SEGMETASI BAYESIAN HIRARKI UNTUK MODEL AR STASIONER KONSTAN PER SEGMEN MENGGUNAKAN ALGORITMA *REVERSIBLE JUMP* MCMC

Suparman
Pendidikan Matematika FKIP UAD
Jl. Prof. Dr. Soepomo, SH Warungboto Yogyakarta
Telp : 081328201198
E-mail : suparmancict@yahoo.co.id

Abstrak

Makalah ini membahasmasalahsegmentasidata dalamkerangkaBayesiandengan menggunakansampling reversibeljumpMCMC.Datadimodelkanolehmodel autoregresif konstan sepotong demi sepotong, di mana banyaknya segmen,orde dankoefisien prosesARuntuk setiap segmentidak diketahui. Algoritmareversible jump MCMCkemudian digunakanuntuk menghasilkan sampelyang didistribusikansesuai dengan distribusiposteriorgabungan dariparameter yang tidak diketahui. Sampel inimemungkinkanuntuk menghitungbeberapa fiturmenarik daridistribusiposterior.Kinerjametodeini diilustrasilan denganbeberapahasilsimulasi. Dari hasil simulasi menunjukkan bahwa algoritma reversible jump MCMC dapat mengestimasi parameter model AR stasioner konstan per segmen dengan baik.

***Kata Kunci**:* Bayesian, *Reversible jump* MCMC, *segmentasi,* AR.

Abstract

This paper addresses the problem of the data segmentation within a Bayesian framework by using reversible jump MCMC sampling. The data is modeled by piecewise constant Autoregressive (AR) processes where the numbers of segments, the time of abrupt, the order and the coefficients of the AR processes for each segment are unknown. The reversible jump MCMC algorithm is then used to generate samples distributed according to the joint posterior distribution of the unknown parameters. These samples allow to compute some interesting features of the a posterior distribution. The performance of the this methodology is illustrated via several simulation results.The results of simulation show that the reversible jump MCMC algorithm can estimate the parameters of piecewise constant autoregressive well.

***Keywords** :* Bayesian, *Reversible jump* MCMC, *segmentation,* AR.

1. PENDAHULUAN

Model *autoregresif* (AR) stasioner konstan per segmen merupakan model yang sering digunakan untuk memodelkan berbagai jenis data. Data indeks Dow-Jones, data indeks harga konsumen (IHK) dan data laju inflasi merupakan dua contoh data riil yang dapat dimodelkan oleh model AR stasioner konstan per segmen. Apabila model AR stasioner konstan per segmen dicocokkan terhadap data riil, umumnya parameter model tidak diketahui. Parameter model di sini meliputi : banyaknya segmen, waktu terjadinya perubahan model AR dan parameter model AR untuk tiap-tiap segmen. Parameter model AR meliputi : orde, koefisien dan variansi gangguan stokhastik.

Pendekatan yang digunakan dalam penelitian ini adalah pendekatan Bayesian. Parameter model dipertimbangkan sebagai variabel random yang mempunyai distribusi tertentu. Distribusi ini dikenal sebagai distribusi prior. Distribusi prior dari parameter model dan fungsi kemungkinan dari sinyal dikombinasikan untuk mendapatkan distribusi posterior dari parameter model. Estimasi Bayesian didasarkan pada distribusi posterior.

Distribusi posterior mempunyai bentuk yang sangat rumit menyebabkan penentuan estimator tidak dapat dilakukan secara analitis. Untuk mengatasi masalah ini, digunakan algoritma *reversible jump Markov Chain Monte Carlo* (MCMC).

Uraian dalam artikel ini disusun sebagai berikut. Dalam Seksi 2, dibahas metode yang mencakup model AR Stasioner konstan per segmen, pendekatan Bayesian, algoritme *reversible jump* MCMC dan penurunan rumus. Sedangkan hasil dan diskusi dijelaskan dalam Seksi 3. Dalam seksi 3 diuraikan implementasi dari algoritma *reversible jump* MCMC pada data síntesis dan data riil. Akhirnya kesimpulan dari hasil penelitian disajikan dalam Seksi 4.

