we choose  $k = 40, 60, 80, 100, 120, 140, 160$ , respectively, and consider the following six error distributions:

- (i) Simulation 1: normal distribution:  $\epsilon_i \sim N(0, 1)$ ;
- (ii) Simulation 2: Laplace distribution:  $\epsilon_i \sim L(0, 1)$ ;
- (iii) Simulation 3:  $t$  distribution with three degree of freedom:  $\epsilon_i \sim t(3)$ ;
- (iv) Simulation 4:  $\chi^2$  distribution with three degree of freedom:  $\epsilon_i \sim \chi^2(3)$ ;
- (v) Simulation 5: log-normal distribution:  $\epsilon_i \sim LN(0, 1)$ ;
- (vi) Simulation 6: Cauchy distribution.

In order to evaluate the finite sample performance of the proposed method, 500 replications are conducted for different error distribution, respectively. In each replication, the initial values of  $\beta$  and  $\beta_1, \beta_2$  are set to be their ordinary least squares estimates, and the EM procedure is stopped when the absolute differences between two successive values of maximum log-likelihood functions are less than a preassigned small number (e.g.,  $10^{-6}$ ). The final values of  $\beta$  and  $\beta_1, \beta_2$  are taken as the median regression coefficient and the estimate of  $\sigma$  is given by (31). The position  $k$  of change-point is estimated by the SIC method denoted as  $\hat{k}^{(i)}$  in the  $i$ th replication.

Finally, the mean and standard difference of the estimated change-point position  $k$  are given by

$$\hat{k} = \frac{1}{500} \sum_{i=1}^{500} \hat{k}^{(i)}, \quad \text{sd} = \sqrt{\frac{1}{499} \sum_{i=1}^{500} (\hat{k}^{(i)} - \hat{k})^2}. \quad (40)$$

In order to compare the results of our proposed method with those of the normal method proposed in Chen [3], the same simulations are conducted using normal method. The results are presented in Tables 1 and 2, where  $k$  is the true value of change-point position in simulation, estimate. $L$  and diff. $L$  are the estimate of  $k$  and the absolute difference of the estimate and the true value of  $k$ , and finally, sd. $L$  is the standard difference of the estimate. The result by normal method is denoted as estimate. $N$ , diff. $N$ , and sd. $N$ , respectively. We compare diff. $L$  with diff. $N$  and sd. $L$  with

sd. $N$ , and the smaller the better. From Tables 1 and 2, we have the following findings.

- (1) In the  $N(0, 1)$  and  $L(0, 1)$  cases, both the Laplace method and normal method behave quite well with small diff. Besides that, the normal method is better than the Laplace method with smaller sd in the normal error case, and the result is inverse in the Laplace error case.
- (2) In the skew  $\chi^2(3)$  and heavy-tailed  $t(3)$ ,  $LN(0, 1)$ , and Cauchy cases, the Laplace method is better than the normal method with smaller diff. and smaller sd.
- (3) In the  $\chi^2(3)$  and  $LN(0, 1)$  cases, when the true position  $k$  of change-point is in the middle of the data (80, 100, 120), both of the two methods have good performance, but when  $k$  is in the lower part of the data (40, 60) or upper part (140, 160), the normal method can not detect the change-point effectively with large diff and large sd.
- (4) An interesting phenomenon appears in the case of Cauchy distribution, where the normal method simply uses integers around 100 to estimate the true values of  $k$  whenever  $k$  is small or large; on the contrary, the performance of the Laplace method is better relatively.

Another simulations design with  $\beta_1 = (1, 1, 1)^T$  and  $\beta_2 = (1, 2, 3)^T$  are conducted and the results are not presented in our paper. The results also justify our former findings; that is, our Laplace method is more robust than the normal method to estimate the position of mean change-point when the errors follow skew and heavy-tailed distributions, especially when the true position of change-point is in the head part or in the tail part of the data.

## 5. Application to Stock Market Data

Holbert [18] studied the switching simple linear regression models with changes in the coefficients from a Bayesian point of view. Later on, Chen [3] analyzed this data using the SIC method to detect changes in the mean for normal linear regression models, and Osorio and Galea [6] analyzed the same data also using the SIC method to detect mean change-point but assuming the errors that follow the Student  $t$ -distribution. In this section, we reanalyze the Holbert data using the proposed Laplace linear regression model to detect the mean change-point based on the SIC method. The monthly dollar volume of sales (in millions) on the Boston Stock Exchange (BSE) is considered as the response variable  $y$ , and the combined New York American Stock Exchange (NYAMSE) is considered as the regressor  $x$ . There are  $n = 35$  observations corresponding to the period between January 1967 and November 1969. The model we consider is that

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + \epsilon_i, & i = 1, \dots, k, \\ y_i &= \beta_0^* + \beta_1^* x_i + \epsilon_i, & i = k + 1, \dots, n, \end{aligned} \quad (41)$$

TABLE 1: Comparison of results for Laplace model and normal model based on 500 replications in Simulations 1–3.

