Hierarchical Bayesian segmentation for piecewise stationary autoregressive model based on reversible jump MCMC

By SUPARMAN

Hierarchical Bayesian segmentation for piecewise stationary autoregressive model based on reversible jump MCMC

Suparman[1*]

[1]Department of Mathematics Education, University of Ahmad Dahlan, Indonesia
suparman@pmat.uad.ac.id[*]

Abstract. This paper aims to decompose time series data in segments where many segments are unknown. The data in each segment is modeled as a stationary autoregressive where the model order is unknown. The model parameters include the number of segments, the location of segment changes, the order of each segment, and the autoregressive coefficients of each segment. The Bayesian method is used to estimate parameters, but Bayesian estimator cannot be calculated analytically. The Bayesian estimator is calculated using the reversible jump Markov Chain Monte Carlo algorithm. The performance of the algorithm is tested using synthesis data. The simulation results show that the algorithm estimates the model parameters well.

1. Introduction

A constant model per segment is a model that is often used to model various types of data. Kaipio & Karjalainen [1] uses a constant autoregressive model per segment to model the electroencephalogram (EEG) data. Evans et al. [2] uses a constant model per segment to model DNA data. Fryzlewicz & Rao [3] uses a constant autoregressive model per segment to model the Financial Times Stock Exchange (FTSE) data. Shao et al. [4] uses a constant linear regression model per segment to model precipitation data. Yau & Zhao [5] use stationary models per segment to model electroencephalogram data.

The method for constant model per segment segmentation is examined by several authors. Punskaya et al. [6] uses the reversible jump MCMC method to segment sound signals. Nitanda et al. [7] uses fuzzy c-means clustering to segment audio signals. Tai et al. [8] used Bayesian methods to segment tumor data. Phojanamongkolkij [9] uses a spectogram to segment sismic data. Zimroz et al. [10] uses time frequency decomposition to segment seismic data. Kavsaoglu et al. [11] uses an adaptive segmentation method to segment photopletthysmography data. Polak et al. [12] uses the analysis of time frequency maps of group delay to segment and cluster seismic data. Hewaarachchi et al. [13] uses the Bayesian minimum description length (BMDL) method to segment temperature data. Kim et al. [14] used a domain assisted parameter semi-free wave mining (DAPs) model to segment data epileptic activity data.

The AR model that has stationary properties is a useful model in forecasting. Punskaya et al. [6] does not discuss stationary AR models. If the piecewise stationary constant AR model is matched to real data, generally the model parameters are unknown. The parameters here include: number of segments, location of the model changes, and AR model parameters for each segment. The AR model parameters include: order, coefficient, and variance in stochastic disturbances. This paper aims to estimate the stationary-constant-per-segment AR model using the reversible jump MCMC algorithm.

2. Research Method

Suppose $x = (x_1, \dots, x_n)$ is n observation. This data is said to have a constant AR model per segment k ($k = 0,1,\dots,k_{max}$) if for t = 1, ..., n this data has the following stochastic equation:

$$x_t = z_t - \sum_{j=1}^{p} \phi_{i,j,k}^{(p_{i,k})} x_{t-j}, \quad \tau_{i,k} < t \le \tau_{i+1,k}, \quad i = 0,1,\dots,k \tag{1}$$

where under the assumption of k segment: $\tau_{i,k}$ is the location of the ith AR model change, with conversions $\tau_{0,k} = 0$ and $\tau_{k+1,k} = n$ where each ith segment, $p_{i,k}$ and $\phi_{i,k}^{(p_{i,k})} = \left(\phi_{i,k,1}^{(p_{i,k})}, \dots \phi_{i,k,q_{i,k}}^{(p_{i,k})}\right)$ is the AR model coefficient corresponding to the ith segment. z_t is a stochastic error value at t corresponding to the ith segment. The z_t is modeled as a normal distribution with mean 0 and variance $\sigma_{i,k}^2$. Next the ith AR model (i = 0.1, ..., k) is called stationary if and only if the equation

$$\phi(a) = 1 - \sum_{j=1}^{p_{i,k}} \phi_{i,k,j}^{p_{i,k}} a \tag{2}$$

is 0 for the value of a outside the circle with the radius equal to one.