2. METODE

2.1 MODEL AR STASIONER KONSTAN PER SEGMEN

Misalkan X_1, X_2, …, X_n adalah data. Data ini dikatakan mempunyai model AR konstan per segmen dengan banyaknya segmen k (k = 0, 1, …, k_{max}) apabila (untuk t = 1, 2, …, n) data tersebut memenuhi persamaan stokhastik berikut ([11]) :

$$X_t = Z_t - \sum_{j=1}^{q_{i,k}} \theta_{i,k,j}^{(q_{i,k})} X_{t-j}, \quad \tau_{i,k} < t \le \tau_{i+1,k}, \quad i = 0,1,\cdots,k \qquad (1)$$

di mana di bawah asumsi k segmen : $\tau_{i,k}$ adalah waktu terjadinya perubahan model AR ke-i, dengan konvensi $\tau_{0,k} = 0$ dan $\tau_{k+1,k} = n$ dan untuk tiap-tiap segmen ke-i :

- $q_{i,k}$ dan $\theta_{i,k}^{(q_{i,k})} = \left(\theta_{i,k,1}^{(q_{i,k})},\cdots,\theta_{i,k,q_{i,k}}^{(q_{i,k})}\right)$ adalah orde dan koefisien model AR yang bersesuaian dengan segmen ke-i.
- Z_t adalah nilai gangguan stokhastik pada saat t yang bersesuaian dengan segmen ke-i. Z_t dimodelkan sebagai distribusi normal dengan mean 0 dan variansi $\sigma_{i,k}^2$.

Selanjutnya model AR ke-i (i=0,1, …, k) disebut stasioner jika dan hanya jika persamaan suku banyak

$$\phi(b) = 1 - \sum_{j=1}^{q_{i,k}} \theta_{i,k,j}^{(q_{i,k})} b$$

bernilai nol untuk nilai b di luar lingkaran dengan jari-jari sama dengan satu ([3]).

Apabila banyaknya segmen k diasumsikan diketahui, waktu terjadinya perubahan model AR diasumsikan diketahui dan orde yang diasumsikan diketahui, maka permasalahan inferensi model AR konstan per segmen menjadi permasalahan identifikasi orde dan estimasi parameter model AR untuk tiap-tiap segmen.

Apabila orde model AR diasumsikan diketahui, maka permasalahan identifikasi orde model AR dan estimasi parameter model AR menjadi permasalahan estimasi parameter model AR. Estimasi parameter model AR dapat dilakukan dengan menggunakan berbagai metode. Metode-metode tersebut diantaranya diusulkan oleh [10], [3], [2] dan [13]. Metode Bayesian digunakan untuk mengestimasi parameter AR [10]. Sedangkan ketiga peneliti lainnya, [3], [2] dan [13], menggunakan Metode Kemungkinan Maksimum untuk mengestimasi parameter model AR. Selanjutnya metode identifikasi orde dan estimasi parameter model AR diusulkan oleh [12].

Dalam penelitian ini, banyaknya segmen dan orde model AR untuk masing-masing segmen diasumsikan tidak diketahui. Algoritma *reversible jump* MCMC ([4]) digunakan untuk mendeteksi banyaknya segmen, lokasi perubahan model AR, mengidentifikasi orde model AR dan mengestimasi parameter model AR secara bersamaan dalam satu tahap. Untuk mengatasi masalah hiperparameter yang muncul, diadopsi Bayesian hirarkis ([9]). Kinerja algoritma yang diusulkan akan diuji dengan menggunakan data sintesis.

