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

Suparman*  
Department of Mathematics Education, University of Ahmad Dahlan, Indonesia  
*Corresponding author: 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. A constant exponential regression per segment is used to model the excess mortality rate data [1]. A constant autoregressive model per segment is applied to the cell-cycle data [2]. A constant autoregressive model per segment is used to model the Financial Times Stock Exchange (FTSE) data [3]. A constant regression model per segment is applied to the long-term observational data [4]. A stationary model per segment is used to model electroencephalogram data [5].

The method for constant model per segment segmentation is examined by several authors. A reversible jump MCMC method is used to segment sound signals [6]. A fuzzy c-means clustering is used to segment audio signals [7]. A Bayesian method is used to segment tumor data [8]. A spectogram is used to segment seismic data [9]. A time frequency decomposition is used to segment seismic data [10]. An adaptive segmentation method is used to segment photoplethysmography data [11]. An analysis of time frequency maps of group delay is used to segment and cluster seismic data [12]. The Bayesian minimum description length (BMDL) method is used to segment temperature data [13]. A domain assisted parameter semi-free wave mining (DAPs) model is used to segment data epileptic activity data [14].

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. Methods

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

\[
x_t = \tau_t - \sum_{i=1}^{k} \phi_i \cdot \varphi_t - \xi_t, \quad \tau_{i,k} < t \leq \tau_{i+1,k}, \quad i = 0, 1, ..., k
\]

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

\[
\phi(a) = 1 - \exp(-a \rho)
\]

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 \( s = (x_{b_m-1}, ..., x_n) \) is a realization of a piecewise stationary constant AR model. If the values of \( \sigma_0^2 \) and \( \sigma_{\tau_t}^2 \) are known and

\[
\theta = \{ k, \tau^{(k)}_i \}
\]

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

\[
L(s|\theta, \sigma_0^2) = \prod_{t=1}^{n}(2\pi\sigma_{\tau_t}^2)^{-1/2} \exp\left(-\frac{1}{2\sigma_{\tau_t}^2} \sum_{t=\tau_i+1}^{\tau_i+k} \left(x_t - \sum_{j=1}^{\tau_i+k} G^{-1}(\phi_{i,k}) x_{t-j}\right)^2\right)
\]

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

\[
F: \phi_{i,k} \mapsto \rho_{\phi_{i,k}} \in (-1, 1)_{\phi_{i,k}}
\]

The AR model is stationary if and only if \( \rho_{\phi_{i,k}} \in (-1, 1)_{\phi_{i,k}} \). If

\[
\rho = \{ k, \tau^{(k)}_i, \phi_{i,b} \}
\]

then the likelihood function can be rewritten as:

\[
L(s|\rho, \sigma_0^2) = \prod_{t=1}^{n}(2\pi\sigma_{\tau_t}^2)^{-1/2} \exp\left(-\frac{1}{2\sigma_{\tau_t}^2} \sum_{t=\tau_i+1}^{\tau_i+k} \left(x_t - \sum_{j=1}^{\tau_i+k} G^{-1}(\phi_{i,k}) x_{t-j}\right)^2\right)
\]

3.1. Hierarchical Bayesian Approach

The hierarchical Bayesian approach is used by several authors, for example [16-21]. 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 \( \lambda \) is selected as the prior distribution \( \pi(k|\lambda) \), namely:

\[
\pi(k|\lambda) = (\lambda)^k (1-\lambda)^{k_{\text{max}}-k} \quad \text{for} \quad k = 1, ..., k_{\text{max}} \text{ and is 0 for others.}
\]

2. Suppose that \( \pi(\tau_i|k) \) is the prior distribution for the position \( \tau_i \). 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 \( \pi(\tau_i|k) \).
3. Suppose that $\pi(p_{ik}|k)$ is prior distribution for order $p_{ik}$. A uniform distribution in $\{1, \ldots, p_{max}\}$ is selected as the prior distribution $\pi(p_{ik}|k)$.

4. Suppose that $\rho^\text{upl}_{ik}$ is a conditional prior distribution for the coefficient vector $\rho^\text{upl}_{ik}$ if it is known $p_{ik}$. A uniform distribution in the interval $(-1,1)^{p_{ik}}$ is selected as the prior distribution $\pi(\rho^\text{upl}_{ik}|p_{ik})$.

5. Suppose that $\pi(\sigma^2_{\nu i}|k, \alpha, \beta)$ is the prior distribution for $\sigma^2_{\nu i}$. An inverse gamma distribution with the parameters $\alpha$ and $\beta$ is selected as the prior distribution $\pi(\sigma^2_{\nu i}|k, \alpha, \beta)$, namely:

$$\pi(\sigma^2_{\nu i}|k, \alpha, \beta) = \frac{\Gamma(\alpha/2)}{\Gamma(\alpha)} \frac{1}{(2\pi)^{\alpha/2}} \frac{1}{\sigma^\alpha_{\nu i}} \exp(-\beta(2\sigma^2_{\nu i})^{-\alpha})$$

for $\sigma^2_{\nu i} > 0$ 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 $\lambda$. The distribution $\pi(\lambda)$ is assumed to be uniformly distributed at intervals $(0.1)$. Suppose that $\pi(\beta)$ is the prior distribution for hyperparameter $\beta$. The distribution $\pi(\beta)$ is assumed to be Jeffrey's distribution. Here, the value of $\alpha$ is equal to 2. Let $H_1 = \{k, \tau_{ik}, p_{ik}, \rho^\text{upl}_{ik}, \sigma^2_{\nu i}\}$ 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 $H_1$ can be expressed as

$$\pi(H_1, H_2) = \pi(k|\lambda) \pi(\lambda|\tau_{ik}|k) \pi(p_{ik}|k) \pi(\rho^\text{upl}_{ik}|p_{ik}) \pi(\sigma^2_{\nu i}|k, \alpha, \beta) \pi(\beta)$$

