

# Bayesian Segmentation of Piecewise Linear Regression Models Using Reversible Jump MCMC Algorithm

Suparman<sup>1</sup> and Michel Doisy<sup>2</sup>

1. Department of Mathematical Education, Faculty of Teacher Training and Education, Yogyakarta 55161, Indonesia

2. ENSEEIHT/TESA (Ecole Nationale Supérieure Electronique Electrotechnique Informatique Hydraulique Telecommunications), Toulouse 31071, France

**Abstract:** Piecewise linear regression models are very flexible models for modeling the data. If the piecewise linear regression models are matched against the data, then the parameters are generally not known. This paper studies the problem of parameter estimation of piecewise linear regression models. The method used to estimate the parameters of piecewise linear regression models is Bayesian method. But the Bayes estimator can not be found analytically. To overcome these problems, the reversible jump MCMC (Markov Chain Monte Carlo) algorithm is proposed. Reversible jump MCMC algorithm generates the Markov chain converges to the limit distribution of the posterior distribution of the parameters of piecewise linear regression models. The resulting Markov chain is used to calculate the Bayes estimator for the parameters of piecewise linear regression models.

**Key words:** Piecewise linear regression models, hierarchical bayesian, reversible jump MCMC.

## 1. Introduction

Piecewise linear regression models are a model that is often used in many fields. For example, it is used in the field of econometrics [1], geophysics [2], health [3], and ecology [4]. In the field of econometrics, piecewise linear regression models used to model the commission. For example, a company pays a commission to the sales clerk. The company was paid a commission based on the sales in a way such that up to a certain level, called the threshold one and after a commission structure on top of earlier structures commissions.

If the piecewise linear regression models are matched against the data, then the model parameters are generally unknown. There are many piecewise linear regression models. In this paper, the error distribution for each piece will be assumed and has the gaussian distribution with mean 0 and variance  $\sigma^2$ .

For  $t = 1, 2, \dots, n$ , let  $y_t$  be a dependent variable

and let  $x_t$  be independent variable. Then the piecewise linear regression models can be written in the following equation:

$$y_t = \begin{cases} \alpha_1^{(k)} + \beta_1^{(k)} x_t + z_t & \tau_0^{(k)} < t \leq \tau_1^{(k)} \\ \alpha_2^{(k)} + \beta_2^{(k)} x_t + z_t & \tau_1^{(k)} < t \leq \tau_2^{(k)} \\ \vdots & \\ \alpha_k^{(k)} + \beta_k^{(k)} x_t + z_t & \tau_{k-1}^{(k)} < t \leq \tau_k^{(k)} \end{cases}$$

with

$$\tau_0^{(k)} = 0, \quad \tau_k^{(k)} = n$$

and

$$\begin{aligned} z_t &\sim N(0, \sigma_1^{2(k)}) & \tau_0^{(k)} \leq t < \tau_1^{(k)} \\ z_t &\sim N(0, \sigma_2^{2(k)}) & \tau_1^{(k)} \leq t < \tau_2^{(k)} \\ &\vdots & \\ z_t &\sim N(0, \sigma_k^{2(k)}) & \tau_{k-1}^{(k)} \leq t < \tau_k^{(k)} \end{aligned}$$

Fig. 1 shows the graph of the four piecewise linear regression.

In the above equation: (a)  $k$  is the number of threshold point, (b)  $\tau_0^{(k)}, \tau_1^{(k)}, \dots, \tau_k^{(k)}$  are the point corresponding threshold, (c)  $\alpha_1^{(k)}, \dots, \alpha_k^{(k)}$  and  $\beta_1^{(k)}, \dots, \beta_k^{(k)}$  are the regression coefficients, (d)  $\sigma_1^{2(k)}, \sigma_2^{2(k)}, \dots, \sigma_k^{2(k)}$  are the error variance. If  $\theta$  is

**Corresponding author:** Suparman, Dr, research fields: time series, bayesian, reversible jump MCMC, signal processing. E-mail: suparmancict@yahoo.co.id.