2.2 METODE BAYESIAN HIRARKI

Andaikan $s = \left(x_{q_{makx}+1}, x_{q_{maks}+2}\cdots, x_n\right)$ adalah suatu realisasi dari model AR konstan per segmen. Jika nilai $s_0 = \left(x_1, x_2\cdots, x_{q_{maks}}\right)$ diketahui dan $\theta = \left(k, \tau^{(k)}, \{\theta_{i,k}^{(q_{i,k})}\}_{i=0}^{k}, \sigma^{(k)}\right)$, maka fungsi kemungkinan dari s dapat ditulis kurang lebih sebagai berikut :

$$\ell(s|\theta) = \prod_{i=0}^{k} (2\pi\sigma_{i,k}^2)^{-\frac{1}{2}(\tau_{i+1,k}-\tau_{i,k})} \exp - \frac{1}{2\sigma_{i,k}^2} \sum_{t=\tau_{i,k}+1}^{\tau_{i+1},k} (x_t - \sum_{j=1}^{q_{i,k}} G(\rho_{i,k}^{(q_{i,k})}) x_{t-j})^2 \qquad (2)$$

Untuk $t = q_{maks+1},\cdots,n$.

Misalkan $S_{q_i,k}$ adalah daerah stasionaritas. Dengan menggunakan transformasi

$$F: \theta_{i,k}^{(q_{i,k})} \in I_q \quad \mapsto \quad \rho_{i,k}^{(q_{i,k})} \in (-1,1)^{q_{i,k}} \qquad (3)$$

maka model AR $(X_t)_{t\in Z}$ stasioner jika dan hanya jika $\rho_{i,k}^{(q_{i,k})} \in (-1,1)^{q_{i,k}}$ [1]. Apabila $\rho = \left(k, \tau^{(k)}, \left\{\rho_{i,k}^{(q_{i,k})}\right\}_{i=0}^{k}, \sigma^{(k)}\right)$, maka fungsi kemungkinan dapat ditulis kembali sebagai :

$$\ell(s|\rho) = \prod_{i=0}^{k} (2\pi\sigma_{i,k}^2)^{-\frac{1}{2}(\tau_{i+1,k}-\tau_{i,k})} \exp - \frac{1}{2\sigma_{i,k}^2} \sum_{t=\tau_{i,k}+1}^{\tau_{i+1},k} (x_t - \sum_{j=1}^{q_{i,k}} G^{-1}(\theta_{i,k}^{(q_{i,k})}) x_{t-j})^2 \qquad (4)$$

Penentukan distribusi prior untuk parameter-parameter tersebut di atas adalah sebagai berikut :

a) Banyaknya segmen k berdistribusi Binomial dengan parameter λ

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

b) Posisi τ_k berdistribusi indeks genap dari statistik terurut 2k+1 yang diambil seragam tanpa pengembalian dalam $\{1,2,\cdots,n-1\}$.

c) Orde $p_{i,k}$ berdistribusikan seragam dalam $\{0,1,\cdots,p_{maks}\}$.

d) Untuk orde $p_{i,k}$ ditentukan terlebih dahulu, vektor koefisien $\rho_{i,k}^{p_{i,k}}$ berdistribusikan seragam pada interval $(-1,1)^{p_{i,k}}$.

e) Variansi $\sigma_{i,k}^2$ berdistribusikan invers gamma dengan parameter $\alpha/2$ dan $\beta/2$:

$$\pi(\sigma_{i,k}^2 \mid \alpha,\beta) = \frac{(\beta/2)^{\alpha/2}}{\Gamma(\alpha/2)} (\sigma_{i,k}^2)^{-(1+\alpha/2)} \exp - \beta/(2\sigma_{i,k}^2)$$

Di sini parameter λ diasumsikan berdistribusi seragam pada interval (0,1), nilai α diambil sama dengan 2 dan parameter β diasumsikan berdistribusi Jeffrey. Sehingga distribusi prior untuk parameter $H_1 = (q_{i,k}, \rho_{i,k}^{q_{i,k}}, \sigma_{i,k}^2)$ dan $H_2 = (\lambda,\beta)$ dapat dinyatakan sebagai :

$$\pi(H_1,H_2) = \pi(q_{i,k} \mid \lambda)\pi(\rho_{i,k}^{q_{i,k}} \mid q_{i,k})\pi(\sigma_{i,k}^2 \mid \alpha,\beta)\pi(\lambda)\,\pi(\beta) \tag{5}$$

