ISSN 2229-5518

GSTAR and GSTAR-X Models to Forecast Inflation in Four Cities in North Sumatera

Henny Agustina Siagian, Anik Djuraidah, I Made Sumertajaya

Abstract— Inflation presents the purchasing power of people. One important indicator of economic stability is inflation. Forecasting inflation rate is very important in taking policy for the government, enterprises or other economic actors. North Sumatra is one of the big provinces in Indonesia. In 2016, four cities in North Sumatra: Medan, Siantar, Sibolga and Sidempuan had the highest inflation rate in Indonesia. GSTAR model can forecast inflation in several locations simultaneously by including time and location factors in the model. Modeling the inflation rate in four cities in North Sumatra using GSTAR produced GSTAR (2:1)(1,0,0:1)6. The average values of AIC and RMSEP on GSTAR(2:1)(1,0,0:1)6 were -11.22 and 0.90. In its development, GSTAR model was not just influenced by time and location but there were other influencing variables, called exogenous variables. This model is called GSTAR-X. In this study, the addition of three exogenous variables on GSTAR-X(2:1) (1,0,0:1)6 produced the best model with an average AIC of -89.70 and an average RMSE of 0.57.

Index Terms— Exogenous variables, GSTAR, Inflation.

1 Introduction

INFLATION became one of the macroeconomic indicators that illustrate the economic stability of a country. The central government established an inflation control team in the regions called TPID (Regional Inflation Control Team) to maintain stability in the inflation rate. In its implementation, TPID found several problems that cause instability in the rate of inflation in the region. These issues include government policies on fuel oil prices and taxes, limited food production, limited transportation, limited distribution and inefficient market structure.

North Sumatra is one of the major provinces in Indonesia. North Sumatra Province has the largest population excluding Java, i.e about 13 million people. North Sumatra also has the third largest number of districts / cities after East Java and Central Java, i.e 33 districts / cities. North Sumatra has four inflation cities namely Medan, Siantar, Sibolga and Sidempuan. Cumulative inflation in North Sumatra in 2016 increased 6.34 percent compared to cumulative inflation in 2015 which only reached 3.24 percent. In 2016, Siboga City and Medan City became the cities with the highest and third highest cumulative inflation in Indonesia, respectively[1].

Some researchers have conducted a study using the topic of inflation rate as the subject of their studies. Ardianto (2013) used the General Space Time Autoregressive (GSTAR) model to forecast inflation in five major cities in Java[2]. Faizah (2013) used GSTAR model to forecast inflation of three cities in Central Java[3], and Husna (2014) used GSTAR model to forecast inflation in Jabodetabek [4]. In its development, the GSTAR model was not only influenced by observations in some locations over a period of time, but there were exogenous variables that influenced it, called GSTAR-X model. Hafsyah (2014) stated that inflation in North Sumatra was influenced by the exchange rate of Rupiah to US Dollar [5].

The weighted matrix has a very important role in GSTAR model. Husna (2014) studied inflation using GSTAR model [6], Astuti (2015) examined forecasting CPO exports on Sumatra Island using GSTAR model [7] and Siregar (2015) examined forecasting sugar prices on Sumatra Island using GSTAR model [8]. The three researchers concluded that prediction value using weighted inverse distance and cross-corelation resulted the best predicted value.

This study examined inflation data in four cities in North Sumatra using GSTAR and GSTAR-X models. The weighted matrices used were inverse distance and cross correlation matrices. The three exogenous variables used in GSTAR-X model are two dummy variables and one numerical variable. The selection of two dummy variables in the form of calendar variations was expected to capture the information behind the incidence of rising or falling inflation rate suddenly that cannot be measured quantitatively. Meanwhile, the selection of numerical variable, the exchange rate of Rupiah to US Dollar was expected to capture the state of the world's ongoing economy that can be measured quantitatively.

2 Literature Reviews

2.1 Autoregressive Moving Average (ARMA)

The general form of non-seasonal ARMA model can be written as follows:

$$\Phi_p(B)Z_t = \theta_q(B)\varepsilon_t \quad (1)$$

p is the order of AR, q is the order of MA, and B is the back shift operator, ε_t is white noise residual which is independent.

