International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:10, No:5, 2016

Segmentation of Piecewise Polynomial Regression Model by Using Reversible Jump MCMC Algorithm

Suparman

***Abstract*—**Piecewise polynomial regression model is very flexible model for modeling the data. If the piecewise polynomial regression model is matched against the data, its parameters are not generally known. This paper studies the parameter estimation problem of piecewise polynomial regression model. The method which is used to estimate the parameters of the piecewise polynomial regression model is Bayesian method. Unfortunately, the Bayes estimator cannot be found analytically. Reversible jump MCMC algorithm is proposed to solve this problem. Reversible jump MCMC algorithm generates the Markov chain that converges to the limit distribution of the posterior distribution of piecewise polynomial regression model parameter. The resulting Markov chain is used to calculate the Bayes estimator for the parameters of piecewise polynomial regression model.

***Keywords*—**Piecewise, Bayesian, reversible jump MCMC, segmentation.

I. Introduction

PIECEWISE polynomial regression model is a model that is often used in many fields. For example, it is used in the field of geomagnetic [1], health [2], ecology [3] and genome [4].

If the piecewise linear regression models are matched against the data, the model parameters will be generally unknown. The objective of this paper is to estimate the parameters of the piecewise polynomial regression models. There are so many piecewise polynomial regression models. In this paper, the noise distribution for each piece will be assumed has the Gaussian distribution with mean 0 and variance unknown.

Let y_t be a dependent variable and t be an independent variable with $t = 1, 2, \cdots, n$ and n is the number of samples. A piecewise polynomial regression models can be written as:

$$yt = \begin{cases} \alpha_{11} t^{p_1} + \cdots + \alpha_{1p_1} t + \alpha_{1(p_i+1)} + zt, & \tau_0 < t \le \tau_1 \\ \alpha_{21} t^{p_2} + \cdots + \alpha_{2p_2} t + \alpha_{2(p_2+1)} + zt, & \tau_1 < t \le \tau_2 \\ \vdots & \\ \alpha_{k1} t^{p_k} + \cdots + \alpha_{ap_k} + \alpha_{k(p_k+1)} + zt, & \tau_{k-1} < t \le \tau_k \end{cases} \tag{1}$$

with $\tau_0 = 0$ and $\tau_k = n$ where

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

For example, Fig. 1 shows the graph of the piecewise polynomial regression with threshold at 100, 175 and 300.

In the above two equations: (a) k is the number of pieces, (b) $\tau = (\tau_0, \tau_1, \cdots, \tau_k)$ is the threshold vector, (c) α_{ij} $(i = 1, \cdots, k; j = 1, \cdots, p_i + 1)$ are the regression coefficients, (d) $p = (p_1, \cdots, p_k)$ is the regression order vector, (e) $\sigma^2 = (\sigma_1^2, \sigma_2^2, \cdots, \sigma_k^2)$ is the noise variance. Let Ω be the regression coefficient matrix.

$$\Omega = \begin{pmatrix} \alpha_{11} \cdots \alpha_{1(p_1+1)} \\ \alpha_{21} \cdots \alpha_{1(p_2+1)} \\ \vdots \quad \vdots \quad \vdots \\ \alpha_{k1} \cdots \alpha_{1(p_k+1)} \end{pmatrix} \tag{3}$$

Let θ be the parameter of the above piecewise polynomial regression model. Then this parameter θ can be written as:

$$\theta = (k, \tau, p, \Omega, \sigma^2) \tag{4}$$

Suppose that $y_t\,(t = 1, \cdots, n)$ is a random sample drawn from a population having a piecewise polynomial regression models. Based on the random sample, the main problem is how to estimate the parameter θ. Here, the parameter θ is estimated by using Bayesian method. The study of the Bayesian method can be found in the literature, for example [5]. Unfortunately, the Bayes estimator cannot be determined analytically because the likelihood function of the parameter θ has a complicated form. To overcome these problems, reversible jump MCMC Algorithm [6] is used.

Suparman is with the Mathematics Education Department, Faculty of Teacher Training and Education, Ahmad Dahlan University, Yogyakarta, Indonesia (phone: +6281328201198; e-mail: suparmancict@yahoo.co.id).