Menurut Teorema Bayes, maka distribusi a posteriori untuk parameter H_1 dan H_2 dapat dinyatakan sebagai :

$$\pi(H_1,H_2 \mid s) \propto \ell(s \mid H_1)\pi(H_1,H_2) \tag{6}$$

Distribusi a posteriori merupakan gabungan dari fungsi kemungkinan dan distribusi prior yang kita asumsikan sebelum sampel diambil. Dalam kasus ini, distribusi a posteriori $\pi(H_1,H_2 \mid s)$ mempunyai bentuk yang sangat rumit sehingga tidak dapat diselesaikan secara analitis. Untuk mengatasi masalah tersebut, diusulkan metode *reversible jump* MCMC.

2.3 METODE REVERSIBLE JUMP MCMC

Misalkan M = (H_1, H_2). Secara umum, metode MCMC merupakan suatu metode sampling, yaitu dengan cara membuat rantai Markov homogen $M_1, M_2, \cdots, M_m$ yang memenuhi sifat aperiodik dan irreduktibel ([8]) sedemikian hingga $M_1, M_2, \cdots, M_m$ dapat dipertimbangkan sebagai variabel acak yang mengikuti distribusi $\pi(H_1,H_2 \mid s)$. Dengan demikian $M_1, M_2, \cdots, M_m$ dapat digunakan sebagai sarana untuk menaksir parameter M. Untuk merealisasikan itu diadopsi algoritma Gibbs Hibrida ([8]) yang terdiri dari dua tahap : tahap 1, simulasi distribusi $\pi(H_2 \mid H_1, s)$ dan tahap 2, simulasi distribusi $\pi(H_1 \mid H_2, s)$

Distribusi $\pi(H_2 \mid H_1, s)$ mempunyai bentuk eksplisit. Sehingga Algoritma Gibbs dapat digunakan untuk mensimulasikan distribusi $\pi(H_2 \mid H_1, s)$. Distribusi marginal posterior dari H_2 dapat dituliskan sebagai :

$$\pi(H_2 \mid H_1, s) = B(k+1, k_{maks} - k + 1) \otimes G(\frac{\alpha}{2}(k+1), \sum_{i=0}^{k} \frac{1}{2\sigma_{i,k}^2})$$

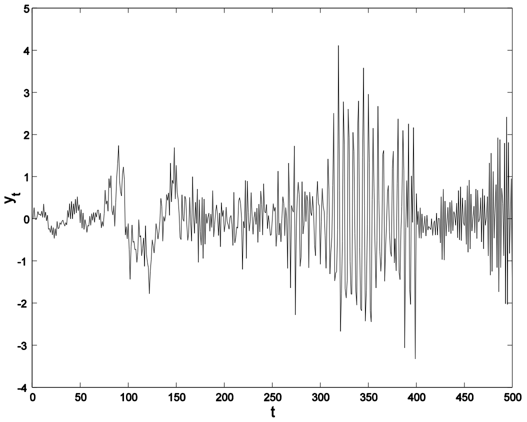
Sebaliknya, distribusi $\pi(H_1 \mid H_2, s)$ tidak mempunyai bentuk eksplisit. Sehingga simulasi eksak tidak mungkin dilakukan. Untuk itu, dusulkan algoritma hibrida, yang mengabungkan algoritma *Reversible Jump* MCMC ([4]) dengan algoritma Gibbs, untuk mensimulasikan distribusi $\pi(H_1 \mid H_2, s)$. Algoritma *Reversible Jump* MCMC merupakan rampatan dari algoritma Metropolis-Hastings ([7]; [6]). Algoritma hibrida ini terdiri dari tiga tahap : simulasi $\pi(k, \tau_k, p_k, \rho_k^{(p_k)} \mid H_2, s)$, simulasi $\pi(p_k, \rho_k^{(p_k)} \mid k, \tau_k H_2, s)$ dan simulasi $\pi(\sigma_k^2 \mid k, \tau_k, p_k, \rho_k^{(p_k)}, H_2, s)$. Karena harga k dan p_k tidak diketahui maka pada Tahap 2.1 dan Tahap 2.2 digunakan Algoritma *reversible jump* MCMC.