- Henny Agustina Siagian is currently pursuing masters degree program in applied statistics in Bogor Agricultural University, Indonesia, PH +628536130218. E-mail: siagianhenny08@gmail.com
- Anik Djuraidah is Lecturer, Departement of Statistics, Bogor Agricultural University, Bogor, Indonesia. E-mail: anikdjuraidah@gmail.com
- I Made Sumartejaya is Lecturer, Departement of Statistics, Bogor Agricultural University, Bogor, Indonesia. E-mail: imsjaya@gmail.com

In seasonal ARMA (SARMA) model with s is the length of the seasonal period can generally be written as follows:

$$\Phi^*_P(B^s)Z_t = \Theta^*_Q(B^s)\varepsilon_t \tag{2}$$

P is the order of the seasonal AR, Q is the seasonal order of MA, and B is the back shift operator, ε_t is white noise residual which is independent.[9]

Multiplicative models for non-seasonal ARMA and seasonal ARMA :

$$\Phi_p(B)\Phi^*_P(B^s)Z_t = \theta_q(B)\theta^*_Q(B^s)\varepsilon_t \tag{3}$$

2.2 General Space Time Autoregressive Moving Average (GSTARMA)

The GSTARMA model (p; λ_k; q; ν_k) with n location can be written as follows:

$$Z_t = \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi^k_{kl}W^l Z_{t-k} - \sum_{k=1}^{q}\sum_{l=0}^{\nu_k}\theta^k_{kl}W^l\varepsilon_{t-k} + \varepsilon_t \tag{4}$$

p is the order of AR, q is the order of MA, λ_k is the spatial order of k-th AR, and ν_k is the spatial order of k-th MA.

2.3 General Space Time Autoregressive Moving Average using Exogenous Variables (GSTARMA-X)

The GSTARMA-X model (p; λ_k; q; ν_k) with n location can be written as follows:

$$Z_t = \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi^k_{kl}W^l Z_{t-k} + \sum_{k=1}^{q}\sum_{l=0}^{\nu_k}\theta^k_{kl}W^l\varepsilon_{t-k} + \alpha X(t) + \varepsilon_t \tag{5}$$

p is the order of AR, q is the order of MA, λ_k is the spatial order of k-th AR, ν_k is the spatial order of k-th MA, and α is parameter of exogenous variable.

2.4 Weighted Distance Matrices

2.4.1 Inverse Distance Weighted Matrix

The inverse distance weighted matrix is one of the weighted matrices that refers to the distance between locations. The value of the inverse weight of the distance is obtained based on the calculation of latitude and longitude coordinates distances between the observed locations with the calculation of Euclidean distance. Nearby locations get greater weight values and vice versa.[10]

$$w_{ij} = \begin{cases} \dfrac{w_{ij}^*}{\sum_{j=1}^{N} w_{ij}^*}, & for\ i \neq j \\ 0\ , & for\ i = j \end{cases}$$

$$\text{with } w_{ij}^* = 1/d_{ij} \text{ and } d_{ij} = \left\{\left[x_i(u_i) - x_j(u_j)\right]^2 + \left[x_i(v_i) - x_j(v_j)\right]^2\right\}^{1/2}$$

(u_i, u_j) is the latitude coordinate and (v_i, v_j) is the longitude coordinate at location i and j.

2.4.2 Weighted Cross-Correlation Matrix

Cross-correlation is generally formulated by:

$$r_{ij}(k) = \frac{\sum_{i=k+1}^{n}(Z_i(t) - \bar{Z}_i)(Z_j(t) - \bar{Z}_j)}{\sqrt{\sum_{i=k+1}^{n}(Z_i(t) - \bar{Z}_i)^2 \sum_{i=k+1}^{n}(Z_j(t) - \bar{Z}_j)^2}}$$

The determination of elements in weighted cross-correlation matrix is by:

$$w_{ij} = \frac{|r_{ij}(k)|}{\sum_{j \neq i} |r_{ij}(k)|}$$

3 Methods

3.1 Data

The data used in this study are the inflation data in four cities and the Rupiah to US Dollar exchange rate data. Inflation data in four cities include Medan, Siantar, Sibolga and Sidempuan. The data used are monthly data from January 2008 to December 2016. Inflation data were obtained from publication of Central Agency on Statistics of North Sumatera Province and Rupiah to US Dollar exchange rate data were obtained from publication of Ministry of Trade of The Republic of Indonesia.