If the number of segments is assumed to be known, the location of the AR model change is assumed to be known and the assumed order is known, then a problem of the piecewise constant stationary AR model estimation becomes a problem of order identification and AR model parameter estimation for each segment. If the AR model order is assumed to be known, the problem of identifying the AR model order and AR parameter estimation becomes a problem of AR model parameter estimation. In this study, the number of segments and the order of the AR model for each segment is assumed to be unknown. The reversible jump algorithm MCMC is used to detect the number of segments, detect the location of the AR model changes, identify the AR model order and estimate the AR model parameters simultaneously. The Hierarchical Bayesian is adopted to estimate the hyperparameter that appears. The performance of the reversible jump MCMC algorithm will be tested by using synthesis data.

3. Results and Discussion

Suppose $s = \left(x_{p_{max+1}}, \dots, x_n\right)$ is a realization of a piecewise stationary constant AR model. If the values of $s_0 = \left(x_1, \dots, x_{p_{max}}\right)$ are known and

$$\theta = \left(k, \tau^{(k)}, p_{i,k}, \left\{\phi_{i,k}^{(p_{i,k})}\right\}_{i=0}^{k}, \sigma^{(k)}\right),$$

then the likelihood function of s can be written more or less as follows:

$$L(s|\phi, s_0) = \prod_{i=0}^{k} \left(2\pi\sigma_{i,k}^2\right)^{-\frac{1}{2}} \left(\tau_{i+1,k} - \tau_{i,k}\right) exp - \frac{1}{2\sigma_{i,k}^2} \sum_{t=\tau_{i,k}+1}^{\tau_{i+1,k}} \left(x_t - \sum_{j=1}^{p_{i,k}} G\left(\rho_{i,k}^{(p_{i,k})}\right) x_{t-j}\right)^2 \tag{3}$$

for $t = p_{max+1}, \dots, n$. Suppose $S_{p_{i,k}}$ is the area of stationarity. By using transformation

$$F: \phi_{i,k}^{(p_{i,k})} \in S_{p_{i,k}} \mapsto \rho_{i,k}^{(p_{i,k})} \in (-1,1)^{p_{i,k}} \tag{4}$$

The AR model is stationary if and only if $\rho_{i,k}^{(p_{i,k})} \in (-1,1)^{p_{i,k}}$ [21]. If

$$\rho = \left(k, \tau^{(k)}, p_{i,k}, \left\{\rho_{i,k}^{(p_{i,k})}\right\}_{i=0}^{k}, \sigma^{(k)}\right)$$

then the likelihhod fuction can be rewritten as:

$$L(s|\rho, s_0) = \prod_{i=0}^{k} \left(2\pi\sigma_{i,k}^2\right)^{-\frac{1}{2}} \left(\tau_{i+1,k} - \tau_{i,k}\right) exp - \frac{1}{2\sigma_{i,k}^2} \sum_{t=\tau_{i,k}+1}^{\tau_{i+1,k}} \left(x_t - \sum_{j=1}^{q_{i,k}} G^{-1}\left(\phi_{i,k}^{(p_{i,k})}\right) x_{t-j}\right)^2 \tag{5}$$

3.1. Hierarchical Bayesian Approach

The hierarchical Bayesian approach is used by several authors, for example He et al. [15], Millar [16], Cross et al. [17], Shelton et al. [18], Grzegororezyk & Husmeier [19], Glassen & Nitsch [20]. The selection of prior distributions for the above parameters is as follows:

1. Suppose that $\pi(k|\lambda)$ is a prior distribution for the number of segments k. A binomial distribution with parameter λ is selected as the prior distribution $\pi(k|\lambda)$, namely: $\pi(k|\lambda) = C_k^{k_{max}}\lambda^k(1-\lambda)^{k_{max}-k}$ for $k = 1, \dots, k_{max}$ and is 0 for others.
2. Suppose that $\pi(\tau_k|k)$ is the prior distribution for the position τ_k. The even distribution of indexes from 2k + 1 sequential statistics is taken uniformly without returns in {1, ..., n-1} is selected as the prior distribution $\pi(\tau_k|k)$.
3. Suppose that $\pi(p_{i,k}|k)$ is prior distribution for order $p_{i,k}$. A uniform distribution in $\{1, \dots, p_{max}\}$ is selected as the prior distribution $\pi(p_{i,k}|k)$.
4. Suppose that $\pi\left(\rho_{i,k}^{(p_{i,k})}\middle|k, p_{i,k}\right)$ is a conditional prior distribution for the coefficient vector $\rho_{i,k}^{(p_{i,k})}$ if it is known $p_{i,k}$. A uniform distribution in the interval $(-1,1)^{q_{i,k}}$ is selected as the prior distribution $\pi\left(\rho_{i,k}^{(p_{i,k})}\middle|k, p_{i,k}\right)$.
5. Suppose that $\pi(\sigma_{i,k}^2|k,\alpha,\beta)$ is the prior distribution for $\sigma_{i,k}^2$. An inversion gamma distribution with the parameters $\frac{\alpha}{2}$ and $\frac{\beta}{2}$ is selected as the prior distribution $\pi(\sigma_{i,k}^2|k,\alpha,\beta)$, namely: $\pi(\sigma_{i,k}^2|k,\alpha,\beta) = \frac{(\beta/2)^{\alpha/2}}{\Gamma(\alpha/2)}(\sigma_{i,k}^2)^{-(1+\alpha/2)} exp - \beta/(2\sigma_{i,k}^2)$ for $\sigma_{i,k}^2 > 0$ and is 0 for the others.

The parameter that appears in the prior distribution is called hyperparameter. Suppose that $\pi(\lambda)$ is the prior distribution for the hyperparameter λ. The distribution $\pi(\lambda)$ is assumed to be uniformly distributed at intervals (0.1). Suppose that $\pi(\beta)$ is the prior distribution for hyperparameter β. The distribution $\pi(\beta)$ is assumed to be Jeffrey's distribution. Here, the value of α is equal to 2. Let $H_1 = \left(k, \tau_k, p_{i,k}, \rho_{i,k}^{(p_{i,k})}, \sigma_{i,k}^2\right)$, and $H_2 = (\lambda, \beta)$, and $\pi(H_1, H_2)$ be the prior distribution for parameter (H_1, H_2). The prior distribution for parameter (H_1, H_2) can be expressed as

$$\pi(H_1, H_2) = \pi(k|\lambda)\pi(\lambda)\pi(\tau_k|k)\,\pi(p_{i,k}|k)\,\pi\left(\rho_{i,k}^{(p_{i,k})}\middle|k, p_{i,k}\right)\pi(\sigma_{i,k}^2|k,\alpha,\beta)\pi(\beta) \quad (6)$$

According to Bayes's theorem, the posterior distribution for the parameters H_1 and H_2 can be expressed as

$$\pi(H_1, H_2|s) \propto L(s|\rho, s_0)\pi(H_1, H_2).$$

The posterior distribution is a combination of likelihood function and prior distributions. In this case, the posterior distribution $\pi(H_1, H_2|s)$ has a complicated form so that the Bayes estimator cannot be determined analytically. The MCMC reversible jump algorithm was adopted to determine Bayes estimator.