3. HASIL DAN PEMBAHASAN

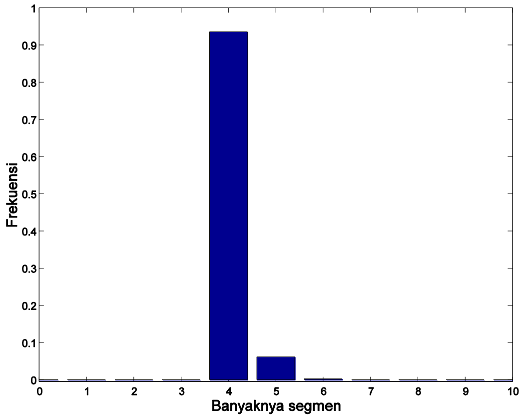
Sebagai ilustrasi, kita akan menerapkan metode ini untuk mengidentifikasi orde dan menaksir parameter data sintesis. Algoritma *reversible jump* MCMC digunakan untuk mengestimasi banyaknya segmen, waktu terjadinya perubahan model AR, orde model AR untuk masing-masing segmen, dan koefisien model AR untuk masing-masing model AR serta variansi gangguan stokhastik yang bersesuaiaan. Untuk keperluan itu, algoritma *reversible jump* MCMC dimplementasikan 70000 iterasi dengan periode pemanasan 10000 iterasi. Nilai orde q_{maks} dibatasi maksimum 10 sehingga p_{maks} = 10.

3.1 Data Sintesis

Gambar 1 merupakan data sintesis dengan model AR konstan per segmen yang dibuat menurut persamaan (1) di atas.



Gambar 1 : Data sintesis model AR konstan per segmen



Gambar 2 : Histogram dari banyaknya segmen

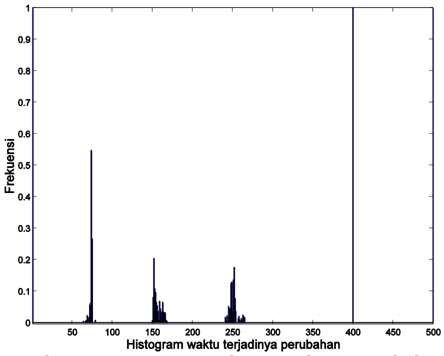
Pembuatan sinyal sintesis dilakukan dengan menggunakan bahasa pemograman MATLAB ([5]), dengan jumlah data n = 500, k = 4 dan waktu terjadinya perubahan model AR adalah $\tau = (75, 150, 250, 400)$. Sedangkan orde, koefisien, dan gangguan stokhastik model AR untuk masing-masing segmen dinyatakan dalam Tabel 1.

Tabel 1 : Nilai parameter model

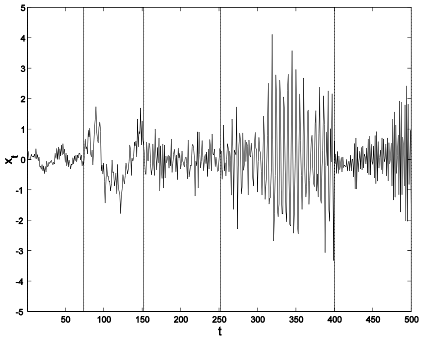
Segmen ke-i	$\sigma_{i,4}$	$p_{i,4}$	$\theta_{i,4}^{(p_i,1)}$
0	0.12	3	(-0.25, -0.79, 0,34)
1	0,5	2	(-1.54, -0.41)
2	0,4	1	(0.19)
3	0,5	4	(0.59, 0.99, 0.64, 0.87)
4	0,12	3	(0.86, -0.83, -0.96)

Berdasarkan data dalam Gambar 1, selanjutnya parameter model diestimasi dengan menggunakan *reversible jump* MCMC. Histogram dari k disajikan pada Gambar 2. Hasilnya adalah $\hat{k} = 4$. Histogram untuk τ yang bersesuaian dengan nilai $\hat{k} = 4$ diberikan pada Gambar 3. Hasilnya adalah $\hat{\tau} = (75, 150, 250, 400)$.

