Jurnal Dinamika, April 2015, halaman 61-66 Vol. 06. No. 1
ISSN 2087-7889

SIMULASI PERBANDINGAN METODE PERAMALAN MODEL *GENERALIZED SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE* (GSARIMA) DENGAN *SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE* (SARIMA)

Asrirawan

Program Studi Matematika, Fakultas Sains
Universitas Cokroaminoto Palopo
Email: enalmantovani@gmail.com

ABSTRAK

Salah satu topik utama dalam kajian pemodelan peramalan deret waktu (*time series*) pada tiga dekade terakhir ini adalah peramalan data jumlahan (*count data*). Peramalan data jumlahan berbasis model stokastik masih belum banyak dilakukan dan menggunakan distribusi Gaussian (normal). Salah satu model stokastik dan non-Gaussian (negatif binomial) untuk peramalan data jumlahan adalah model *generalized autoregressive moving average* (GARMA). Model GARMA menghubungkan komponen ARMA dengan variabel prediktor ke transformasi parameter rata-rata dari distribusi data dengan menggunakan fungsi link (*link function*) tetapi tidak melibatkan efek stasioner dan musiman. Model *generalized seasonal autoregressive integrated moving average* (GSARIMA) merupakan model pengembangan dari GARMA dengan melibatkan efek stasioner dan musiman (*seasonal*). Model GSARIMA diterapkan pada data simulasi dengan membandingkan tingkat akurasi peramalan model tersebut dengan model *seasonal autoregressive integrated moving average* (SARIMA). Berdasarkan analisis yang dilakukan dengan menggunakan simulasi hasil model dan peramalan dengan menggunakan model GSARIMA relatif lebih baik dibandingkan dengan model SARIMA dengan menggunakan nilai *AIC* dan *Maref*.

Kata kunci: *GSARIMA, IRLS, binomial negatif, SARIMA*

PENDAHULUAN

Salah satu topik utama dalam kajian pemodelan peramalan deret waktu (*time series*) pada tiga dekade terakhir ini adalah peramalan data jumlahan (*count data*). Hal ini juga seperti yang telah diuraikan oleh Gooijer dan Hyndman [1] yang telah melakukan kajian literatur berkaitan dengan perkembangan peramalan deret waktu dalam kurun waktu 25 tahun, yaitu mulai 1982 sampai dengan 2005, berdasarkan 940 makalah pada jurnal-jurnal bidang peramalan. Gooijer dan Hyndman menyatakan bahwa kajian tentang peramalan data jumlahan yang berbasis model stokastik masih belum banyak dilakukan oleh para peneliti bidang peramalan.

Yu, Chen, dan Wen [3] mengemukakan bahwa ada dua pendekatan yang sering digunakan untuk memodelkan data deret waktu non-Gaussian. Pendekatan pertama adalah menggunakan kombinasi linier pada variabel acak distribusi non-Gaussian (Gaver & Lewis, 1980; Li & Mcleod, 1987) dan pendekatan yang kedua adalah mentransformasi data deret waktu Gaussian kedalam distribusi marginal yang lebih spesifik sehingga mampu digunakan untuk pendekatan non-Gaussian [2]. Hal ini juga digunakan oleh Benjamin,Rigby, dan Stasinopoulos [3]. Benjamin *et al*. mengembangkan model stokastik untuk data-data yang mengikuti distribusi non-Gaussian seperti distribusi Poisson dan binomial negatif. Model tersebut adalah model *generalized autoregressive moving average* (GARMA).

Model GARMA menghubungkan komponen ARMA dengan variabel prediktor ke transformasi parameter rata-rata dari distribusi data dengan menggunakan fungsi link (*link function*). Fungsi link ini digunakan untuk memastikan bahwa distribusi data tetap dalam domain bilangan riil positif, sehingga memiliki ketepatan prediksi yang lebih akurat. Model GARMA sangat fleksibel untuk memodelkan data jumlahan dengan struktur *autoregressive* dan atau *moving average*. Akan tetapi, hanya dapat diaplikasikan pada data yang dianggap stasioner dan tidak musiman. Benjamin *et al*. menggunakan pendekatan *iteratively reweighted least square* (IRLS) untuk mengestimasi parameter-parameter model GARMA [3].