ISSN 2229-5518

3.2 Method of Analysis

The steps of the analysis data of this research are:

a. Conducting data exploration
b. Dividing data into two parts: training and testing data. The first 102 observations were training data used for the modeling stage and the last 6 observations were testing data used to test the model.
c. Calculating the correlation coefficient between locations
d. Development and estimation of GSTAR model parameters
   - Checking the stationarity of inflation data at each location
   - Identifying the ARMA models in each location
   - Constructing weighted matrices by using weighted cross-correlation and inverse distance weighted matrices
g. Estimating GSTAR-X parameters
h. Examining the assumptions of GSTAR and GSTAR-X residuals
i. Selecting the best model based on the smallest AIC and RMSEP values
j. Forecasting the next 6 months inflation rate based on the best model

4 Result and Discussion

4.1 Data Description

Descriptive statistics for inflation in each city is shown in Table 1. In general, the largest inflation rate of 1.35 was located in Sibolga City, while the smallest inflation rate of 0.63 was located in Medan City.

Table 1 Descriptive Statistics

Variable	Minimum	Maximum	Mean	Variance
Medan	-1.36	2.74	0.53	0.63
Siantar	-1.47	2.88	0.56	0.80
Sibolga	-2.43	3.78	0.60	1.35
Sidempuan	-1.43	3.70	0.45	0.78

4.2 Correlation Coefficient between Locations

Table 2 shows the correlation matrix in four observation locations. All correlation coefficients were positive and greater than 0.7. A correlation value greater than 0.7 indicates a relatively strong relationship between locations.

Table 2 Correlation Coefficient in Four Cities

Variable	Medan	Siantar	Sibolga	Sidempuan
Medan	1			
Siantar	0.792	1		
Sibolga	0.758	0.755	1	
Sidempuan	0.714	0.709	0.712	1

4.3 Development and estimation of GSTAR model parameters

4.3.1 Identification of Data Stationerity

The requirement to model the time series data is that the data used must be stationary. Table 3 shows the ADF test results at four locations. The results of the ADF test showed a significant value in all locations. Based on the results of the ADF test, the data had met the assumptions of stationarity.

Table 3 The Results of ADF Test

Variable	Rho	p-Value
Medan	-63.8134	<.0001
Siantar	-63.6252	<.0001
Sibolga	-68.5101	<.0001
Sidempuan	-81.0128	<.0001

4.3.2 Identification of ARMA Model

The ARMA model identification at each location determined the time order in GSTAR model. The ARMA model at each location was examined based on ACF and PACF plots followed by over fitting some model candidates. The best model criterion was a model that had significant parameter estimation, minimum AIC value, white noise residual and had the smallest residual mean square of error.

Based on those criterias, AR (2) (1,0,0) 6 was the best model for Medan City. AR (1) (1,0,0) 6 became the best model for Siantar City. Meanwhile, the best ARMA model for Sibolga City was not identified. The best model for Sidempuan City was AR (2).

4.3.3 Construction of Weighted Matrices

The weighted cross-correlation matrix was determined by the correlation value of the corresponding lag between locations, meanwhile the inverse distance weighted matrix was determined by the distance between locations. Table 3 is a weighted cross-correlated matrix of corresponding lags.

Table 3 Weighted Cross-Correlation Matrix

Lag	Matrix	lag	Matrix
1	$W = \begin{bmatrix} 0 & 0.368 & 0.313 & 0.320 \\ 0.366 & 0 & 0.296 & 0.338 \\ 0.370 & 0.382 & 0 & 0.247 \\ 0.336 & 0.366 & 0.297 & 0 \end{bmatrix}$	7	$W = \begin{bmatrix} 0 & 0.034 & 0.450 & 0.515 \\ 0.199 & 0 & 0.380 & 0.755 \\ 0.132 & 0.639 & 0 & 0.228 \\ 0.105 & 0.594 & 0.301 & 0 \end{bmatrix}$
2	$W = \begin{bmatrix} 0 & 0.360 & 0.313 & 0.327 \\ 0.330 & 0 & 0.380 & 0.289 \\ 0.337 & 0.361 & 0 & 0.302 \\ 0.337 & 0.256 & 0.407 & 0 \end{bmatrix}$	8	$W = \begin{bmatrix} 0 & 0.507 & 0.363 & 0.130 \\ 0.542 & 0 & 0.279 & 0.178 \\ 0.482 & 0.324 & 0 & 0.194 \\ 0.419 & 0.276 & 0.304 & 0 \end{bmatrix}$
3	$W = \begin{bmatrix} 0 & 0.357 & 0.403 & 0.240 \\ 0.395 & 0 & 0.342 & 0.263 \\ 0.370 & 0.333 & 0 & 0.297 \\ 0.300 & 0.311 & 0.388 & 0 \end{bmatrix}$		

