EPH - International Journal of Mathematics and Statistics
ISSN (Online): 2208-2212
Volume 7 Issue 1 February 2021

DOI:https://doi.org/10.53555/eijms.v7i1.54

BAYESIAN ESTIMATION OF LINEAR REGRESSION MODEL WITH MULTIPLE CHANGE POINTS FOR MISSING DATA

Chaobing He[1*]
*[*1]School of Mathematics and Statistics, Anyang Normal University, Anyang 455000, China*

Corresponding Author:-
Email: chaobing5@163.com

Abstract:-
The missing data is filled in by a random way. The complete data likelihood function of linear regression model with multiple change points for missing data is obtained. The full conditional distributions of change-point positions and other unknown parameters are studied. All the parameters are sampled by Gibbs sampler, and the means of Gibbs samples are taken as Bayesian estimations of the parameters. Random simulation results show that the estimations are fairly accurate.

Keywords:-*Complete-data likelihood function; full conditional distribution; prior distribution; Gibbs sampling; Metropolis-Hastings algorithm*

Mathematics Subject Classification: *62F15*

1. INTRODUCTION

The change point model is very important in mathematical statistics, and it is widely used in industrial quality control, economy, hydrologic statistics and other fields [1–4]. With the rapid development of statistical computing technology, Bayesian methods are increasingly applied, especially the Markov chain Monte Carlo (MCMC) methods [5–7]. Gibbs sampling of MCMC methods makes parameter estimation of change point model very convenient. Regression analysis is a statistical method to determine the quantitative relationship of the interdependence between variables. It is widely used in economy, finance, medicine and social science. Linear regression analysis is the most classical and most commonly used regression analysis method, and many nonlinear models can be transformed into linear regression models for analysis [8–10]. The linear regression change point model with complete data has been studied in literature[11–15],

but the lack of data is rarely studied. This paper mainly studies the parameter estimation of linear regression model with multiple change points for missing data by using MCMC methods. The missing data is filled in by a random way. All the parameters are sampled by Gibbs sampler, and the means of Gibbs samples are taken as Bayesian estimations of the parameters. Random simulation results show that the estimations are fairly accurate.

The rest of this paper is organized as follows. In Section 2, we present a description of linear regression model with missing data and its likelihood function. Section 3 describes multiple change-point problem under linear regression model with missing data and gives change-point estimation by Gibbs sampling and Metropolis-Hastings algorithm. In Section 5, random simulation test is developed and the test results show that Bayesian estimation of each parameter is fairly accurate. Finally, we summarize and conclude the paper in Section 6.

2. Linear regression model with missing data

The linear regression model is described as follows:

$$y_i = \alpha + \beta x_i + \varepsilon_i, i = 1,2,\cdots,n,$$

where x_i is a non-random variable, $\varepsilon_i \sim N(0,\sigma^2)$, y_is are mutually independent, and $\theta = (\alpha,\beta,\sigma^2)$ is the parameter vector. Let $\phi(x)$ denote the probability density function (pdf) of the standard normal distribution $N(0,1)$. Since $y_i \sim N(\alpha + \beta x_i,\sigma^2)$, hence the likelihood function for complete data is given by

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{n} \sigma^{-1}\varphi((y_i-\alpha-\beta x_i)\sigma^{-1}) \\ &= \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left\{-(y_i-\alpha-\beta x_i)^2/(2\sigma^2)\right\} \end{aligned}.$$

Suppose that some y_is are missing in the observation data (y_i,x_i), however, the corresponding x_is can be observed. In order to make full use of the data x_i, it is better to add the missing $y_i = y_{1i}$. Since $y_{1i} \sim N(\alpha + \beta x_i,\sigma^2)$, we can generate y_{1i} by random sampling.

Introduce indicative variable $\delta_i = I(y_i$ is not missing). Let x,y and u_1 be the vectors of $x_i s, y_i s$ and $y_{1i}s$, respectively. Set $D = \{1,2,\cdots,n\}$ and let $L(D)$ denote the number of elements in the set D.