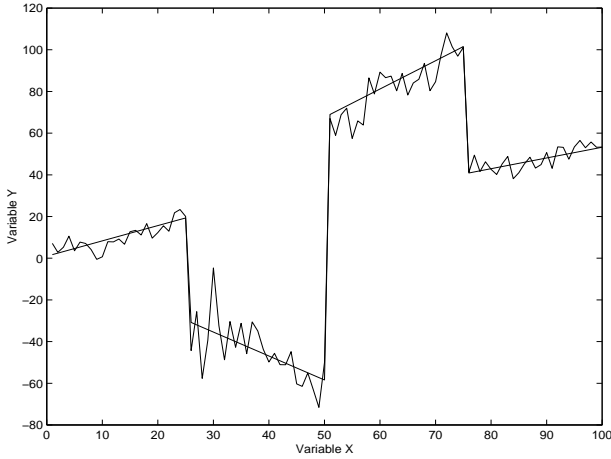


Fig. 1 Four piecewise linear regression.

the parameter of the piecewise linear regression models above, then

$$\theta = (k, \tau_1^{(k)}, \dots, \tau_k^{(k)}, \alpha_1^{(k)}, \dots, \alpha_k^{(k)}, \beta_1^{(k)}, \dots, \beta_k^{(k)}, \sigma_1^{2(k)}, \dots, \sigma_k^{2(k)})$$

Suppose  $x_1, x_2, \dots, x_n$  are a random sample drawn from a population having a piecewise linear regression models. Based on the random sample, the main problem is how to estimate the parameters  $\theta$ . Parameter  $\theta$  is estimated using Bayesian method. The study of the Bayesian method can be found in the literature, for example [5]. Parameter estimation using the Bayesian method can not be determined analytically because the likelihood function for the parameter  $\theta$  has a complicated shape. To overcome these problems, in this study Reversible Jump MCMC Algorithm is used.

## 2. Maximum Likelihood Function

For  $i = 1, 2, \dots, k$  and  $\tau_{i-1}^{(i)} < t \leq \tau_i^{(i)}$ , suppose  $z_t$  has a Gaussian distribution with mean 0 and variance  $\sigma_i^{2(i)}$ . Then the density function of  $z_t$  is

$$f(z_t | \sigma_i^{2(i)}) = \left( \frac{1}{\sqrt{2\pi\sigma_i^{2(i)}}} \right) \exp - \frac{1}{2\sigma_i^{2(i)}} z_t^2$$

Let  $z_i = (z_{\tau_{i-1}^{(i)}+1}, \dots, z_{\tau_i^{(i)}})$ , then it has a joint density function as follow:

$$f(z_i | \sigma_i^{2(i)}) = \left( \frac{1}{\sqrt{2\pi\sigma_i^{2(i)}}} \right)^{\tau_i^{(i)} - \tau_{i-1}^{(i)}}$$

$$\exp - \frac{1}{2\sigma_i^{2(i)}} \sum_{t=\tau_{i-1}^{(i)}}^{\tau_i^{(i)}} z_t^2$$

By using variable transformation

$$y_t = \alpha_i^{(k)} + \beta_i^{(k)} x_t + z_t,$$

then  $z_t = y_t - \alpha_i^{(k)} - \beta_i^{(k)} x_t$  and  $\frac{dz_t}{dy_t} = 1$ . So

that the density function of  $y_i$  is

$$f(y_i | \alpha_i^{(i)}, \beta_i^{(i)}, \sigma_i^{2(i)}) = \left( 2\pi\sigma_i^{2(i)} \right)^{-\frac{1}{2}(\tau_i^{(i)} - \tau_{i-1}^{(i)})} \exp - \frac{1}{2\sigma_i^{2(i)}} \sum_{t=\tau_{i-1}^{(i)}}^{\tau_i^{(i)}} (y_t - \alpha_i^{(k)} - \beta_i^{(k)} x_t)^2$$

Finally, the maximum likelihood function for  $y = (y_1, y_2, \dots, y_n)$  is as follow:

$$L(y | \theta) = \prod_{i=1}^k \prod_{t=\tau_{i-1}^{(k)}+1}^{\tau_i^{(k)}} \frac{1}{\sqrt{2\pi\sigma_i^{2(k)}}} \exp - \frac{1}{2\sigma_i^{2(k)}} (y_t - \alpha_i^{(k)} - \beta_i^{(k)} x_t)^2$$

or

$$L(y | \theta) = \prod_{i=1}^k \left\{ \left( 2\pi\sigma_i^{2(k)} \right)^{-\frac{1}{2}(\tau_i^{(k)} - \tau_{i-1}^{(k)})} \exp - \frac{1}{2\sigma_i^{2(k)}} \sum_{t=\tau_{i-1}^{(k)}}^{\tau_i^{(k)}} (y_t - \alpha_i^{(k)} - \beta_i^{(k)} x_t)^2 \right\}$$