Berdasarkan uraian di atas maka dalam penelitian ini akan dilakukan kajian tentang tingkat akurasiperamalan model GSARIMA dengan SARIMA dengan menggunakan data simulasi.

METODE PENELITIAN

Data Simulasi

Simulasi pertama adalah simulasi perbandingan estimasi model GSARIMA dengan transformasi ZQ1 dan ZQ2 [2]. Sedangkan pada simulasi kedua dilakukan perbandingan tingkat akurasi peramalan model GSARIMA dan model ARIMA musiman. Program yang digunakan untuk estimasidata simulasi adalah R, SAS dan MATLAB.

Model Binomial Negatif Data Jumlahan

Regresi binomial negatif merupakan salah satu model regresi terapan dari GLM Distribusi binomial negatif memiliki ketiga komponen yaitu komponen random,komponen sistematik dan fungsi link [2]. Adapun bentuk fungsi massa peluang binomial negatif adalah:

$$f(y_t, \mu, \alpha) = \frac{\Gamma(y + 1/\alpha)}{\Gamma(1/\alpha) y_t!} \left(\frac{1}{1 + \alpha\mu_t}\right)^{1/\alpha} \left(\frac{\alpha\mu_t}{1 + \alpha\mu_t}\right)^{y_t}$$

dengan $y_t = 0,1,2, \dots, N$, $E(y_t) = \mu_t$ dan $Var(y_t) = \mu_t + \frac{\mu_t^2}{\alpha}$.

Saat $\alpha \to 0$ maka distribusi negatif binomial memiliki variansi μ_t. Distribusi binomial negatif akan mendekati suatu distribusi Poisson yang mengasumsikan *mean* dan variansi sama yaitu$E(y_t) = Var(y_t) = \mu_t$.

Kontribusi variabel prediktor dalam model regresi binomial negatif dinyatakan dalam bentuk kombinasi linier antara parameter(η)dengan parameter regresi yang akan diestimasi yaitu:

$$\eta_t = \beta_0 + \beta_1 x_{t-1} + \cdots + \beta_k x_{t-k}$$

atau dalam matriks dituliskan dalam bentuk

$$\boldsymbol{\eta} = \boldsymbol{X}^T \boldsymbol{\beta}$$

dengan $\boldsymbol{\eta}$ adalah vektor $(n\ x\ 1)$ dari observasi, $\boldsymbol{X}$ adalah matriks $(n\ x\ c)$ dari variabel bebas,$\boldsymbol{\beta}$adalah matriks $(c\ x\ 1)$ dari koefisien regresi, dengan $c = p + 1$. Nilai ekspetasi dari variabel respon Y adalah diskrit dan bernilai positif. Maka, untuk mentransformasikan nilai η_t(bilangan riil) ke rentang yang sesuai dengan rentang pada respon y_t diperlukan suatu fungsi link $g(.)$ yaitu:

$$g(\mu_t) = \ln \mu_t = \boldsymbol{X}^T \boldsymbol{\beta}$$

Model GSARIMA

Model GSARIMA dikembangkan berdasarkan model GARMA dengan melibatkan efek musiman dan *differencing*. Diberikan$y^T = (y_t, y_{t+1}, \dots, y_{t+n})$ adalah model data deret waktu. Misalkan $y_t \sim NegBin(\mu_t, \psi)$dengan $E(y_t) = \mu_t$ dan $V(y_t) = \mu_t + \frac{\mu_t^2}{\psi}$. Jika $\cdot \to \infty$ maka y_t akan mengikuti distribusi Poisson. Adapun model GARMA (p, q)dapat dilihat pada model (2):

$$g(\mu_t) = \phi_p(B)[\boldsymbol{X}_t^T \boldsymbol{\beta} - g(y_t)] + g(y_t) - \theta_p(B)[\boldsymbol{X}_t^T \boldsymbol{\beta} - g(y_t)] + g(y_t) - g(\mu_t)$$

dimana$g(.)$adalah fungsi link,$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^pdan\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$.$B$ merupakan operator *backshift* dengan $B^d y_t = y_{t-d}$ dan vektor $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \beta_2, \dots, \beta_v)$adalah koefisien untuk vektor $\boldsymbol{X}_t^T = (x_0, x_{1,t}, \dots, x_{v,t})$dengan x_0 merupakan *intercept* yang biasanya digunakan nilai $x_0 = 1$. Pada kasus GARMA, datajumlahan dimodelkan dengan menggunakan sebuah fungsi link *logarithmic* atau *identity* [2]. Jika y_t mengikuti distribusi Poisson dengan nilai *mean* sebagai berikut:

$$\mu_t = E(y_t|y_{t-1}) = \mu[1 + \exp(-\beta_0 - \beta_1 y_{t-1})], \beta_i > 0$$