Based on the calculation of Euclidean distance, the inverse distance weighted matrix was constructed:

$$W = \begin{bmatrix} 0 & 0.58 & 0.23 & 0.19 \\ 0.47 & 0 & 0.3 & 0.23 \\ 0.17 & 0.27 & 0 & 0.56 \\ 0.15 & 0.22 & 0.63 & 0 \end{bmatrix}$$

4.3.4 Estimation of GSTAR Parameters

GSTARMA model was constructed based on the maximum time order on ARMA model. The maximum time order was obtained in AR model (2)(1,0,0)6. Meanwhile, spatial order for the GSTARMA model was restricted to one order. The form of GSTARMA model was GSTAR (2:1)(1,0,0:1)6. The parameter estimation value if translated into GSTAR (2:1)(1,0,0:1)6 using inverse distance weighted matrix for Medan City was as follows:

$$\begin{aligned} Z_1(t) = {} & 0.273Z_1(t-1) + 0.053\,Z_2(t-1) + 0.053\,Z_3(t-1) + 0.053\,Z_4(t-1) \\ & + 0.027Z_1(t-2) - 0.083\,Z_2(t-2) - 0.033\,Z_3(t-2) - 0.027\,Z_4(t-2) \\ & + 0.199Z_1(t-6) + 0.112\,Z_2(t-6) + 0.045\,Z_3(t-6) + 0.037\,Z_4(t-6) \\ & - 0.406Z_1(t-7) + 0.141\,Z_2(t-7) + 0.056\,Z_3(t-7) + 0.046\,Z_4(t-7) \\ & - 0.178Z_1(t-8) + 0.154\,Z_2(t-8) + 0.061\,Z_3(t-8) + 0.050\,Z_4(t-8) \end{aligned}$$

Table 4 Parameter Estimations in Medan City

Variable	Parameter	Estimation[a]	Estimation[b]
Medan	ϕ^1_{10}	0.270	0.273
	ϕ^1_{11}	0.114	0.092
	ϕ^2_{10}	0.037	0.027
	ϕ^2_{11}	-0.174	-0.143
	ϕ^6_{10}	0.238	0.199
	ϕ^6_{11}	0.161	0.194
	ϕ^7_{10}	0.296	0.406
	ϕ^7_{11}	-0.114	-0.244
	ϕ^8_{10}	0.125	0.178
	ϕ^8_{11}	-0.234	-0.265
[a] cross-corelation matrix		[b] inverse distance matrix	

4.4 Estimation of GSTAR-X Parameters

In this study, GSTAR-X model used 3 exogenous variables i.e D_1, D_2 and X_3. D_1 were dummy variables which category 1 for the month of government policy to raise fuel prices, TDL, ELPIJI, etc., and 0 for others. D_2 were dummy variables which category 1 for the month of government policy to reduce fuel prices, ELPIJI, TDL, etc., and 0 for others. Meanwhile, X_3 was exchange rate of Rupiah to US Dollar. GSTAR-X model produced 14 models, one of which was GSTAR-$D_1D_2X_3$ (2:1)(1,0,0:1)6. The very small difference in parameter estimation between two weighted matrices using GSTAR-$D_1D_2X_3$ (2:1)(1,0,0:1)6 showed the parameter estimation values between the weighted cross-correlation matrix which was almost equal to the inverse distance weighted matrix.

Table 5 Parameter Estimations in Medan City of GSTAR-$D_1D_2X_3$ (2:1)(1,0,0:1)6

Variable	Parameter	Estimation[a]	Estimation[b]	Difference
Medan	ϕ^1_{10}	-0.06585	-0.05091	0.01494
	ϕ^1_{11}	0.202348	0.178999	0.02335
	ϕ^2_{10}	-0.14455	-0.13916	0.00539
	ϕ^2_{11}	0.003459	0.008505	0.00505
	ϕ^6_{10}	-0.0407	-0.05734	0.01664
	ϕ^6_{11}	0.073553	0.082596	0.00904
	ϕ^7_{10}	0.242683	0.315112	0.07243
	ϕ^7_{11}	-0.08937	-0.17576	0.08639
	ϕ^8_{10}	0.139569	0.15809	0.01852
	ϕ^8_{11}	-0.03034	-0.03758	0.00724
	α_1	1.102128	1.106054	0.00393
	β_1	-0.99029	-0.96715	0.02314
	γ_1	0.000042	0.000041	0.00000
[a] cross-corelation matrix			[b] inverse distance matrix	

4.5 Diagnostic of Residual

The residual assumption was needed to declare GSTAR and GSTAR-X models constructed was appropriate models. Table 6 shows that GSTAR (2:1)(1,0,0:1)6 had a minimum AIC residual located at AR (0) and MA (0) in both weighted matrices. This means that the residual assumption of GSTAR (2:1)(1,0,0:1)6 had been fulfilled.