Error distribution	$k$	Estimate. $L$	Estimate. $N$	diff. $L$	diff. $N$	sd. $L$	sd. $N$
$N(0, 1)$	40	40.07	40.11	0.07	0.11	1.92	1.77
	60	59.85	59.81	0.14	0.18	1.83	1.93
	80	79.92	79.91	0.07	0.08	1.88	1.85
	100	99.92	99.90	0.07	0.10	1.78	1.74
	120	120.00	119.97	0.00	0.02	1.78	1.74
	140	139.87	139.90	0.12	0.09	1.64	1.53
	160	159.92	159.85	0.08	0.14	1.86	1.77
$L(0, 1)$	40	39.98	38.89	0.02	0.11	2.50	2.96
	60	59.94	60.08	0.06	0.08	2.47	2.83
	80	80.06	80.14	0.06	0.14	2.95	3.19
	100	100.21	99.99	0.21	0.01	2.83	3.18
	120	120.03	120.07	0.03	0.07	2.23	2.91
	140	140.13	140.23	0.13	0.23	2.29	2.52
	160	160.03	159.81	0.03	0.19	2.55	5.29
$t(3)$	40	40.14	40.79	0.14	0.79	2.54	11.26
	60	59.96	61.10	0.04	1.10	2.51	13.22
	80	80.22	80.80	0.22	0.80	2.56	12.67
	100	100.05	99.32	0.05	0.68	2.99	12.68
	120	120.03	119.53	0.03	0.47	2.99	10.06
	140	140.11	139.51	0.11	0.49	2.82	12.15
	160	159.67	159.27	0.33	0.73	7.54	11.82

TABLE 2: Comparison of results for Laplace model and normal model based on 500 replications in Simulations 4–6.

Error distribution	$k$	Estimate. $L$	Estimate. $N$	diff. $L$	diff. $N$	sd. $L$	sd. $N$
$\chi^2(3)$	40	42.00	46.42	2.00	6.42	16.74	28.94
	60	60.67	61.58	0.67	1.58	9.62	19.85
	80	80.37	81.44	0.37	1.44	7.76	17.26
	100	99.18	99.03	0.82	0.97	8.76	14.99
	120	119.82	120.36	0.18	0.36	11.13	16.98
	140	139.26	137.25	0.74	2.75	10.98	22.95
	160	158.21	152.70	1.79	7.30	15.74	30.94
$LN(0, 1)$	40	41.23	49.78	1.23	9.78	13.49	38.92
	60	60.40	63.87	0.40	3.87	4.83	27.14
	80	80.58	83.08	0.58	3.08	4.66	25.60
	100	100.00	101.35	0.00	1.35	6.50	21.20
	120	120.44	117.74	0.44	2.26	6.03	24.65
	140	139.92	136.16	0.08	3.84	5.48	27.81
	160	158.80	152.27	1.20	7.73	13.71	35.13
Cauchy	40	51.75	98.10	11.75	58.10	46.99	67.04
	60	64.81	98.77	4.81	38.77	34.23	66.56
	80	81.63	81.44	1.63	1.44	7.76	17.26
	100	98.61	99.34	1.39	0.66	29.76	64.17
	120	119.83	105.36	0.17	14.64	28.86	65.80
	140	134.43	107.23	5.57	32.78	36.47	63.55
	160	149.95	105.12	10.05	54.90	46.53	66.94

TABLE 3: SIC calculated by Laplace method and normal method for Holbert data.