Hence the likelihood function after adding data is given by

$$\begin{aligned} L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u}_1|\boldsymbol{\theta}) &= \sigma^{-n}\prod_{i=1}^{n}[\varphi((y_i-\alpha-\beta x_i)\sigma^{-1})]^{\delta_i}[\varphi((y_i-\alpha-\beta x_i)\sigma^{-1})]^{(1-\delta_i)} \\ &\propto (\sigma^2)^{-n/2}\exp\left\{-g(\alpha,\beta,D)/(2\sigma^2)\right\}, \end{aligned} \tag{1}$$

where

$$g(\alpha,\beta,D) = \alpha^2 L(D) + 2\alpha\beta\sum_{i\in D} x_i + \beta^2\sum_{i\in D} x_i^2 - 2\alpha\sum_{i\in D} b_i - 2\beta\sum_{i\in D} x_i b_i + \sum_{i\in D} c_i,$$

$$b_i = \delta_i y_i + (1-\delta_i)y_{1i}, c_i = \delta_i y_i^2 + (1-\delta_i)y_{1i}^2.$$

3. Linear regression model with multiple change points

Linear regression model with multiple change points is as follows:

$$y_i = \begin{cases} \alpha_1 + \beta_1 x_i + \varepsilon_i^{(1)},\ i = 1,2,\cdots,k_1, \\ \alpha_2 + \beta_2 x_i + \varepsilon_i^{(2)},\ i = k_1+1,\cdots,k_2, \\ \alpha_3 + \beta_3 x_i + \varepsilon_i^{(3)},\ i = k_2+1,\cdots,n, \end{cases} \tag{2}$$

where $\varepsilon_i^{(m)} \sim N(0,\sigma_m^2), m = 1,2,3,\ (\alpha_1,\beta_1,\sigma_1^2) \neq (\alpha_2,\beta_2,\sigma_2^2),(\alpha_2,\beta_2,\sigma_2^2) \neq (\alpha_3,\beta_3,\sigma_3^2)$, and k_1 and k_2 are called change-point position parameters.

4. Bayesian multiple change-point estimation

Set $\gamma = (k_1,k_2,\alpha_1,\alpha_2,\alpha_3,\beta_1,\beta_2,\beta_3,\sigma_1^2,\sigma_2^2,\sigma_3^2), D_1 = \{1,2,\cdots,k_1\}, D_2 = \{k_1+1,\cdots,k_2\}, D_3 = \{k_2+1,\cdots,n\}$, and $l_i = L(D_i)$, hence $l_1 = k_1, l_2 = k_2 - k_1$, and $l_3 = n - k_2$.

From Equations (1) and (2), under linear regression model with multiple change points, the likelihood function is given by

$$L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u}_1|\gamma)=\prod_{i=1}^{3}(\sigma_i^2)^{-l_i/2}\exp\left\{-g(\alpha_i,\beta_i,D_i)/(2\sigma_i^2)\right\}.$$

When $\delta_i = 0$,

$$\pi(y_{1i}|\boldsymbol{\gamma},\boldsymbol{x},\boldsymbol{y},\boldsymbol{u}_{-1i})\propto\psi(y_{1i};\boldsymbol{\gamma},x_i)\sim N(\alpha_m+\beta_m x_i,\sigma_m^2)),\ i\in D_m, m=1,2,3,$$

where $u_{-1i} = \{y_{1j} : j \neq i\}$.

The prior distributions of all parameters are as follows.
Let us consider a non informative joint prior distribution for k_1 and k_2

$$\pi(k_1,k_2)=\frac{1}{C_{n-2}^2}=\frac{2}{(n-2)(n-3)},\ \ 2\le k_1<k_2\le n-1.$$

Let us consider a normal prior distribution $N(\mu_i,\tau_i^2)$ for α_i

$$\pi(\alpha_i)\propto\exp\left\{-(\alpha_i-\mu_i)^2/2\tau_i^2\right\}.$$

Let us consider a normal prior distribution $N(\rho_i,\omega_i^2)$ for β_i

$$\pi(\beta_i)\propto\exp\left\{-(\beta_i-\rho_i)^2/2\omega_i^2\right\}.$$

Let us consider a inverse gamma prior distribution $IGa(\gamma_i,\lambda_i)$ for σ_i^2

$$\pi(\sigma_i^2)\propto(\sigma_i^2)^{-(\gamma_i+1)}\exp\left\{-\lambda_i/\sigma_i^2\right\},\gamma_i>0,\lambda_i>0.$$