## 3. Prior Distribution

To obtain the posterior distribution, first it must be determined the prior distribution of parameter

$$\theta = (k, \tau^{(k+1)}, \alpha^{(k)}, \beta^{(k)}, \sigma^{2(k)}),$$

as follow [6]:

$$\pi(k) = C_k^{k_{maks}} \lambda^k (1 - \lambda)^{k_{maks} - k}$$

for  $k = 1, 2, \dots, k_{maks}$ .

$$\pi(\tau^{(k)} | k) = \frac{(2k+1)!}{n^{2k}} \frac{1}{2^k} \prod_{i=1}^k (\tau_i - \tau_{i-1})$$

$$\pi(\alpha^{(k)} | k) = \prod_{i=1}^k \frac{1}{\sqrt{2\pi a^2}} \exp - \frac{1}{2a^2} \alpha_i^2$$

$$\pi(\beta^{(k)} | k) = \prod_{i=1}^k \frac{1}{\sqrt{2\pi b^2}} \exp - \frac{1}{2b^2} \beta_i^2$$

$$\pi(\sigma^{2(k)} | k) = \prod_{i=1}^k \frac{a}{2} (\sigma_i^{2(i)})^{-2} \exp - \frac{a}{2\sigma_i^{2(i)}}$$

Let  $\varphi = (\lambda, c, d, a, b)$  is hyperparameter of the prior distributions. Generally this hyperparameter is unknown. Furthermore hyperparameter  $\varphi$  viewed as a random variable with a certain distribution, i.e.:

$$\begin{aligned}\pi(\lambda) &\propto \frac{1}{\lambda} & \pi(c) &\propto \frac{1}{c} \\ \pi(a) &\propto \frac{1}{a} & \pi(b) &\propto \frac{1}{b}\end{aligned}$$

Suppose  $\pi(\theta, \varphi)$  expressed a prior distribution for  $(\theta, \varphi)$ . Because  $\pi(\theta | \varphi) = \frac{\pi(\theta, \varphi)}{\pi(\varphi)}$  then the prior distribution of  $(\theta, \varphi)$  can be determined as follows

$$\pi(\theta, \varphi) = \pi(\theta | \varphi) \pi(\varphi)$$

or

$$\begin{aligned}\pi(\theta, \varphi) &\propto C_k^{k_{maks}} \lambda^k (1-\lambda)^{k_{maks}-k} \frac{(2k+1)!}{n^{2k}} \\ &\quad \frac{1}{2^k} \prod_{i=1}^k (\tau_i - \tau_{i-1}) \\ &\quad \prod_{i=1}^k \frac{1}{\sqrt{2\pi a^2}} \exp\left(-\frac{1}{2a^2} \alpha_i^2\right) \\ &\quad \prod_{i=1}^k \frac{1}{\sqrt{2\pi b^2}} \exp\left(-\frac{1}{2b^2} \beta_i^2\right) \\ &\quad \prod_{i=1}^k \frac{a}{2} (\sigma_i^{2(i)})^{-2} \exp\left(-\frac{a}{2\sigma_i^{2(i)}}\right) \\ &\quad \frac{1}{\lambda} \frac{1}{c} \frac{1}{a} \frac{1}{b}\end{aligned}$$

#### 4. Posterior Distribution

Suppose  $\pi(\theta, \varphi | y)$  is a posterior distribution for  $\theta$ . By using Bayes theorem  $(\theta, \varphi)$ , then the posterior distribution for the parameter  $(\theta, \varphi)$  can be expressed as the product of the likelihood function and the prior distribution

$$\begin{aligned}\pi(\theta, \varphi | y) &\propto f(y | \theta) \pi(\theta, \varphi) \\ &\propto f(y | \theta) \pi(\theta, \varphi) \pi(\theta | \varphi) \pi(\varphi) \\ &\propto \prod_{i=1}^k \left\{ (\sigma_i^{2(k)})^{-\frac{1}{2}(\tau_i^{(k)} - \tau_{i-1}^{(k)})} \right\}\end{aligned}$$

