IJCSN - International Journal of Computer Science and Network, Volume 6, Issue 6, December 2017
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Forecasting Simulation with ARIMA and Combination of Stevenson-Porter-Cheng Fuzzy Time Series

[1] Wahyu Dwi Sugianto; [2] Agus Mohamad Soleh; [3] Farit Mochamad Afendi

[1] Department of Statistics, Bogor Agricultural University, IPB
Bogor, 16680, Indonesia

[2] Department of Statistics, Bogor Agricultural University, IPB
Bogor, 16680, Indonesia

[3] Department of Statistics, Bogor Agricultural University, IPB
Bogor, 16680, Indonesia

Abstract - The simulation was implemented to find out the perfomance for a combination of methods in Stevenson-Porter-Cheng Fuzzy Time Series (FTS) based on 100 replicates on 100 generated data following the model of ARIMA (1,0,0) or AR (1). There are 9 scenarios used as a combination between 3 data generation error variance values (0.5, 1, 3) and 3 AR(1) parameter values i.e. 0.3, 0.5, and 0.7. The results of the simulation showed the greater variance of error and the value of the of AR(1) parameter then the variance of the MSE results with ARIMA will be even greater (0.0634 – 15.7633). While the variance of the MSE results forecasting with Cheng and Cheng2 (no sub interval) FTS tend to be more stable (0.0712 – 2.9648 and 0.0640 – 2.7157). By using the percentage change of historical data as the set of universe, SP Cheng FTS produces MSE forecasting range values ranging from 0.0722 – 14.7045. While SP Cheng2 FTS using the difference of historical data resulted in MSE forecasting values ranging from 0.0759 – 4.6803. Although both MSE values do not look much better than Cheng and Cheng2 FTS, but the greater the AR(1) parameter then MSE forecasting of Cheng and Cheng2 FTS will be better than ARIMA and even closer to the Cheng and Cheng2 FTS.

Keywords - *ARIMA, Cheng Fuzzy Time Series, Simulation, Stevenson-Porter.*

1. Introduction

Autoregressive Integrated Moving Averages (ARIMA) is still a known method used in forecasting for a univariate time series. ARIMA has the assumption that must be met in the determination of their best model used for forecasting. Another approach to forecasting time series introduced by Song and Chissom [5] i.e. fuzzy time series (FTS). It's forecast system is to capture the patterns of past data as a basis for the future projection. Application of FTS by using historical data as the percentage change the set universe done by Stevenson and Porter [6] to predict the number of students at the University of Alabama. Dan et al. [2] using fuzzy local trend transform for forecasting with FTS.

In another research, K-means clustering applied by Zhiqiang and Qiong [8] as the basis for intervals formation of FTS. Hasbiollah and Hakim [3] implemented Stevenson-Porter FTS with modifications on the interval formation.

Other FTS method can be used in forecasting was a Cheng method [1]. Adaptive forecasting on the Cheng method generates smaller error size than ARIMA methods on composite stock price index by Tauryawati and Irawan [7].

Based on the previous explaination, this research will conducted a forecasting simulation on an ARIMA generated data i.e. AR (1) in order to know the performance of the combination Stevenson-Porter FTS and Cheng algorithm model. Cheng FTS applied sub-interval (Cheng FTS) and no sub interval (Cheng2 FTS). As for the method of Stevenson-Porter FTS combined with Cheng algorithm without sub-interval which adopt the concept of percentage changes to historical data (SPC FTS) and methods of Stevenson-Porter FTS combined with Cheng algorithm without sub-interval which use the difference concept of historical data (SPC2 FTS) as universe of discourse.

This simulation study is a part of inflation forecasting research in Sumatera with fuzzy time series approach.

2. Literature Reviews

2.1 Simulation of Data Generation

In this simulation, data is generated following AR(1) model with 9 different scenarios. Three combination of AR(1) parameter is 0.3, 0.5, and 0.7. While variance of error generation used 3 value i.e. 0.5, 1, and 3. The desired end result is 100 data series with 100 replication.

2.2 ARIMA

ARIMA has the assumption that must be met prior to a data modeling i.e. the stationary assumption. In general the stages of ARIMA model formation as follows:

a. Identification, starting with an examination of the stationary assumption with time series plot, *Autocorrelation Funcion* (ACF), *Partial Autocorrelation Funcion* (PACF), and *augmented* Dickey-Fuller (ADF) test or Box-Cox transformation [4]. Next is determined tentative model after the data is stationary. The model used in this simulation is an AR (1) stated as follows:

$$Z_t - \mu = \phi_1(Z_{t-1} - \mu) + e_t \quad (1)$$

with
μ : constant
ϕ_1 : autoregressive of 1st order parameter
e_t : error in t

b. Parameter estimation, started with parameter significance test with:

H_0 : $\phi = 0$
H_1 : $\phi \neq 0$
Statistical test:

$$t = \frac{\hat{\phi}}{Se(\hat{\phi})} \quad (2)$$

Rejection region: reject H_0 if $|t| > t_{\alpha/2,df}$
with df is degrees of freedom (subtract from data length and the sum of it's parameter)

c. Model diagnostics, checks the residual assumption using Ljung-Box test. The hypotheses being tested is:
H_0 : uncorrelated residuals
H_1 : correlated residuals
Q statistics :

$$Q_{LB} = T(T+2)\sum_{k=1}^{K}\left(\frac{r_k^2}{T-k}\right) \quad (3)$$

with:
T : length training data set
r_k : residual correlation at k-lag
K : number of lags being tested
Rejection of H_0 is when the p-value Q_{LB} is smaller than $\alpha = 5\%$.