Suppose that $(k_1,k_2),\alpha_1,\alpha_2,\alpha_3,\beta_1,\beta_2,\beta_3,\sigma_1^2,\sigma_2^2$ and σ_3^2 are independent, hence

$$\begin{aligned}\pi(\gamma|\boldsymbol{x},\boldsymbol{y},\boldsymbol{u}_1)&\propto L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u}_1|\gamma)\pi(k_1,k_2)\prod_{i=1}^{3}\pi(\alpha_i)\pi(\beta_i)\pi(\sigma_i^2)\\&\propto\prod_{i=1}^{3}(\sigma_i^2)^{-(l_i/2+\gamma_i+1)}\exp\left\{-\frac{g(\alpha_i,\beta_i,D_i)}{2\sigma_i^2}-\frac{\lambda_i}{\sigma_i^2}\right.\\&\qquad\left.-\frac{(\alpha_i-\mu_i)^2}{2\tau_i^2}-\frac{(\beta_i-\rho_i)^2}{2\omega_i^2}\right\}.\end{aligned}$$

Let $\pi(\alpha_1|\cdot)$ denote $\pi(\alpha_1|k_1,k_2,\alpha_2,\alpha_3,\beta_1,\beta_2,\beta_3,\sigma_1^2,\sigma_2^2,\sigma_3^2,\boldsymbol{x},\boldsymbol{y},\boldsymbol{u}_1)$.

The full conditional distributions of all parameters are as follows.

$$\begin{aligned}\pi(\alpha_i|\cdot)&\propto\exp\left\{-\frac{1}{2\sigma_i^2\tau_i^2}\{(\sigma_i^2+\tau_i^2)\alpha_i^2-2[\mu_i\sigma_i^2-(\beta_i\sum_{j\in D_i}x_j-\sum_{j\in D_i}y_j)\tau_i^2]\alpha_i\}\right\}\\&\propto\exp\left\{-\frac{1}{2B_i^2}(\alpha_i-A_i)^2\right\}\propto N(A_i,B_i^2),\end{aligned}$$

where

$$A_i=\frac{\mu_i\sigma_i^2-(\beta_i\sum\limits_{j\in D_i}x_j-\sum\limits_{j\in D_i}y_j)\tau_i^2}{\sigma_i^2+a_i\tau_i^2},\qquad B_i^2=\frac{\sigma_i^2\tau_i^2}{\sigma_i^2+\tau_i^2}.$$

$$\begin{aligned}\pi(\beta_i|\cdot)&\propto\exp\left\{-\frac{1}{2\sigma_i^2\tau_i^2}\{[\sigma_i^2+(\sum_{j\in D_i}x_j^2)\omega_i^2]\beta_i^2-2[\rho_i\sigma_i^2-(\alpha_i\sum_{j\in D_i}x_j-\sum_{j\in D_i}x_jy_j)\omega_i^2]\beta_i\}\right\}\\&\propto\exp\left\{-\frac{1}{2E_i^2}(\beta_i-C_i)^2\right\}\propto N(C_i,E_i^2),\end{aligned}$$

where

$$C_i=\frac{\rho_i\sigma_i^2-(\alpha_i\sum\limits_{j\in D_i}x_j-\sum\limits_{j\in D_i}x_jy_j)\omega_i^2}{\sigma_i^2+(\sum\limits_{j\in D_i}x_j^2)\omega_i^2},\qquad D_i^2=\frac{\sigma_i^2\tau_i^2}{\sigma_i^2+(\sum\limits_{j\in D_i}x_j^2)\omega_i^2}.$$

$$\pi(\sigma_i^2|\cdot)\propto(\sigma_i^2)^{-(l_i/2+\gamma_i+1)}e^{-[g(\alpha_i,\beta_i)/2+\lambda_i]/\sigma_i^2}\propto IGa(l_i/2+\gamma_i,g(\alpha_i,\beta_i)/2+\lambda_i).$$

The full conditional distributions of k_1 and k_2 are both not common standard distributions:

$$\pi(k_1|\cdot) \propto (\sigma_2^2/\sigma_1^2)^{k_1/2} \exp\left\{-(s_1/\sigma_1^2 + s_2/\sigma_2^2)/2\right\},\ 1 \le k_1 \le k_2 - 1,$$

$$\pi(k_2|\cdot) \propto (\sigma_3^2/\sigma_2^2)^{k_2/2} \exp\left\{-(s_2/\sigma_2^2 + s_3/\sigma_3^2)/2\right\},\ k_1 + 1 \le k_2 \le n - 1,$$

where $s_i = g(\alpha_i, \beta_i, D_i), i = 1,2,3$.