Table 6 AIC Residual on GSTAR (2:1)(1,0,0:1)6

AIC[a]					AIC[b]				
Lag	MA 0	MA 1	MA 2	MA 3	Lag	**MA 0**	MA 1	MA 2	MA 3
AR 0	**-4.11**	-3.91	-3.62	-3.41	AR 0	**-5.86**	-5.67	-5.38	-5.25
AR 1	-3.90	-3.69	-3.36	-3.15	AR 1	-5.69	-5.47	-5.21	-5.03
AR 2	-3.57	-3.38	-3.16	-2.86	AR 2	-5.37	-5.18	-4.99	-4.77
AR 3	-3.30	-3.11	-2.86	-2.47	AR 3	-5.13	-5.03	-4.72	-4.34
[a] cross-corelation matrix					[b] inverse distance matrix				

4.6 Selection of The Best Models

The best models were measured by the smallest AIC and RMSEP values. This study obtained 16 models, therefore to facilitate the analysis, model naming efficiency was performed.

Table 7 Model Naming Efficiency using Average Values of AIC and RMSEP

Alphabet	Model	AIC Value	RMSEp Value	Alphabet	Model	AIC Value	RMSEp Value
a	GSTAR*	-11.71	0.90	b	GSTAR**	-11.22	0.90
c	GSTAR-D_1*	-68.25	0.66	d	GSTAR-D_1**	-69.43	0.66
e	GSTAR-D_2*	-20.76	0.85	f	GSTAR-D_2**	-19.45	0.86
g	GSTAR-X_3*	-17.62	0.87	h	GSTAR-X_3**	-17.21	0.87
i	GSTAR-D_1D_2	-81.75	0.61	j	GSTAR-D_1D_2**	-81.86	0.61
k	GSTAR-D_1X_3*	-69.24	0.66	l	GSTAR-D_1X_3**	-70.41	0.66
m	GSTAR-D_2X_3*	-35.72	0.78	n	GSTAR-D_2X_3**	-34.06	0.79
o	GSTAR-D_1D_2X3*	-89.50	0.59	p	GSTAR-D_1D_2X3**	**-89.70**	**0.57**
* cross-corelation matrix				** inverse distance matrix			

ISSN 2229-5518

Figure 1 is the average value of AIC interpreted in graphical form for each weighted matrix. The largest average value of AIC was in the GSTAR (2:1)(1,0,0:1)6. The average value of AIC in GSTAR (2:1)(1,0,0:1)6 using cross-correlation matrix was -11.71 and using inverse distance matrix was -11.22. Meanwhile the best model based on the smallest average value of AIC was GSTAR-$D_1D_2X_3$ (2:1)(1,0,0:1)6. The average value of AIC in GSTAR-$D_1D_2X_3$ (2:1)(1,0,0:1)6 using cross-correlation matrix was -89.50 and using inverse distance matrix was -89.70.

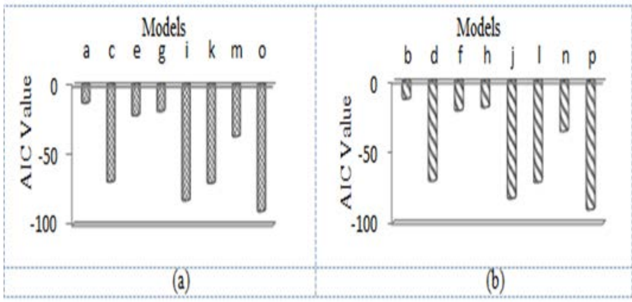

Figure 1 The Average Value of AIC (a) Cross-Correlation and (b) Inverse Distance

Figure 2 shows the largest average value of RMSEP was in GSTAR (2:1)(1,0,0:1)6. The smallest average value of RMSEP was GSTAR-$D_1D_2X_3$ (2:1)(1,0,0:1)6. Based on the smallest average values of AIC and RMSEP, GSTAR-$D_1D_2X_3$ (2:1)(1,0,0:1)6 using inverse distance weighted matrix was the best model of the 16 models constructed.