3.2. Reversible Jump MCMC Algorithm

The reversible jump algorithm MCMC [22] was proposed to determine the Bayes estimator. Suppose that $M = (H_1, H_2)$. The general idea of the MCMC reversible jump algorithm is to make a homogeneous Markov chain $M_1, \dots, M_m$ that satisfies aperiodic and irreductible properties so that the Markov chain $M_1, \dots, M_m$ can be considered as random variables that follow the distribution $\pi(H_1, H_2|s)$. The Markov chain $M_1, \dots, M_m$ can be used to calcalute the estimation of parameter M. The Markov chain is made in two stages: stage 1 simulates the distribution $\pi(H_2|H_1, s)$ and stage 2 simulates the distribution $\pi(H_1|H_2, s)$. Distribution $\pi(H_2|H_1, s)$ has an explicit form. So the Gibbs algorithm can be used to simulate the distribution $\pi(H_1|H_2, s)$. The conditional distribution of H_1 if known (H_2, s) can be written as

$$\pi(H_2|H_1, s) = B(k+1, k_{max} - k + 1) \otimes G\left(\frac{\propto}{2}(k+1), \sum_{i=1}^{k} \frac{1}{2\sigma_{i,k}^2}\right) \quad (7)$$

Conversely, the distribution $(H_1|H_2,s)$ does not have an explicit form. So that the exact simulation is not possible. The simulation of distribution $\pi(H_1|H_2,s)$ is done in three stages, namely: the stage one simulates distribution $\pi\left(k,\tau^{(k)},p_{i,k},\left\{\rho_{i,k}^{(p_{i,k})}\right\}_{i=0}^{k}\middle|H_2,s\right)$, the stage two simulates distribution $\pi\left(p_{i,k},\left\{\rho_{i,k}^{(p_{i,k})}\right\}_{i=0}^{k}\middle|k,\tau^{(k)},H_2,s\right)$, and the stage three simulates distribution $\pi\left(\sigma_k^2\middle|p_{i,k},\left\{\rho_{i,k}^{(p_{i,k})}\right\}_{i=0}^{k},k,\tau^{(k)},H_2,s\right)$. The reversible jump MCMC algorithm is used in stage one and stage two. The Gibbs algorithm is used in stage three.

3.3. Simulation Study

The Figure 1 shows a synthesis data. This synthetis data is made according to equation (1).

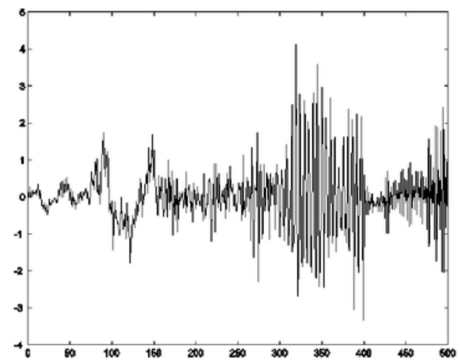

Fig. 1: Synthesis data

Making the synthesis data is done by taking the parameter value k = 4 and the location of the AR model change is $\tau = (75, 150, 250, 400)$. The Table 1 gives the order, coefficients and variances of the AR model for each segment.

Table 1: Parameter value for synthesis data

ith segment	$\sigma_{i,4}$	$p_{i,4}$	$\theta_{i,4}^{(p_{i,4})}$
0	0.12	3	(-0.25, -0.79, 0,34)
1	0,5	2	(-1.54, -0.41)
2	0,4	1	(0.19)
3	0,5	4	(0.59, 0.99, 0.64, 0.87)
4	0,12	3	(0.86, -0.83, -0.96)

Based on the data in Figure 1, then the model parameters are estimated using reversible jump MCMC. The reversible jump algorithm MCMC is used to estimate the number of segments, the location of the AR model changes, the AR model order for each segment, the AR model coefficient for each AR model, and the stochastic error variance. For this purpose, the MCMC reversible jump algorithm is implemented with 70.000 iterations with a 10.000 iteration burn-in period. The AR model order value is limited to a maximum of 10 so that p_{max}. The histogram for k is presented in Figure 2.