Since full conditional distributions are obtained, we can get the stationary distributions of every parameter using MCMC methods. α_is, β_is and σ_i^2s are all generated directly by Gibbs sampling. However, it is difficult to conduct Gibbs sampling directly for k_1 and k_2, so we use Metropolis-Hastings algorithm to sample k_1 and k_2, and the two proposed distributions are selected as discrete uniform distributions.

The MCMC methods can be described by the following iterative steps; where $\gamma^{(t)}$ is the vector of generated values in t iteration of the algorithm:

(1) Set $\gamma^{(0)} = (k_1^{(0)}, k_2^{(0)}, \alpha_1^{(0)}, \alpha_2^{(0)}, \alpha_3^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}, \beta_3^{(0)}, \sigma_1^{2(0)}, \sigma_2^{2(0)}, \sigma_2^{2(0)})$.

(2) For $t = 1,2,\cdots,M$, repeat the following steps:

1. Set $\gamma = \gamma(t-1)$.

2. When $\delta_i = 0$, generate $y_{1i}^{(t)}$ from $\psi(y_{1i};\gamma,x_i)$ and set $y_{1i} = y_{1i}^{(t)}$.

3. Generate $\alpha_1^{(t)}$ from $\pi(\alpha_1|\cdot)$ and set $\alpha_1 = \alpha_1^{(t)}$
$\alpha_2^{(t)}, \alpha_3^{(t)}, \beta_i^{(t)} 2(t)$ $\alpha_1^{(t)}$

4. The generations ofs and σ_i s are similar to.

5. Sample k_1' from 2, $3,\cdots,k_2^{(t-1)} - 1$ and generate u from uniform $U(0,1)$.
Set $r_1(k_1^{(t-1)}, k_1') = \min\left\{\dfrac{\pi(k_1'|\cdot)}{\pi(k_1^{(t-1)}|\cdot)}, 1\right\}$, set $k_1^{(t)} = k_1'$ if $u \le r_1(k_1^{(t-1)}, k_1')$,
set $k_1^{(t)} = k_1^{(t-1)}$ otherwise, and set $k_1 = k_1^{(t)}$.

6. Sample k_2' from $k_1^{(t)} + 1, \cdots, n - 1$ and generate u from uniform $U(0,1)$.
Set $r_2(k_2^{(t-1)}, k_2') = \min\left\{\dfrac{\pi(k_2'|\cdot)}{\pi(k_2^{(t-1)}|\cdot)}, 1\right\}$, set $k_2^{(t)} = k_2'$ if $u \le r_2(k_2^{(t-1)}, k_2')$,
set $k_2^{(t)} = k_2^{(t-1)}$ otherwise, and set $k_2 = k_2^{(t)}$.

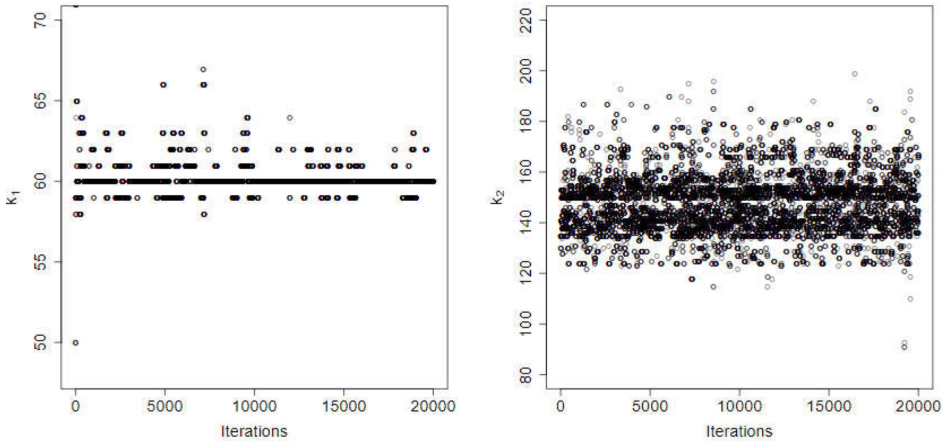
Assume that $\gamma^{(t)} = (k_1^{(t)}, k_2^{(t)}, \alpha_1^{(t)}, \alpha_2^{(t)}, \alpha_3^{(t)}, \beta_1^{(t)}, \beta_2^{(t)}, \beta_3^{(t)}, \sigma_1^{2(t)}, \sigma_2^{2(t)}, \sigma_3^{2(t)})$
$(t = 1,2,\cdots,B,\cdots,M)$ is a Gibbs sample, and in the burnin period the first B iterations are eliminated from the sample in order to avoid the influence of the initial values. Consider $\{\gamma^{(B+1)},\gamma^{(B+2)},\cdots,\gamma^{(M)}\}$ as the sample for the posterior analysis. From this sample, in the following simulation test, we will estimate the posterior mean, median, 2.5% and 97.5% quantiles of γ.