$$\begin{aligned}&\exp\left\{-\frac{1}{2\sigma_i^{2(k)}} \sum_{t=\tau_{i-1}^{(k)}}^{\tau_i^{(k)}} (y_t - \alpha_i^{(k)} - \beta_i^{(k)} x_t)^2\right\} \\ &\quad C_k^{k_{maks}} \lambda^k (1-\lambda)^{k_{maks}-k} \frac{(2k+1)!}{n^{2k}} \\ &\quad \prod_{i=1}^k \frac{1}{\sqrt{a^2}} \exp\left(-\frac{1}{2a^2} \alpha_i^2\right) \\ &\quad \prod_{i=1}^k \frac{1}{\sqrt{b^2}} \exp\left(-\frac{1}{2b^2} \beta_i^2\right) \\ &\quad \prod_{i=1}^k \frac{a}{2} (\sigma_i^{2(i)})^{-2} \exp\left(-\frac{a}{2\sigma_i^{2(i)}}\right) \\ &\quad \frac{1}{\lambda} \frac{1}{c} \frac{1}{a} \frac{1}{b}\end{aligned}$$

#### 5. Reversible Jump MCMC Algorithm

Suppose  $M = (\theta, \varphi)$ . In general, The MCMC algorithm is a method of sampling to make a homogeneous Markov Chain  $M_1, M_2, \dots, M_m$  that satisfies aperiodic and irreducible [7] such that  $M_1, M_2, \dots, M_m$  can be considered as a random variable whose distribution  $\pi(\theta, \varphi | y)$ . Thus  $M_1, M_2, \dots, M_m$  it can be used to estimate parameters  $M$ . To realize this, the Gibbs algorithm is adopted [7] which consists of two phases :

- (1) Simulate  $\pi(\varphi | \theta, y)$
- (2) Simulate  $\pi(\theta | \varphi, y)$

Distribution  $\pi(\varphi | \theta, y)$  has an explicit form, so that the Gibbs algorithm can be used to simulate the distribution of  $\pi(\varphi | \theta, y)$ . On the contrary, the distribution  $\pi(\theta | \varphi, y)$  has not an explicit form. So the exact simulation is possible. The solution is to use a hybrid algorithm consisting of three stages as follows :

- (2.1) Simulate  $\pi(\sigma^{2(k)} | k, \tau^{(k)}, \varphi, y)$
- (2.2) Simulate  $\pi(k, \tau^{(k)} | \varphi, y)$

The distribution  $\pi(\sigma^{2(k)} | k, \tau^{(k)}, \varphi, y)$  has the form explicit, so that the Gibbs algorithm can be used to simulate  $\pi(\sigma^{2(k)} | k, \tau^{(k)}, \varphi, y)$ . On the contrary,

because the value  $k$  is not known then the MCMC algorithm can not be used to simulate the distribution  $\pi(k, \tau^{(k)} | \varphi, y)$ . Here, reversible jump MCMC algorithm [8] is used to simulate  $\pi(k, \tau^{(k)} | \varphi, y)$ .

Let  $\omega = (k, \tau)$  is the actual point of the Markov chain. There are three types of transformations are used, namely: the birth of the threshold point, the death of the threshold point and the change of the threshold point. Further suppose that  $N_k$  is the probability of transformation from  $k$  to  $k + 1$ ,  $D_k$  is the probability of transformation from  $k + 1$  to  $k$ , and  $P_k$  is the probability of transformation from  $k$  to  $k$ .

### 5.1 Birth/Death of the Threshold Point

The transdormstion of the birth of the threshold will change the number of threshold point, from  $k$  to the  $k + 1$ . If the birth of the threshold is selected, then the birth of the threshold from a point  $\omega = (k, \tau)$  is defined in the following way: Choose a random point  $z$  in the  $\{1, \dots, n-1\} \setminus \tau$ . Suppose the point  $z$  in the  $\{\tau_i + 1, \dots, \tau_{i+1} - 1\}$ . Next, create a new point  $\omega^* = (k + 1, \tau)$  with

$$\tau_1, \dots, \tau_i, \tau_{i+1} = z, \tau_{i+2}, \dots, \tau_{k+2}$$

Otherwise, the transdormstion of the birth of the threshold will change the number of threshold point, from  $k + 1$  to the  $k$ . If the death of the threshold is selected, then the death of the threshold from a point  $\omega^* = (k + 1, \tau)$  is defined in the following way: Choose randomly a point in  $\tau$ . Suppose then that point is  $\tau_{i+1}$ . Next, create a new point  $\omega = (k, \tau)$  with

$$\tau_1, \dots, \tau_i, \tau_{i+2}, \dots, \tau_{k+1}$$

Suppose that  $a_n$  and  $a_d$  are respectively a probability of acceptance for birth and death. Then the probability of acceptance for birth is as follows:

$$a_n(\omega, \omega^*) = \min \left\{ 1, \frac{\pi(\omega^* | \varphi, y)}{\pi(\omega | \varphi, y)} \frac{q(\omega, \omega^*)}{q(\omega^*, \omega)} \right\}$$