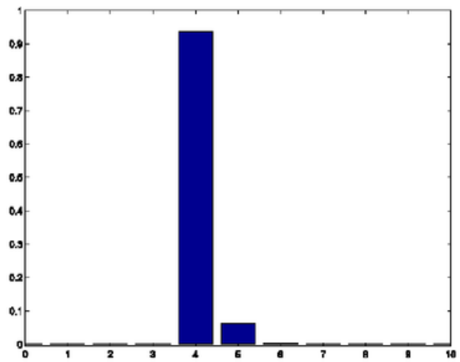

Fig. 2: Histogram for k

The histogram in Figure 2 shows that the maximum value for the number of segments occurs at $\hat{k} = 4$. So that the estimator for k is $\hat{k} = 4$. The histogram for τ corresponding to k = 4 is given in Figure 3. The result is $\hat{\tau} = (75, 150, 250, 400)$.

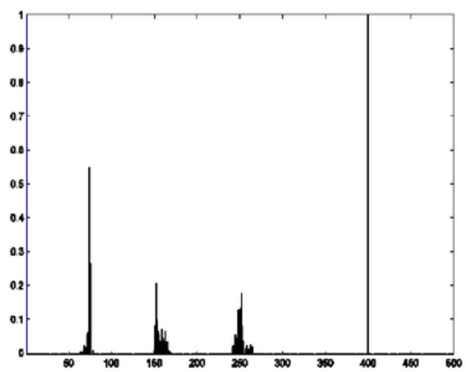

Fig. 3: Histogram for τ if known $\hat{k} = 4$

The segmentation results are presented in Figure 4.

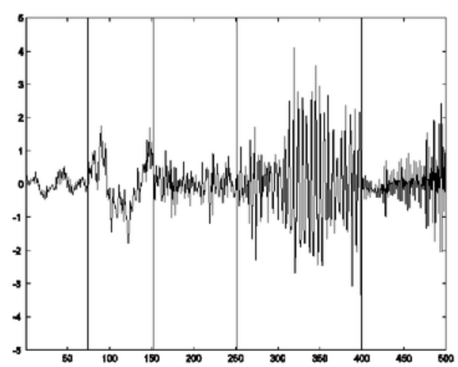

Fig. 4: Segmentation of data

Estimation results of coefficient and deviation of stochastic error standard for each segment are written in Table 2.

Table 2: Estimation results of coefficients and deviations of stochasticerror standards for synthetic data

ith segment	$\hat{\sigma}_{i,4}$	$\hat{p}_{i,4}$	$\hat{\theta}_{i,4}^{\hat{p}_{i,4}}$
0	0.13	3	(-0.23, -0.76, 0.23)
1	0.47	2	(-0.50, -0.27)
2	0.41	1	(0.34)
3	0.52	4	(0.57, 0.93, 0.62, 0.83)
4	0.13	3	(0.86, -0.79, -0.94)

Based on the output of the reversible jump MCMC algorithm, the data in Figure 1 is divided into 5 segments. In the first segment data is modeled by AR (3), in the second segment data is modeled by AR (2), in the third segment data is modeled by AR (1), in the fourth segment data is modeled by AR (4), and in the fifth segment data is modeled by AR (3).

3.4. Application

The Figure 5 shows a real data. This data is the Dow-Jones utilities index [23].

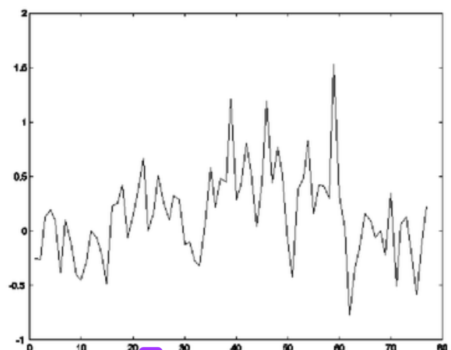

Fig. 5: Dow-Jones Utilities Index

Based on the data in Figure 5, the model parameters are estimated using the reversible jump MCMC algorithm. For this purpose, the MCMC reversible jump algorithm is implemented with 70.000 iterations with a 10.000 iteration burn-in period. The histogram for k is presented in Figure 6.