Time point	Calendar month	NYAMSE	BSE	SIC Laplace	SIC normal
1	Jan. 1967	10581.6	78.8	—	—
2	Feb. 1967	10234.3	69.1	364.2829	368.5739
3	Mar. 1967	13299.5	87.6	363.3368	367.8817
4	Apr. 1967	10746.5	72.8	363.3110	367.7757
5	May 1967	13310.7	79.4	361.7661	366.4980
6	Jun. 1967	12835.5	85.6	359.5049	365.7947
7	Jul. 1967	12194.2	75.0	357.4092	364.8795
8	Aug. 1967	12860.4	85.3	354.8215	363.9410
9	Sep. 1967	11955.6	86.9	<b>353.8327</b>	363.5574
10	Oct. 1967	13351.5	107.8	354.3397	363.5818
11	Nov. 1967	13285.9	128.7	357.2243	364.6607
12	Dec. 1967	13784.4	134.5	360.1891	365.4162
13	Jan. 1968	16336.7	148.7	361.0186	365.3077
14	Feb. 1968	11040.5	94.2	362.3622	365.5670
15	Mar. 1968	11525.3	128.1	363.4817	366.6527
16	Apr. 1968	16056.4	154.1	364.0982	366.8008
17	May 1968	18464.3	191.3	362.8076	366.9825
18	Jun. 1968	17092.2	191.9	360.4358	367.2177
19	Jul. 1968	15178.8	159.6	359.1133	367.3715
20	Aug. 1968	12774.8	185.5	359.5155	368.4097
21	Sep. 1968	12377.8	178.0	359.5954	368.3030
22	Oct. 1968	16856.3	271.8	356.8656	363.5156
23	Nov. 1968	14635.3	212.3	355.3476	<b>358.1847</b>
24	Dec. 1968	17436.9	139.4	357.0433	361.1139
25	Jan. 1969	16482.2	106.0	360.8022	364.8916
26	Feb. 1969	13905.4	112.1	360.8808	365.1567
27	Mar. 1969	11973.7	103.5	360.4217	365.0086
28	Apr. 1969	12573.6	92.5	360.9285	365.3012
29	May. 1969	16566.8	116.9	363.5041	367.3072
30	Jun. 1969	13558.7	78.9	364.3949	368.2468
31	Jul. 1969	11530.9	57.4	362.7110	368.2235
32	Aug. 1969	11278.0	75.9	360.3198	367.7685
33	Sep. 1969	11263.7	109.8	362.6290	368.1350
34	Oct. 1969	15649.5	129.2	—	—
35	Nov. 1969	12197.1	115.1	358.0474	361.4956

where

$$\epsilon_i \sim L(0, \sigma), \quad i = 1, 2, \dots, n. \quad (42)$$

The computed SIC values are listed in Table 3 along with the original BSE and NYAMSE values given in Holbert [18]. The SIC values calculated using the normal method by Chen [3] are also listed in the last column.

The bold SIC value in Table 3 is the minimum SIC value, which corresponds to time point 9, using the Laplace method and time point 23 using the normal method in Chen [3]. So, Chen [3] believed the regression model change starts at the time point 24, which is December 1968. Osorio and Galea [6] obtained the minimum values of SIC using the Student- $t$  model, which are at time point 9 for degree of freedom 1 and 23 for degree of freedom 4. Osorio and Galea [6] finally drew the same conclusion as Chen [3] that the regression model

TABLE 4: Minimum SIC values in Holbert data based on different models.

Method	Laplace	Normal	$t(4)$	$t(1)$
SIC( $n$ )	358.0474	361.4956	358.082	363.358
min SIC( $k$ )	353.8327	358.1847	355.035	357.416
$\hat{k}$	9	23	23	9

change starts at the time point 24. We list the minimum values of SIC in different regression models in Table 4. Table 4 tells us that the minimum SIC value is 353.8327 for time point 9 in our Laplace model; hence, we draw the conclusion that the regression model change starts at the time point 10, which is October 1967. The median estimates for the parameters are  $\hat{\beta}_0 = 10.500$ ,  $\hat{\beta}_1 = 0.0058$  for observations 1–9 and



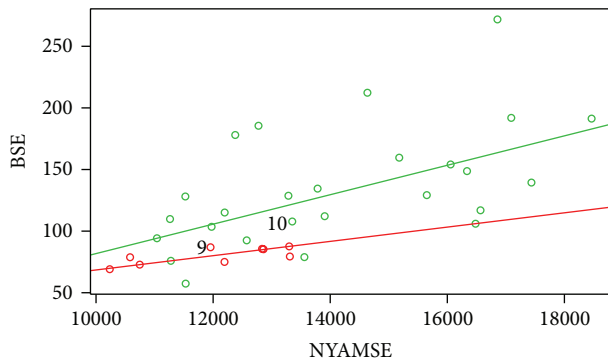


FIGURE 1: Scatter plot and regression lines of BSE versus NYAMSE. The red points and line correspond to the first 9 observations; the green points and line correspond to the 10–35 observations.