Maka ada dua metode transformasi sebagai berikut yang dapat digunakan:

a. Transformasi ZQ1 yang mempunyai bentuk:

$$\log(\mu_t) = \boldsymbol{X}_t^T \boldsymbol{\beta} + \sum_{i=1}^{q} \theta_i [\log(y_{t-1}^T) - \boldsymbol{X}_{t-i}^T \boldsymbol{\beta}]$$

dimana $y_t^T = \max(y_t, c), 0 < c \le 1$.

b. Transformasi ZQ2 yang mempunyai bentuk:

$$\log(\mu_t) = \boldsymbol{X}_t^T \boldsymbol{\beta} + \sum_{i=1}^{q} \theta_i [\log(y_{t-1} + c) - \log[\exp(\boldsymbol{X}_{t-i}^T \boldsymbol{\beta} + \boldsymbol{c})]]$$

Sehingga model GSARIMA binomial negatif untuk ZQ1 diberikan persamaan 3:

$$\log(\mu_t) = \phi_p(B)(1-B)^d(1 - B^S)^D \Phi_P(B^S)\{\boldsymbol{X}_t^T \boldsymbol{\beta} - \log(y_t^T)\} + \log(y_t^T) - \theta_q(B)\Theta(B^S) \log\left(\frac{y_t^T}{\mu_t}\right) + \log\left(\frac{y_t^T}{\mu_t}\right)$$

Sedangkan untuk ZQ2 adalah

$$\log(\mu_t) = \phi_p(B)(1-B)^d(1-B^S)^D\Phi_P(B^S)\{\log[\exp(\boldsymbol{X}_t^T\boldsymbol{\beta})+c] - \log(y_t+c)\}+\log(y_t+c) - \theta_q(B)\Theta(B^S)\log\left(\frac{y_t+c}{\mu_t+c}\right) + \log\left(\frac{y_t+c}{\mu_t+c}\right)$$

LAYOUT DAN SPESIFIKASI

Simulasi Perbandingan Transformasi Fungsi Link ZQ1 dan ZQ2

Model data yang dibangkitkan adalah model Poisson AR (1) dengan menggunakan dua transformasi ZQ1 dan ZQ2. Data series yang dibangkitkan sebesar N=1000 dengan $\exp(\beta_0)$sebesar 5. Sedangkan nilai *c* ada tiga yaitu 0,1;0,5; dan 1 serta nilai parameter AR ϕ_1 -0,5 dan 0,5. Adapun hasil perbandingan antara fungsi link ZQ1 dan ZQ2 dapat dilihat pada tabel 1 di bawah:

Tabel 1 Simulasi Perbandingan Estimasi Fungsi Link ZQ1 dan ZQ2 dengan Menggunakan Model Poisson AR (1) pada $\exp(\beta_0) = 100$, $c = 0,1$; $c = 1,0$;$\phi_1 = 0,5$

Model	***c***	***intercept***	[illegible]	***Maref***	**DIC**
Log-ZQ1	0,1	101,37	0,48	0,079	7431,5
Log-ZQ1	1,0	100,45	0,47	0,081	7491,7
Log-ZQ2	0,1	99,93	0,49	0,081	7453,4
Log-ZQ2	1,0	101,04	0,49	0,080	7463,5

Berdasarkan Tabel 1 dapat dilihat bahwa ketika nilai $\exp(\beta_0)$ 100 dengan nilai c=0,1 maka transformasi fungsi link ZQ1 memiliki nilai *maref* dan DIC yang relatif kecil dibandingkan dengan transformasi fungsi link ZQ2. Tetapi sebaliknya ketika nilai *c* lebih besar (c=1,0) maka nilai transformasi dari ZQ2 memiliki nilai *maref* dan DIC yang relatif kecil. Selain itu jika nilai c lebih besar maka dari kedua fungsi link tersebut relatif menaikkan nilai DIC.