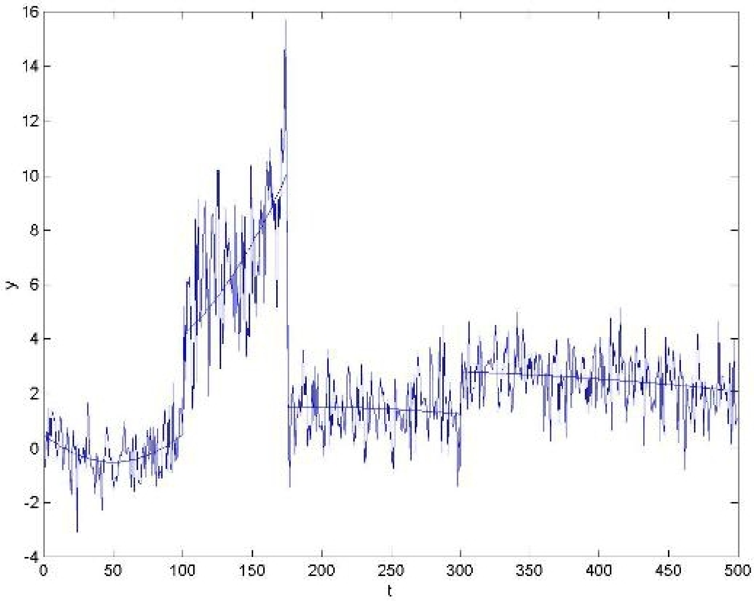

Fig. 1 Four piecewise polynomial regression

II. Method

A. Maximum Likelihood Function

Because the variable random z_t has a Gaussian distribution with mean 0 and variance σ_i^2 for $i = 1, 2, \cdots, k$ and $\tau_{i-1} < t \le \tau_i$, the density function of z_t is

$$f(z_t | \sigma_i^2) = \left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right) \exp - \frac{1}{2\sigma_i^2} z_t^2 \tag{5}$$

Let $z_i = (z_{\tau_{i-1}+1}, \cdots, z_{\tau_i})$ be the noise of the piece $[\tau_{i-1}+1 \quad \tau_i]$. Then the z_i has a joint density function is:

$$f(z_i | \tau_{i-1}, \tau_i, \sigma_i^2) = \left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right)^{\tau_i - \tau_{i-1}} \exp - \frac{1}{2\sigma_i^2} \sum_{t=\tau_{i-1}}^{\tau_i} z_t^2 \tag{6}$$

The variable transformation:

$$y_t = \alpha_{i1} t^{p_1} + \cdots + \alpha_{ip_1} t + \alpha_{i(p_i+1)} + z_t \tag{7}$$

is used. Then $z_t = y_t - \alpha_{i1} t^{p_1} - \cdots - \alpha_{ip_i} t - \alpha_{i(p_i+1)}$ and $\frac{dz_t}{dy_t} = 1$. Let $\Omega_i = (\alpha_{i1}, \cdots, \alpha_{i(p_i+1)})$ be the regression coefficient vector. Thus, the density function of y_t is

$$f(y_t | \tau_{i-1}, \tau_i, p_i, \Omega_i, \sigma_i^2) = \left(2\pi\sigma_i^2\right)^{-\frac{1}{2}(\tau_i - \tau_{i-1})} \exp - \frac{1}{2\sigma_i^2} \sum_{t=\tau_{i-1}+1}^{\tau_i} \left(y_t - \alpha_{i1} t^{p_1} - \cdots - \alpha_{ip_i} t - \alpha_{i(p_i+1)}\right)^2 \tag{8}$$

Finally, the maximum likelihood function of the $y = (y_1, y_2, \cdots, y_n)$ can be expressed in:

$$L(y | \theta) = \prod_{i=1}^{k} \left\{ (2\pi\sigma_i^2)^{-\frac{1}{2}(\tau_i - \tau_{i-1})} \exp - \frac{1}{2\sigma_i^{2(k)}} \sum_{t=\tau_{i-1}+1}^{\tau_i} \left(y_t - \alpha_{i1} t^{p_1} - \cdots - \alpha_{ip_i} t - \alpha_{i(p_i+1)}\right) \right\}^2 \tag{9}$$

B. Prior Distribution

To obtain a posterior distribution, a prior distribution must be determined. As in [7] and [8], the prior distribution is chosen as follows:

A binomial distribution is chosen for the number of pieces k ($k = 1, 2, \cdots, k_{max}$).

$$\pi(k|\mu) = C_k^{k_{max}} \mu^k (1-\mu)^{k_{max} - k} \tag{10}$$

where k_{max} is the maximum of k and μ is a hyper parameter. The prior distribution for τ given k is chosen as follows:

$$\pi(\tau \mid k) = \frac{(2k+1)!}{n^{2k}} \frac{1}{2^k} \prod_{i=1}^{k} (\tau_i - \tau_{i-1}) \quad (11)$$

A binomial distribution is chosen for the number of order p ($p = 1, 2, \cdots, p_{max}$).