While the probability of death is as follows:

$$a_d(\omega, \omega^*) = \min \left\{ 1, \frac{1}{a_n(\omega^*, \omega)} \right\}$$

where

$$\frac{q(\omega^*, \omega)}{q(\omega, \omega^*)} = \frac{D_{k+1}}{N_k} \frac{n-1-k}{k+1}$$

### 5.2 Change of Threshold Point

The transformation of the change of threshold will not change the number of threshold point, but this transformation will change the position of the threshold point. If the change of the threshold is selected, then the change of the threshold point from a  $\omega = (k, \tau)$  is defined in the following way: Choose a random point in  $\tau$ . Suppose that point is  $\tau_i$ . Next, create a new point  $\omega^* = (k, \tau)$  where this point  $\tau_i$  is replaced with  $z$  generated from the uniform distribution on the set  $\{1, \dots, n-1\} \setminus \tau$ .

Let  $a_p$  is the probability of acceptance to the change. Then the probability of acceptance for change is as follows:

$$a_p(\omega, \omega^*) = \min \left\{ 1, \frac{\pi(\omega^* | \varphi, y)}{\pi(\omega | \varphi, y)} \frac{q(\omega, \omega^*)}{q(\omega^*, \omega)} \right\}$$

where

$$\frac{q(\omega^*, \omega)}{q(\omega, \omega^*)} = 1$$

## 6. Conclusion

The purpose of this study is to examine how to estimate the parameters of piecewise linear regression models when the number of regression is unknown. If the number of regression is unknown, the estimated parameters cannot be done by Markov chain Monte Carlo algorithm.

The reversible jump Markov chain Monte Carlo algorithm is one of the new methods that can be used to estimate the parameters of piecewise linear regression models although the number of regression is unknown. The advantages of this method are both the number of regression and the estimation of parameter of linear regression models per piece which can be estimated simultaneously.

## Acknowledgements

This research activity will not be able to work without the financial support of the Directorate of Research and Community Service, the Directorate General of Higher Education Ministry of Education and Culture, Republic of Indonesia, which has provided research grants through the scheme of Fundamental Research in 2014 with the Number of Research Contract: PF.02/LPP-UAD/V/2014. The authors would like to thank profusely for these research grants.

## Reference

- [1] Gujarati, D. N. 2006 *Essentials of Econometrics*. New York: McGraw-Hill.
- [2] Stewart, D. N., and Whaler, K. A. 1995. "Optimal Piecewise Regression Analysis and Its Application to Geomagnetic Time Series." *Geophysical Journal International* 121 (3):710-24.
- [3] Shi, H. Y., Lee, H. H., Tsai, M. H., Chiu, C. C., Uen, Y. H., and Lee, K. T. 2011. "Predicting Long-Term Outcomes of Laparoscopic Cholecystectomy: A Prospective Piecewise Linear Regression Analysis." *Surg Endosc* 25(7): 2132-40.
- [4] Toms, J. D., and Lesperance, M. L. 2003. "Piecewise Regression: A Tool Identifying Ecological Thresholds." *Ecology* 84: 2034-41.
- [5] Robert, C. P. 2002. *The Bayesian Choice: A Decision-Theory Motivation*. New York: Springer.
- [6] Suparman, S., Doisy, M., and Tourneret, J. Y. 2002. "Changeoint Detection Using Reversible Jump MCMC Methods." In *Proceedings of IEEE ICASSP*, 1569-72.
- [7] Robert, C. P., and Casella, G. 1999. *Monte Carlo Statistical Methods*. New York: Springer.
- [8] Green, P. J. 1995. "Reversible Jump MCMC Computation and Bayesian Model Determination." *Oxford Journals* 82 (4): 711-32.