Hasil segmentasi disajikan dalam Gambar 4.

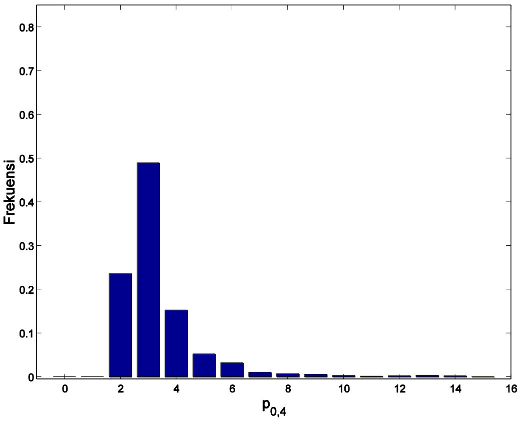


Gambar 3 : Histogram waktu terjadinya perubahan Model

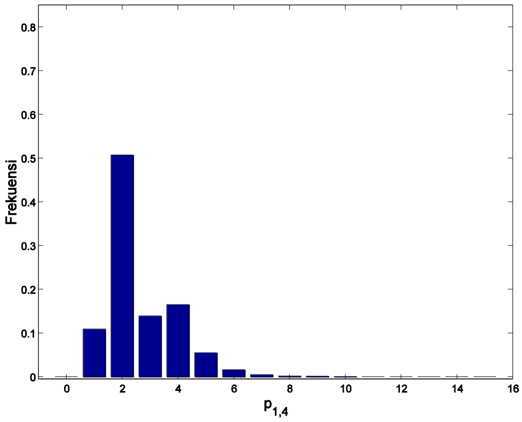


Gambar 4 : Segmentasi data

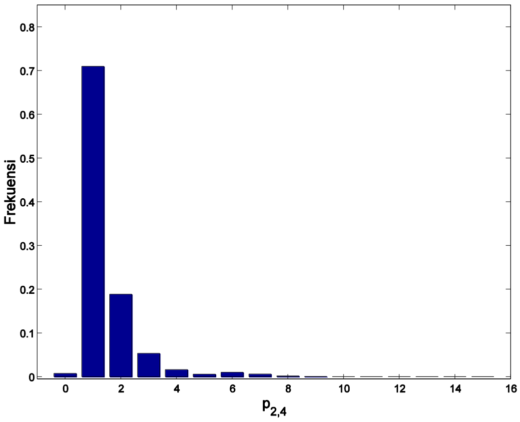
Histogram untuk orde yang bersesuaian dengan nilai $\hat{k} = 4$ dan diberikan pada Gambar 5 - 9.