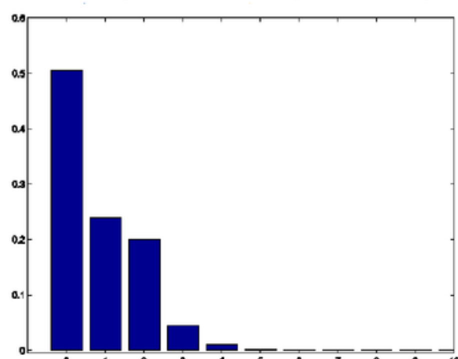

Fig. 6: Histogram for k

The histogram in Figure 6 shows that the maximum value for the number of segments occurs at $\hat{k} = 0$. So that the estimator for k is $\hat{k} = 0$. Estimation results for coefficients and standard deviation of stochastic errors are presented in Table 3.

Table 3: Estimation results from coefficients and deviations of stochastic error standards for the Dow-Jones utilities index

ith segment	$\hat{\sigma}_{i,0}$	$\hat{p}_{i,0}$	$\hat{\theta}_{i,0}^{\hat{p}_{i,0}}$
0	0.39	1	-0.46

4. Conclusion

The above description is a theoretical study of the reversible jump MCMC algorithm and its application to estimate the piecewise stationary constant AR model. By comparing the parameter values and their estimated values from the synthesis data, it shows that the reversible jump algorithm can estimate the piecewise stationary constant AR model parameters well. The advantage of reversible jump MCMC algorithm is that this algorithm can estimate stationary AR model parameters simultaneously. Another advantage is that this algorithm produces a stationary AR model for each segment. The reversible jump MCMC algorithm applied to the Dow-Jones utilities index data. This Dow-Jones utilities index data is modeled by AR(1).

References

[1] Kaipio J P & Karjalainen P A 1997, Simulation of nonstationary EEG. *Biol. Cybern* 76, 349-356.

[2] Evans G E, Sofronov G Y, Keith J M & Kroese D P 2011 Estimating change-points in biological sequences via the cross-entropy method. *Ann Oper Res* 189, 155-165.

[3] Fryzlewicz P & Rao S S 2014 Multiple-change-point detection for auto-regressive conditional heteroscedastic processes. *J.R. Statist. Soc. B* 76, 903-924.

[4] Shao Q, Li Z & Xu Z 2010 Trend detection in hydrological time series by segment regression with application to Shiyang River Basin. *Stoch Environ Res Risk Assess* 24, 221-233.

[5] Yau C Y & Zhao Z 2016 Inference for multiple change point in time series via likelihood ratio scan statistics. *J.R. Statist. Soc. B* 78, 895-916.

[6] Punskaya E, Andrieu C, Doucet A & Fitzgerald W J 2002, Bayesian curve fitting using MCMC with applications to signal segmentation. *IEEE Trans. on Signal Processing* 50, 747-758.

[7] Nitanda N, Haseyama M & Kitajima H 2006, Audio signal segmentation and classification using fuzzy c-means clustering. *Systems and Computers in Japan* 37, 302-312.

[8] Tai Y C, Kvale M N & Witte J S 2010 Segmentation and estimation for SNP microarrays : a Bayesian multiple change-point approach. *Biometrics* 66, 675-683.

[9] Phojanamongkolkij N 2014 A comparison of climate signal tren detection uncertainty analysis methods. *Journal of climate* 27, 3363-3376.

[10] Zimroz R, Madziarz M, Zak G, Wylomanska A & Obuchowski J 2015 Seismic signal segmentation procedure using time-frequency decomposition and statistical modelling. *Journal of Vibroengineering* 17, 3111-3121.

[11] Kavsaoglu A R, Polat K & Bozkurt M R 2016 An innovative peak detection algorithm for photoplethysmography signals : an adaptive segmentation method. *Turkish Journal of Electrical Engineering & Computer Sciences* 24, 1782-1796.