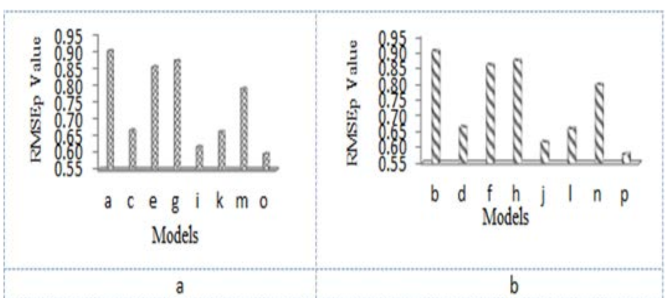

Figure 2 The Average Value of RMSEP (a) Cross-Correlation and (b) Inverse Distance

4.7 Forecasting

Forecasting inflation rate in four cities with the best model was performed for the next 6 months from January 2017 to June 2017. The inflation rate graph for each location shows the pattern of forecast data following the actual data pattern.

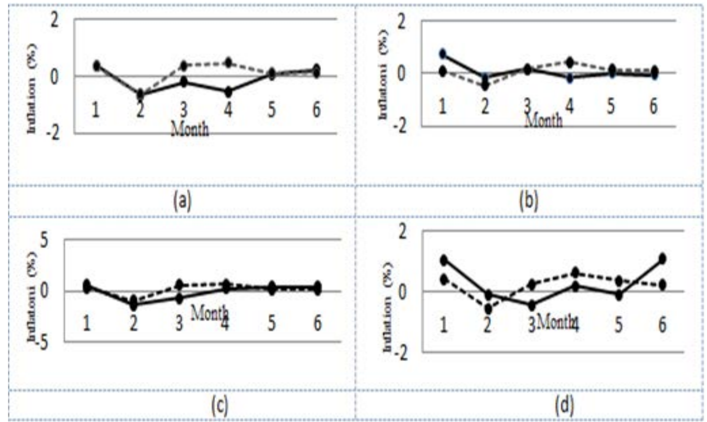

Figure 4 The Graph of Actual Data (----) and Forecast Data (______) Based on The Best Model (a) Medan, (b) Siantar, (c) Sibolga and (d) Sidempuan

5 Conclusion

The addition of exogenous variables on the GSTAR model was able to decrease the average values of AIC from -11.22 to -89.70 and RMSEP from 0.90 to 0.57. GSTAR-$D_1D_2X_3$ (2:1)(1,0,0:1)6 using the inverse distance weighted matrix was the best model.

References

[1] [BPS] Badan Pusat Statistik.2016. Statistik Indonesia 2016, Jakarta(ID):Badan Pusat Statistik.
[2] Ardianto MP. 2014. Pemodelan GSTAR Pada Tiga Periode Waktu (Studi Kasus Inflasi di Lima Kota Besar di Pulau Jawa). Journal Mahasiswa Statistik. Vol 2
[3] Faizah LA. 2013. Pemodelan Inflasi di Kota Semarang, Yogyakarta, dan Surakarta Dengan Pemodelan GSTAR. Journal Sains dan Seni. Surabaya (ID): Institut Teknologi Sepuluh Nopember, Vol. 2.
[4] Husna F. 2014. Pemodelan GSTAR dengan Menggunakan Empat Jenis Pembobot Lokasi. Journal Mahasiswa Statistik. Malang(ID): Universitas Brawijaya.
[5] Hafsyah A. 2014. Analisis Inflasi Sumatera Utara dengan Model Sisaan Correction (ECM). Journal QE, Vol. 01.
[6] Husna F. 2014. Pemodelan GSTAR dengan Menggunakan Empat Jenis Pembobot Lokasi. Journal Mahasiswa Statistik. Malang(ID): Universitas Brawijaya.
[7] Astuti Dewi. 2015. Penerapan Model GSTAR-X untuk Peramalan Volume Ekspor CPO [Tesis]. Bandung(ID): Universitas Padjajaran
[8] Siregar D. 2015. Perbandingan Pemodelan dan Peramalan Harga Gula Berdasarkan Model STARIMA dan GSTARIMA [Tesis]. Bogor (ID): Institut Pertanian Bogor..
[9] Montgomery DC, Jennings CL, Kulahci M. 2008. Introduction to Time Series Analysis and Forecasting. Canada (CA): Wiley Interscience.
[10] Suhartono, Atok RM. 2006. Pemilihan Bobot Lokasi yang Optimal pada Model GSTAR, Presented at National Mathematic Conference XIII. Semarang (ID): Universitas Negeri Semarang.