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

As in [22], the reversible jump algorithm MCMC 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 $M_1, \ldots, M_m$ that satisfies aperiodic and irreducible properties so that the Markov chain $M_1, \ldots, M_m$ can be considered as random variables that follow the distribution $\pi(H_1, H_2|s)$. The Markov chain $M_1, \ldots, M_m$ can be used to calculate the estimation of parameter $M$. The Markov chain is made in two stages: stage 1 simulates the distribution $\pi(H_1|H_2, s)$ and stage 2 simulates the distribution $\pi(H_1|H_2, s)$. Distribution $\pi(H_1|H_2, s)$ 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 $H_1$ known $(H_2, s)$ can be written as:

$$\pi(H_1|H_2, s) = B(k + 1, k_{max} - k + 1) \otimes G\left(\frac{-}{\tau}, (k + 1), \frac{-}{\tau} \right)$$

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 $\pi(k, \tau(\cdot), p_{ik}, \rho^\text{upl}_{ik}|(p_{ik})^k)$, the stage two simulates distribution $\pi(p_{ik}, \rho^\text{upl}_{ik}|(p_{ik})^k)$, and the stage three simulates distribution $\pi(\sigma^2_{\nu i}|(p_{ik})^k)$.

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

| $i^{th}$ segment | $\sigma_{i,4}$ | $p_{i,4}$ | $\theta_{i,4}^{[\sigma_{i,4}]}$ |
|------------------|----------------|-----------|--------------------------|
| 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.
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 $\tau$ corresponding to $k = 4$ is given in Figure 3. The result is $\hat{\tau} = (75, 150, 250, 400)$.

The segmentation results are presented in Figure 4.

![Figure 3. Histogram for $\tau$ if known $k = 4$](image)

![Figure 4. Segmentation of data](image)

The estimation results of coefficient and deviation of stochastic error standards for each segment are written in Table 2.

| ith segment | $\hat{\sigma}_{i,4}$ | $\hat{p}_{i,4}$ | $\hat{\theta}_{i,4}$ |
|-------------|---------------------|-----------------|---------------------|
| 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].

![Figure 5. Dow-Jones Utilities Index](image)

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.

![Figure 6. Histogram for k](image)

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.

| ith segment | $\hat{\theta}_{i,0}$ | $\hat{\rho}_{i,0}$ | $\hat{\sigma}_{i,0}$ |
|-------------|----------------|----------------|----------------|
| 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).

Acknowledgment
The author would like to thank the reviewers who have provided suggestions for improving the quality of this paper.

References
[1] Luque-Fernandez M A, Belot A, Quaresma M, Maringe C, Coleman M P, and Rachet B 2016 16 1-8
[2] Rueda C, Fernandez M A, Barragan S, Mardia K V and Peddada S D 2016 72 1266-74
[3] Fryzlewicz P and Rao S S 2014 J.R. B 76 903-24
[4] Buscot M-J, Wotherspoon S S, Magnusses C G, Juonala M, Sabin M A, Burgner D P, Lehtimaki T, Vikari J S A, Hutri-Kahonen N, Raitakari O T and Thomson R J 2017 17 1-15
[5] Yau C Y and Zhao Z 2016 J.R. Statist. Soc. B 78 895-916
[6] Punskaya E, Andrieu C, Doucet A and Fitzgerald W J 2002 IEEE Trans. 50 747-58
[7] Nitanda N, Haseyama M and Kitajima H 2006 Syst. Comput. Jpn. 37 302-12
[8] Tai Y C, Kvale M N and Witte J S 2010 Biometrics 66 675-83
[9] Phojarongmongkolkij N 2014 J. clim. 27 3363-76
[10] Zimroz R, Madziarz M, Zak G, Wylomanska A and Obuchowski J 2015 J. Vibroengineering 17 3111-21
[11] Kayosaoglu A R, Polat K and Bozkurt M R 2016 Turkish J. Electr. Eng. Comput. Sci. 24 1782-96
[12] Polak M, Obuchowski J, Madziarz M, Wylomanska A and Zimroz R 2016 J. Vibroengineering 18 267-75
[13] Hewaarachchi A P, Li Y, Lund R and Rennie J 2017 Am. Meteorol. Soc. 30 985-99
[14] Kim S-H, Li L, Faloutsos C, Yang H-J and Lee S-W 2017 J. Inf. Sci. Eng. 33 517-36
[15] Suparman 2018 Inter. J. of GEOMATE 15, 85-91
[16] Cheng W, Ma L, Yang T, Liang J, and Zhang Y 2016 PloS ONE 11 1-13
[17] Rodriguez de Rivera O, Lopez-Quilez and Blangiardo M 2018 Forests 9 1-17
[18] Arzaghi E, Abaei M M, Abbassi R, Garaniya V, Binns J, Chin C and Khan F 2018 Process Saf. Environ. Prot. 118 307-15
[19] Shelton A O, Dick E J, Pearson D E, Ralston S and Mangel M 2012 Can. J. Fish. Aquat. Sci. 69 231-46
[20] Grzegorczyk M and Husmeier D 2013 Mach Learn 91 105-54
[21] Glassen T and Nitsch V 2016 Biol Cybern 110, 217-27
[22] Suparman and Doisy M 2018 TELKOMNIKA 16, 673–80
[23] Brockwell P J and Davis R A 2009 Time Series: Theory and Methods (New York: Springer Science & Business Media)