$$\pi(p|\varphi) = C_p^{p_{max}} \lambda^p (1-\lambda)^{p_{max}-p} \quad (12)$$

where p_{max} is the maximum of p and λ is a hyper parameter. A non-informative prior distribution for α and σ_i^2 is:

$$\pi(\alpha, \sigma^2 \mid k, \tau) \propto \prod_{i=1}^{k} \sigma_i^{-2}. \quad (13)$$

Furthermore, the hyperprior distribution for μ is:

$$\pi(\mu) \propto [\mu(1-\mu)]^{-1/2}. \quad (14)$$

The hyperprior distribution for λ is:

$$\pi(\lambda) \propto [\lambda(1-\lambda)]^{-1/2}. \quad (15)$$

Let $\varphi = (\mu, \lambda)$ be the hyper parameter vector. Let $\pi(\theta, \varphi)$ be a prior distribution for (θ, φ). Because the distribution of θ given φ is $\pi(\theta \mid \varphi) = \frac{\pi(\theta, \varphi)}{\pi(\varphi)}$, the prior distribution for (θ, φ) can be written as

$$\pi(\theta, \varphi) = \pi(\theta \mid \varphi)\pi(\varphi) \quad (16)$$

C. Posterior Distribution

Let $\pi(\theta, \varphi \mid y)$ be a posteriori distribution for (θ, φ). According to the Bayes Theorem, the posterior distribution for (θ, φ) is given as

$$\pi(\theta, \varphi \mid y) \propto f(y \mid \theta)\pi(\theta, \varphi)$$

$$\propto f(y \mid \theta)\pi(\theta, \varphi)\ \pi(\theta \mid \varphi)\pi(\varphi) \quad (17)$$

D. Reversible Jump MCMC

Suppose that $M = (\theta, \varphi)$. A MCMC method for the simulation of a distribution $\pi(\theta, \varphi \mid y)$ is any method producing an ergodic Markov chain $M_1, M_2, \cdots, M_m$ whose stationary distribution $\pi(\theta, \varphi \mid y)$ [9]. This Markov chain $M_1, M_2, \cdots, M_m$ can be considered as a random variable whose distribution $\pi(\theta, \varphi \mid y)$. Furthermore, the Markov chain $M_1, M_2, \cdots, M_m$ is used to estimate the parameter M. To realize this, the Gibbs sampling algorithm is adopted. It consists of five steps:

1) Simulate $\mu \sim B(k + \frac{1}{2}, k_{max} - k + \frac{1}{2})$
2) Simulate $\lambda \sim B(p + \frac{1}{2}, p_{max} - p + \frac{1}{2})$
3) Simulate $\sigma_i^2 \sim IG(\frac{n-p}{2}, \frac{n-p}{2} s^2)$ where $s^2 = \frac{1}{n-p} y'[I - x(x'x)^{-1}x']y$.
4) Simulate $\Omega_i \sim N((x'x)^{-1}y, \sigma_i^2 (x'x)^{-1})$
5) Simulate $(k, \tau, p) \sim \pi(k, \tau \mid p, \Omega, \sigma^2, \lambda, \mu, y)$

Unfortunately, the distribution $\pi(k, \tau, p \mid p, \Omega, \sigma^2, \lambda, \mu, y)$ have not an explicit form. The exact simulation is not possible to be done. Since the value k is not known, the MCMC algorithm cannot be used to simulate $\pi(k, \tau \mid p, \Omega, \sigma^2, \lambda, \mu, y)$. Here, reversible jump MCMC algorithm [6] is adopted.

Let $\omega = (k, \tau, p)$ be an actual point of the Markov chain. There are 3 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. Furthermore, let N_k be the probability of transformation from k to k + 1, let D_k be the probability of transformation from k + 1 to k, and let P_k be the probability of transformation from k to k.

1. Birth/Death of The Threshold Point

The transformation 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, the birth of the threshold from a point $\omega = (k, \tau, p)$ is defined in the following way. Choose a random point z from a set $\{1, \cdots, n-1\} \setminus \tau$. Suppose that the point z is on the interval $[\tau_i + 1 \quad \tau_{i+1} - 1]$. Next, create a new point $\omega^* = (k+1, \tau, p)$ with:

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

The order p_i is replaced by two orders p_i^* and p_{i+1}^* according to this formula [10]:

$$\begin{cases} p_i^* = u \\ p_{i+1}^* = p_i - p_i^* \end{cases}$$

where $u \sim U\{0, \cdots, p_i\}$. The coefficient vector Ω_i is replaced by $\Omega_i \sim N((x'x)^{-1}y, \sigma_i^2 (x'x)^{-1})$ with $\sigma_i^2 \sim IG(\frac{n-p}{2}, \frac{n-p}{2} s^2)$. The coefficient vector Ω_{i+1} is replaced by $\Omega_{i+1} \sim N((x'x)^{-1}y, \sigma_{i+1}^2 (x'x)^{-1})$ with $\sigma_{i+1}^2 \sim IG(\frac{n-p}{2}, \frac{n-p}{2} s^2)$. Otherwise, the transformation of the birth of the threshold will change the number of threshold point, from k+1 to k. If this transformation is selected, the death of the threshold from a point $\omega^* = (k+1, \tau, p)$ is defined in the following way: Choose randomly a point from τ. Suppose that this point is τ_{i+1}. Next, create a new point $\omega = (k, \tau, p)$ with:

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

Two orders p_i^* and p_{i+1}^* are replaced by:

$$p_i = p_{i+1}^* + p_i^*.$$

The coefficient vector Ω_i is replaced by $\Omega_i \sim N((x'x)^{-1}y, \sigma_i^2(x'x)^{-1})$ with $\sigma_i^2 \sim IG(\frac{n-p}{2}, \frac{n-p}{2}s^2)$.

Suppose that a_n and a_d are respectively a probability of acceptance for birth and death. The probability of acceptance for birth is as:

$$a_n(\omega, \omega^*) = \min\left\{1, \frac{\pi(\omega^* | \varphi, y)}{\pi(\omega | \varphi, y)} \frac{q(\omega^*, \omega)}{q(\omega, \omega^*)}\right\} \quad (18)$$

While the probability of death is as:

$$a_d(\omega, \omega^*) = \min\left\{1, \frac{1}{a_n(\omega^*, \omega)}\right\} \quad (19)$$

where;

$$\frac{q(\omega^*, \omega)}{q(\omega, \omega^*)} = \frac{D_{k+1}}{N_k} \frac{n-1-k}{k+1} \quad (20)$$

2. Change of The Threshold Point

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

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

$$a_p(\omega, \omega^*) = \min\left\{1, \frac{\pi(\omega^* | \varphi, y)}{\pi(\omega | \varphi, y)} \frac{q(\omega^*, \omega)}{q(\omega, \omega^*)}\right\} \quad (21)$$

where;

$$\frac{q(\omega^*, \omega)}{q(\omega, \omega^*)} = 1. \quad (22)$$

III. Conclusion

The purpose of this paper was to estimate the parameters of piecewise polynomial regression models when the number of regression is unknown. The parameters cannot be estimated by Markov chain Monte Carlo algorithm, because the number of regression is unknown.

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

References

[1] D.N. Stewart and K.A. Whaler, Optimal Piecewise Regression Analysis and Its Application to Geomagnetic Time Series, Geophysical Journal International, 1995, 710-724.
[2] H.-Y. Shi, H.-H. Lee, M.-H. Tsai, C.-C. Chiu, Y.-H. Uen and K.-T. Lee, Long-term outcomes of laparoscopic cholecystectomy: a prospective piecewise linear regression analysis, Surg Endosc, 2011, 2132-2140.
[3] J.D. Toms and M.L. Lesperance, Piecewise Regression: A Tool Identifying Ecological Thresholds, Ecology, 2003, 2034-2041.
[4] M. Tarabichi, V. Detours, and T. Konopka. Piecewise Polynomial Representations of Genome Data, Plos One, 2012, 1-10.
[5] C.P. Robert, The Bayesian Choice: A Decision-Theory Motivation, Springer, New York, 2001.
[6] P.J. Green, Reversible Jump MCMC Computation and Bayesian Model Determination, Biometrika, 1995, 711-732.
[7] Suparman and M. Doisy. Bayesian Segmentation of Piecewise Linear Regression Models Using Reversible Jump MCMC Algorithm, Computer Technology and Application, 2014, 14-18.
[8] Suparman, M. Doisy, J.Y. Tourneret, Changepoint Detection Using Reversible Jump MCMC Methods, Proceedings of IEEE ICASSP, 2002, pp. 159-1573.
[9] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer, New York, 1999.
[10] E. Punskaya, C. Andrieu, A. Doucet, and W.J. Fitzgerald. Bayesian Curva Fitting Using MCMC with Applications to Signal Segmentation, IEEE Transactions on Signal Processing, 2002, 747-758.

Suparman was born in Yogyakarta, Indonesia, on Avril 17, 1969. He received the B.Sc degree from the University of Lampung, Lampung, Indonesia, in 1992, the M.S. degree from the Gadjah Mada University, Yogyakarta, Indonesia, in 1997 and the M.S and Ph.D. degrees from the University of Toulouse III, Toulouse, France in 2003. Since February 2011, he has been a Lecturer with the Department of Mathematics Education, University of Ahmad Dahlan, Yogyakarta, Indonesia.

His research interests include reversible jump Markov chain Monte Carlo, Bayesian Statistics and their application to time series, regression and signal processing.