5. Simulation results

The simulation test is developed using the R software. The sample size used in this study is $n = 200$. The vector $\gamma = (k_1, k_2, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \sigma_1^2, \sigma_2^2, \sigma_3^2)$ is taken as $(60,150,-2,4,1.5,3,-0.8,2.5,2.8,1,3.6)$. First take the constants $x_1,x_2,\cdots,x_n$, then select the majority of x_is and generate no missing data y_i from $N(\alpha_m + \beta_m x_i, \sigma_m^2), i \in D_m, m = 1,2,3$. The y_is corresponding to the remaining small part of x_is are taken as the missing data. We consider normal prior distributions $N(-2.3,0.5)$ for α_1, $N(3.8,1.5)$ for α_2, and $N(1.3,0.7)$ for α_3. Consider prior distributions $N(2.7,0.6)$ for β_1, $N(-0.75,0.2)$ for β_2, and $N(2.8,1.2)$ for β_3. Consider prior distributions $IGa(2,2.6)$ for σ_1^2, $IGa(1.5,0.7)$ for σ_2^2, and $IGa(3,7.5)$ for σ_3^2.
Using MCMC methods, our analysis focuses on k_1 and k_2. In the process of simulation, total number of iterations is M = 20000, and the first 10000 iterations are eliminated in the burnin period. The simulation results are presented in Table 1. The iterations of k_1 and k_2 are presented in Figure 1.
To monitor the convergence of the algorithm, we run multiple chains with different starting points. When we observe that the lines of different chains mix or cross in trace, then convergence is ensured. In the simulation study, we run two chains with different starting points. The two iterative chains of k_1 and k_2 are presented in Figure 2.
The analysis of simulation results is as follows. First, it can be observed from Table 1 that the relative errors of k_1 and k_2 are both no more than 2%, the relative errors of the other parameters are no more than 6%, and MC errors is small. The difference in Gibbs sample median and mean is very small, so sample median is also taken Bayesian estimation of the parameter. A 95% credible interval is approximately obtained as [2.5% quantile, 97.5% quantile],

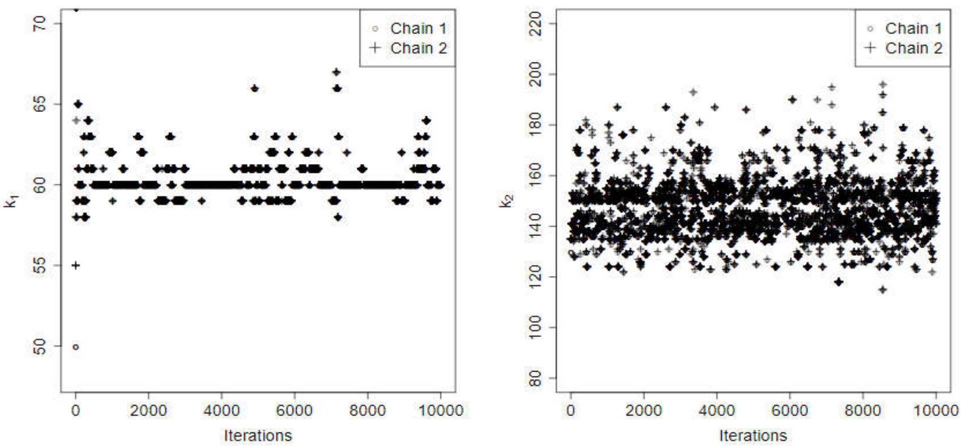
Table 1 The results of Bayesian estimations of parameters