$\hat{\beta}_0^* = -37.646$ ,  $\hat{\beta}_1^* = 0.0119$  for observations 10–35. Finally, the predictive regression lines are plotted in Figure 1. The scatter plot shows that there are outliers in the response observations, so the median regression line estimated from Laplace model is more robust than the ordinary least square regression line estimated from the normal model.

## 6. Summary

In this paper, we proposed the Laplace linear regression model with a mean change-point and developed the EM algorithm with SIC model selection criterion to estimate the position of mean change-point. We investigated the performance of the algorithm for different simulations, finding that the algorithm behaved quite well when the errors follow the Laplace distribution. Besides that, our Laplace method is more robust than the normal method to estimate the position of mean change-point when the errors follow skew and heavy-tailed distributions, especially when the true position of change-point is in the head part or in the tail part of the data. Finally, we applied our method to the Holbert data and detected a mean change-point. Considering the difficulty in estimating the unknown degree of freedom in the Student  $t$ -distribution, we did not compare our Laplace model with the Student  $t$  model in Osorio and Galea [6], where the degree of freedom was predetermined by the authors. As to the multiple mean change-points problem in Laplace linear regression, the stepwise and dichotomy methods Vostrikova [19] may work and further research is needed in the future.

## Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

## References

[1] J. Chen and A. K. Gupta, *Parametric Statistical Change Point Analysis: with Applications to Genetics, Medicine, and Finance*, Birkhäuser, Boston, Mass, USA, 2nd edition, 2012.

[2] G. Schwarz, “Estimating the dimension of a model,” *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[3] J. Chen, “Testing for a change point in linear regression models,” *Communications in Statistics. Theory and Methods*, vol. 27, no. 10, pp. 2481–2493, 1998.

[4] J. Chen and A. K. Gupta, “Change point analysis of a Gaussian model,” *Statistical Papers*, vol. 40, no. 3, pp. 323–333, 1999.

[5] J. Chen and Y.-P. Wang, “A statistical change point model approach for the detection of DNA copy number variations in array CGH data,” *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 6, no. 4, pp. 529–541, 2009.

[6] F. Osorio and M. Galea, “Detection of a change-point in Student-t linear regression models,” *Statistical Papers*, vol. 47, no. 1, pp. 31–48, 2006.

[7] J. G. Lin, J. Chen, and Y. Li, “Bayesian analysis of student t linear regression with unknown change-point and application to stock data analysis,” *Computational Economics*, vol. 40, no. 3, pp. 203–217, 2012.

[8] S. Kotz, T. J. Kozubowski, and K. Podgorski, *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*, Birkhäuser, Boston, Mass, USA, 2001.

[9] E. Purdom and S. P. Holmes, “Error distribution for gene expression data,” *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, pp. 1–33, 2005.

[10] M.-I. Pop, “Distribution of the daily sunspot number variation for the last 14 solar cycles,” *Solar Physics*, vol. 276, no. 1-2, pp. 351–361, 2012.

[11] P. L. Noble and M. S. Wheatland, “Origin and use of the laplace distribution in daily sunspot numbers,” *Solar Physics*, vol. 282, no. 2, pp. 565–578, 2013.

[12] S. van Sanden and T. Burzykowski, “Evaluation of Laplace distribution-based ANOVA models applied to microarray data,” *Journal of Applied Statistics*, vol. 38, no. 5, pp. 937–950, 2011.

[13] R. F. Phillips, “Least absolute deviations estimation via the EM algorithm,” *Statistics and Computing*, vol. 12, no. 3, pp. 281–285, 2002.

[14] T. Park and G. Casella, “The Bayesian lasso,” *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, 2008.

[15] W. Song, W. Yao, and Y. Xing, “Robust mixture regression model fitting by Laplace distribution,” *Computational Statistics & Data Analysis*, vol. 71, pp. 128–137, 2014.

[16] D. F. Andrews and C. L. Mallows, “Scale mixtures of normal distributions,” *Journal of the Royal Statistical Society B*, vol. 36, no. 1, pp. 99–102, 1974.

[17] O. E. Barndorff-Nielsen and N. Shephard, “Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics,” *Journal of the Royal Statistical Society B Statistical Methodology*, vol. 63, no. 2, pp. 167–241, 2001.

[18] D. Holbert, “A Bayesian analysis of a switching linear model,” *Journal of Econometrics*, vol. 19, no. 1, pp. 77–87, 1982.

[19] L. J. Vostrikova, “Detecting disorder in multidimensional random processes,” *Soviet Mathematics Doklady*, vol. 24, no. 1, pp. 55–59, 1981.