[12] Polak M, Obuchowski J, Madziarz M, Wylomanska A & Zimroz R 2016, Time-varying group delay as a basis for clustering and segmentation of seismic signals. *Journal of Vibroengineering* 18, 1392-8716.

[13] Hewaarachchi A P, Li Y, Lund R & Rennie J 2017 Homogenization of daily temperature data.

American Meteorological Society 30, 985-999.
[14] Kim S-H, Li L, Faloutsos C, Yang H-J & Lee S-W 2017 DAPs : mining using change-point detection of epileptic activity time series data. *Journal of Information Science and Engineering* 33, 517-536.
[15] He H S, Dey D C, Fan X, Hooten M B, Kabrick J M, Wikle C K & Fan Z 2007 Mapping pre-European settlement vegetation at fine resolutions using a hierarchical Bayesian model and GIS. *Plant Ecol* 191, 85-94.
[16] Millar R B 2009 Comparison of Hierarchical Models for Overdispersed Count Data Using DIC and Bayes'Factors. *Biometrics* 65, 962-969.
[17] Cross P C, Heisey D M, Scurlock B M, Edwards W H, Ebinger M R & Brennan A 2010 *PloS ONE* 5, 1-9.
[18] Shelton A O, Dick E J, Pearson D E, Ralston S & Mangel M 2012 Estimating species composition and quantifying uncertainty in multispecies fisheries : hierarchical Bayesian models for stratified sampling protocols with missing data. *Can. J. Fish. Aquat. Sci.* 69, 231-246.
[19] Grzegorczyk M & Husmeier D 2013 Regularization of non-homogeneous dynamic Bayesian networks with global information-coupling based on hierarchical Bayesian models. *Mach Learn* 91, 105-154.
[20] Glassen T & Nitsch V 2016 Hierarchical Bayesian models of cognitive development 110, 217-227.
[21] Barndorff-Nielsen O & Schou G 1973, On the parametrization of autoregressive models by partial autocorrelation. *J. Multivar. Anal.* 3, 408-419.
[22] Green P J (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 711–732.
[23] Brockwell P J & Davis R A 1991, *Time Series: Theory and Methods*. (New York: Springer-Verlag).

Hierarchical Bayesian segmentation for piecewise stationary autoregressive model based on reversible jump MCMC

ORIGINALITY REPORT

6%

SIMILARITY INDEX

PRIMARY SOURCES

1	Alex G. Papadopoulos. "A hierarchical approach to the study of the exponential failure model", Communications in Statistics - Theory and Methods, 1989 Crossref	24 words — 1%
2	Cong Han. "Markov Chain Monte Carlo Methods for Computing Bayes Factors : A Comparative Review", Journal of the American Statistical Association, 09/2001 Crossref	20 words — 1%
3	Thulasyammal Ramiah Pillai, Murali Sambasivan. "Chapter 9 Comparative Analysis of ARMA and GARMA Models in Forecasting", Springer Nature, 2017 Crossref	19 words — 1%
4	Keizo Mizuno. "Industrial diversity and metropolitan unemployment rate", The Annals of Regional Science, 03/2006 Crossref	12 words — 1%
5	J.P. Kaipio. "Simulation of nonstationary EEG", Biological Cybernetics, 05/30/1997 Crossref	12 words — 1%
6	Richard A. Davis. "Break Detection for a Class of Nonlinear Time Series Models", Journal of Time Series Analysis, 09/2008 Crossref	9 words — < 1%
7	M. Sambridge, K. Gallagher, A. Jackson, P. Rickwood. "Trans-dimensional inverse problems, model comparison and the evidence", Geophysical Journal	9 words — < 1%

International, 2006
Crossref

8	wei.perso.enseeiht.fr Internet	8 words — < 1%
9	pure.uva.nl Internet	8 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	ON		