θ	True value	Mean	Relative error	MC error	2.5%	Median	97.5%
$k1$	60	60.20040	0.00334	0.00940	59	60	63
$k2$	150	147.42580	0.01716	0.09788	128	150	169
$\alpha 1$	-2	-2.03618	0.01809	0.00118	-2.22702	-2.03826	-1.84056
$\alpha 2$	4	4.03084	0.00771	0.00234	3.65179	4.02820	4.41804
$\alpha 3$	1.5	1.55510	0.03674	0.00090	1.40752	1.55532	1.70255
$\beta 1$	3	3.06701	0.02234	0.00178	2.77640	3.06665	3.36025
$\beta 2$	-0.8	-0.83712	0.04640	0.00048	-0.91665	-0.83713	-0.75803
$\beta 3$	2.5	2.56922	0.02769	0.00148	2.32633	2.56783	2.81476
σ_1^2	2.8	2.91736	0.04192	0.00168	2.63921	2.91815	3.19179
σ_2^2	1	1.05222	0.05222	0.00061	0.95243	1.05309	1.15209
	σ_3^2 3.6	3.72434	0.03454	0.00215	3.37304	3.72209	4.08226



(a) (b)

Figure:1 Gibbs iterations of k_1 and k_2



(a) (b)

Figure 2 Two iterative chains of k_1 and k_2

whose length is very short, so the effect of interval estimation is good. In general, Bayesian estimations are fairly accurate. From Figure 1, it can be observed that the Gibbs sampling has small fluctuation, and tends to be stable and presents a strong regularity. Second, monitor the convergence of the algorithm. It can be observed from Figure 2 that the trace plots of the two chains mix, which ensures the convergence. Above all, the effect of simulation test is good using MCMC methods.

6. Conclusions

In this paper, we consider Bayesian multiple change-point estimation under linear regression model with missing data. The full conditional distributions of all parameters are dicussed. Bayesian estimations of parameters are studied by Gibbs sampling and Metropolis-Hastings algorithm of MCMC methods. Our simulation results show that Bayesian estimations of the parameters are fairly accurate and the effect of simulation is good using MCMC methods.

Reference

[1]. E. S. Page, Continuous inspection schemes, Biometrika 41 (1/2) (1954) 100–115.
[2]. H. Chernoff, S. Zacks, et al., Estimating the current mean of a normal distribution which is subjected to changes in time, The Annals of Mathematical Statistics 35 (3) (1964) 999–1018.
[3]. P. Fearnhead, Exact and efficient bayesian inference for multiple changepoint problems, Statistics and computing 16 (2) (2006) 203–213.
[4]. T. Yuan, Y. Kuo, Bayesian analysis of hazard rate, change point, and cost-optimal burn-in time for electronic devices, IEEE Transactions on Reliability 59 (1) (2010) 132–138.
[5]. F. Liang, W. H. Wong, Real-parameter evolutionary monte carlo with applications to bayesian mixture models, Journal of the American Statistical Association 96 (454) (2001) 653–666.
[6]. M. Lavielle, E. Lebarbier, An application of mcmc methods for the multiple change-points problem, Signal processing 81 (1) (2001) 39–53.
[7]. J. Kim, S. Cheon, Bayesian multiple change-point estimation with annealing stochastic approximation monte carlo, Computational Statistics 25 (2) (2010) 215–239.
[8]. G. A. Seber, A. J. Lee, Linear regression analysis, John Wiley & Sons, 2012.
[9]. B. Jann, The blinder–oaxaca decomposition for linear regression models, The Stata Journal 8 (4) (2008) 453–479.
[10]. C.-H. Zhang, J. Huang, et al., The sparsity and bias of the lasso selection in high-dimensional linear regression, The Annals of Statistics 36 (4) (2008) 1567–1594.
[11]. B. Zhang, J. Geng, L. Lai, Multiple change-points estimation in linear regression models via sparse group lasso, IEEE Transactions on Signal Processing 63 (9) (2015) 2209–2224.
[12]. M. D¨oring, U. Jensen, Smooth change point estimation in regression models with random design, Annals of the Institute of Statistical Mathematics 67 (3) (2015) 595–619.
[13]. Y. Li, L. Qian, W. Zhang, Estimation in a change-point hazard regression model with long-term survivors, Statistics & Probability Letters 83 (7) (2013) 1683–1691.
[14]. J.-G. Lin, J. Chen, Y. Li, Bayesian analysis of student t linear regression with unknown change-point and application to stock data analysis, Computational Economics 40 (3) (2012) 203–217.
[15]. C. W. Chen, J. S. Chan, R. Gerlach, W. Y. Hsieh, A comparison of estimators for regression models with change points, Statistics and Computing 21 (3) (2011) 395–414.