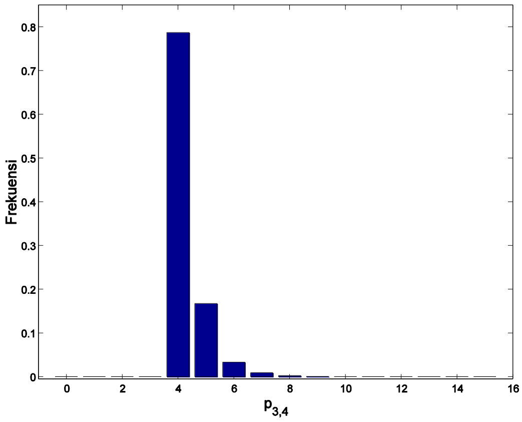
Gambar 5 : Histogram orde segmen ke-0



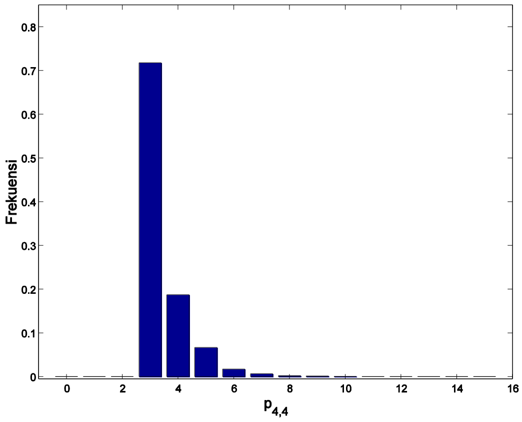
Gambar 6 : Histogram orde segmen ke-1



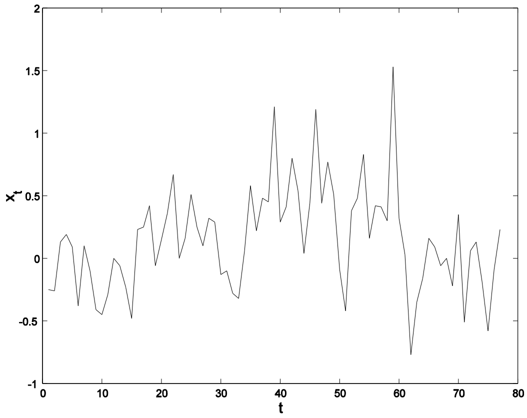
Gambar 7 : Histogram orde segmen ke-2



Gambar 8 : Histogram orde segmen ke-3



Gambar 9 : Histogram orde segmen ke-4



Gambar 10 : Data real model AR

Sedangkan hasil estimasi dari koefisien dan simpangan baku gangguan stokhastik tiap-tiap segmen ditulis dalam Tabel 2.

Tabel 2 : Estimator untuk orde, koefisien dan simpangan baku gangguan stokhastik.

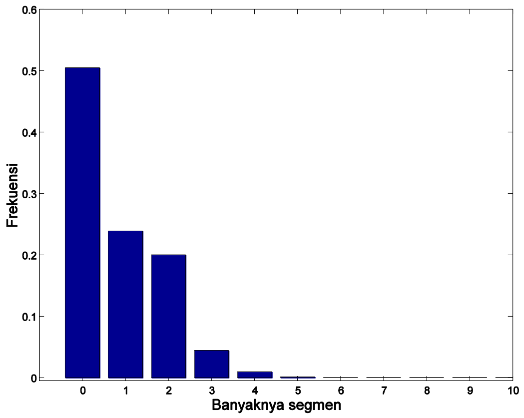
Segmen ke-i	$\hat{\sigma}_{i,4}$	$\hat{p}_{i,4}$	$\hat{\theta}_{i,4}^{(\hat{p}_{i,4})}$
0	0.13	3	(-0.23, -0.76, 0.23)
1	0.47	2	(-0.50, -0.27)
2	0.41	1	(0.34)
3	0.52	4	(0.57, 0.93, 0.62, 0.83)
4	0.13	3	(0.86, -0.79, -0.94)

Berdasarkan *output* dari algoritma, pada data Gambar 1 terbagi atas 5 segmen. Pada segmen pertama (t = 1, 2, ..., 74) data bermodel AR(3), segmen kedua (t = 75, 76, ..., 149) data bermodel AR (2), segmen ketiga (t

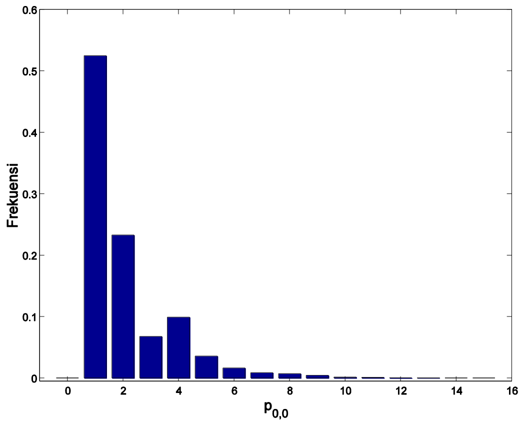
= 150, 151, ..., 249) data bermodel AR(1), segmen keempat (t = 250, 251, ..., 399) data bermodel AR (4) dan segmen kelima (t = 401, 402, ..., 499) data bermodel AR (3).