Sehingga diperoleh kesimpulan bahwa untuk pada saat*intercept* yang kecil maka fungsi link ZQ1 relatif lebih baik dibandingkan dengan ZQ2.

Simulasi Perbandingan Estimasi Parameter Model GSARIMA

Pada bagian ini akan dilakukan simulasi estimasi parameter pada model GSARIMA yang mengikuti distribusi binomial negatif. Model GSARIMA yang dibangkitkan terdiri atas dua model yaitu model GSARIMA $(2,1,0)(0,0,1)^{12}$ dengan GSARIMA $(0,1,2)(1,0,0)^{12}$.

Tabel 2. Simulasi Perbandingan Tingkat Akurasi Peramalan GSARIMA dengan SARIMA Model Binomial Negatif ARIMA $(2,1,0)(0,0,1)^{12}$

Model	*Maref*		Keterangan*)
	GSARIMA	SARIMA	
$(2,1,0)(0,0,1)_{12}$	0,525	4,623	11%
	0,493	198,241	9 %
	0,542	15,154	13 %
	0,449	74,44	8 %

Keterangan : *) jumlah observasi nol

Dari hasil simulasi estimasi data bangkitan dengan menggunakan distribusi Poisson AR (1) maka diperoleh hasil pada Tabel 4.5. PadaTabel 4.5 dapat dilihat bahwa dari keempat simulasi data yang dibangkitkan ternyata tingkat akurasi peramalan model binomial negatif GSARIMA $(2,1,0)(0,0,1)^{12}$ lebih baik dibandingkan dengan model ARIMA $(2,1,0)(0,0,1)^{12}$. Rata-rata nilai *maref* untuk keempat simulasi data untuk model GSARIMA adalah sebesar 0,502 sedangkan rata-rata nilai *maref* untuk model SARIMA sebesar 73,11.

4. DAFTAR PUSTAKA

[1] Benjamin, M. A., Rigby R. A., dan Stasinopoulos, D. M. (1998) *Fitting Non-Gaussian Time Series Models.COMPSTAT Proceedings in Computationel Statistics, eds. R. Payne dan P.Green, Heldelburg: Physica-Verlag*, 191-196.

[2] Benjamin, M. A., Rigby R. A., dan Stasinopoulos, D. M. (2003). *Generalized Autoregressive Moving Average Models.Journal of the American Statistical Association*, 98, 214-224.

[3] Bowerman, B. L., dan O’Connell, R. T. (2003). *Forecasting and Time Series: An Applied Approach, 3th eds*. New Jersey: Prentice Hall.

[4] Briet, J. T. O., (2009). *Toward Malaria Prediction in Sri Lanka: Modelling Spatial and Temporal Variability of Malaria Case Counts*. Disertasi Doktor. Netherland: Universitas Basel.

[5] Briet, J. T. O., Amerasinghe, H. P., dan Vounatsou, P. (2013). *Generalized Seasonal Autoregressive Integrated Moving Average Models for Count Data with Application to Malaria Time Series with Low Case Numbers*. *Malaria Journal*, PLoS ONE, 8(6): e65761.

[6] Croston, J. D. (1972). *Forecasting and Stock Control for Intermittent Demands*. *Operational Research Quarterly*, 23, 289– 303.

[7] Cryer, J. D. (1986). *Time Series Analysis*. Boston: PWS-KENT Publishing Company.

[8] Enders, W. (2004). *Applied Econometric Time Series*, *2nd eds*. John Wiley & Sons, Inc., New York.

[9] Garbhi, M., Quenel, P., dan Marrama, L. (2011). *Time Series Analysis of Dengue Incidence in Guadeloupe, French West Indies: Forecasting Models Using Climate Variables as Predictors. BMC Infect Disease*, 11, 166.

[10] Gooijer, G. J., Hyndman, J. R. (2006). *25 Years of Time Series Forecasting*. *International Journal of Forecasting*, 22, 443-473.