3.2 Data Riil

Gambar 11 merupakan sinyal riil berupa evolusi indeks Dow-Jones ([3]). Berdasarkan data dalam Gambar 11, selanjutnya parameter model diestimasi dengan menggunakan *reversible jump* MCMC. Histogram dari k disajikan pada Gambar 12. Hasilnya adalah $\hat{k} = 0$.



Gambar 11 : Histogram dari banyaknya segmen



Gambar 12 : Histogram orde segmen ke-0

Oleh karena $\hat{k} = 0$ maka tidak ada estimasi untuk lokasi. Sehingga histogram untuk waktu terjadinya perubahan model adalah tidak ada. Histogram untuk orde yang bersesuaian dengan nilai $\hat{k} = 0$ dan diberikan pada Gambar 12.Hasil estimasi untuk koefisien dan simpangan baku gangguan stokhastik ditulis dalam Tabel 3.

Tabel 3 : Estimator untuk orde, koefisien dan simpangan baku gangguan stokhastik.

Segmen ke-i	$\hat{\sigma}_{i,0}$	$\hat{p}_{i,0}$	$\hat{\theta}_{i,0}^{(\hat{p}_{i,0})}$
0	0.39	1	-0.46

4. SIMPULAN DAN SARAN

Uraian di atas, merupakan kajian teori tentang algoritma *reversible jump* MCMC dan penerapannya pada inferensi model AR konstan per segmen. Dengan membandingkan antara nilai parameter dan nilai estimasinya dari data sintesis menunjukkan bahwa algoritma *reversible jump* MCMC dapat menaksir parameter-parameter itu dengan baik. Estimator untuk orde, koefisien, dan gangguan stokhastik model AR untuk masing-masing segmen disajikan dalam Tabel 2. Sebagai contoh implementasi, algoritma *reversible jump* MCMC diterapkan pada data riil.
Penelitian ini masih dapat diperluas dan dikembangkan dengan cara menggantikan konsep AR konstan per segmen dengan konsep ARMA (*autoregressive moving average*) konstan per segmen.

5. DAFTAR PUSTAKA

[1] Barndorff-Nielsen, O. and Schou, G. 1973. On the parametrization of autoregressive models bypartial autocorrelation, *J. Multivar. Anal.*, Vol. 3, 408-419.
[2] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. 1994. *Time Series Analysis : Forecasting and Control*,Prentice Hall, New Jersey.
[3] Brockwell, P.J. and Davis, R.A. 1991. *Times Series : Theory and Methods*, Springer, New York.
[4] Green, P.J. 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian ModelDetermination, *Biometrika*, Vol. 82, 711-732.
[5] Hanselman, D. and Littlefield, B. 1977. *Matlab*, Pearson Education Asia and Andi.
[6] Hastings, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, Vol. 57, 97-109.
[7] Metropolis, N., Rosenbluth, A.W., Teller, A.H. and Teller, E. 1953. Equations of state calculations by fast coputing machines, *Journal Chemical Physics*, Vol 21, 1087-1091.
[8] Robert, C.P., 1996. *Méthodes de Monte Carlo par Chaînes de Markov*, Economica.
[9] Robert, C.P., 1999. *The Bayesian Choice. A Decision-Theoretic Motivation*, Springer Texts in Statistics.

[10] Shaarawy, S. and Broemeling, L. 1984. Bayesian inferences and forecasts with moving averages processes. *Commun. Statist. – Theory Meth.*, 13(15), 1871-1888.
[11] Suparman and Doisy, M 2002. Bayesian Segmentation of Piecewise Constant Moving-Average Processes using Reversible Jump MCMC Methods*Proc. Of the 7th Indonesian Student's Scientific Meeting*, pp. 481-485, Berlin Germany.
[12] Suparman 2006 Identifikasi dan estimasi Bayesian hierarki dalam runtun waktu AR dengan menggunakan algoritma SA, *Jurnal Pakar*, Vol. 7 No. 1 hal. 17-28.
[13] Suparman dan Soejoeti, Z. 1999. Bayesian Estimation of ARMA Time Series Models, Jurnal *WKSI*, Vol. 2 No. 3 hal. 91-98.