Normality test using Kolmogorov-Smirnov. The hypotesis as follows:
H_0:$F(z)=F_0(z)$ (residual normally distributed)
H_1:$F(z)\neq F_0(z)$ (residual doesn't normally dist)
Statistics test:

$$D = Sup|S(z)\text{-}F_0(z)| \quad (4)$$

with:
$S(z)$: cummulative probability function, calculated from data
$F_0(z)$: cummulative probability function of normal distribution
D : supremum value of all z from $|S(z)\text{-}F_0(z)|$
Rejection region: reject H_0 when $D > D_{\alpha,n}$ with n data length and $D_{\alpha,n}$ is from Kolmogorov-Smirnov D value table.

2.3 Cheng Fuzzy Time Series (C FTS)

In general the stages used in FTS Cheng [7] can be described as follows:

a. Define universe of discourse, then the partition data into 7 same intervals. If there is a number of frequencies in particular interval greater than the average value of the frequency of each interval then the interval need tobe re-partitioned into two. While the Cheng FTS without sub-interval (C2 FTS) does not re-partition and next stages just adapts to the number of intervals/sub-interval.

b. Defines the fuzzy set in the universe and continue with do fuzzy classification on historical data. Suppose $A_1, A_2, \ldots A_k$ is the fuzzy set that has linguistic value from a linguistic variables. The definition of fuzzy sets *fuzzy* $A_1, A_2, \ldots A_k$, as follows:

$$A_1 = a_{11}/u_1 + a_{12}/u_2 + \ldots + a_{1m}/u_n$$
$$A_2 = a_{21}/u_1 + a_{22}/u_2 + \ldots + a_{2m}/u_n$$
$$\ldots$$
$$A_k = a_{k1}/u_1 + a_{k2}/u_2 + \ldots + a_{km}/u_n \quad (5)$$

with a_{ij} range of $[0,1]$, $1 \leq i \leq k$ and $1 \leq j \leq m$.
a_{ij} are the degrees of u_j membership in A_i fuzzy set.

c. Determining fuzzy logic relationship (FLR) based

IJCSN - International Journal of Computer Science and Network, Volume 6, Issue 6, December 2017
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

in historical data. In the data that had been fuzzified, two sequenced fuzzy set $A_i(t-1)$ and $A_j(t)$ can be stated as FLR $A_i \rightarrow A_j$.

d. Determining weights for FLR group. As an example for a similar sequence FLR,
(t=1) $A_1 \rightarrow A_1$, weighted by 1
(t=2) $A_2 \rightarrow A_1$, weighted by 1
(t=3) $A_1 \rightarrow A_1$, weighted by 2
(t=4) $A_1 \rightarrow A_1$, weighted by 3
with t refering to time.
Next, transfering weights to the normalized weighted matrix W_n(t) with equation as follows:

$$W_n(t) = [W'_1, W'_2, \ldots, W'_k]$$

$$\left[\frac{w_1}{\sum_{h=1}^{i} w_h}, \frac{w_2}{\sum_{h=1}^{i} w_h}, \ldots, \frac{w_k}{\sum_{h=1}^{i} w_h}\right] \quad (6)$$

e. Forecasting was made by multiplying normalized weighted matrix W_n(t) with fuzzied matrix (L_{df}). Matrix $L_{df} = [m_1, m_2, \ldots, m_k]$ with m_k is the mean from every intervals.

$$F(t) = [L_{df(t-1)} * W_n(t-1)] \quad (7)$$

f. Modifying forecast with adaptive forecast as follows:

$$\text{Adaptive forecast } (t) = Z_{t-1} + h * (F_t - Z_{t-1}) \quad (8)$$

with Z_{t-1} is the actual value period t-1, F_t is forecasted value, and h is weights parameter in range of 0,001-1.

2.4 Stevenson-Porter Fuzzy Time Series with Cheng Algorithm (SPC FTS)

Stevenson-Porter FTS has not been designed to forecast for t+1 period. Therefore this research will use the combination of percentage of change on historical data (d_t) as universe with Cheng algorithm. Also, for initial value the first data in traning data set is used. So, the first percentage of change (d_1) equals with zero.

$$d_t = (z_t - z_{t-1})*100\%/z_{t-1}, \quad t=2,2,3,\ldots,n$$

The next stage following the Cheng FTS stages until complete. After the adaptive forecasting, reverse transformation is required using the following formula:
$F'(t) = (F(t)*z_{t-1})/100 + z_{t-1}$

with F(t) is the adaptive forecasting for d_t.

From the idea of Stevenson-Porter FTS used percentage of change, this research used the difference of historical data as universe of discourse with Cheng algorithm (SPC2 FTS). The initial value used the first training data set. So, the first difference (d_1) equals zero.

$$d_t = z_t - z_{t-1}, \quad t=2,2,3,\ldots,n$$

The next stage following the Cheng FTS stages until complete. After the adaptive forecasting, reverse transformation is required using the following formula:

$$F'(t) = F(t) + z_{t-1}$$

3. Methods

3.1 Data

This simulation used a generated data from ARIMA i.e. AR(1) with 9 combination of AR(1) parameter and error variance from generated data.

3.2 Method of Analysis

General step of generating data from an AR(1) is as follows:

a. Generate $e \sim N(0,\sigma_e^2)$ with n=100, and variance of error specific to each scenario. The parameter μ=10 is used to generate an AR(1) with non-zero means.

b. Added the generated error to the AR(1) formula in order to create time series data with an AR(1) each scenario of ϕ_1 model parameter. This process is repeated for 100 replication.

c. Splitting 100 row of data as training data set (88 rows) and testing data set (last 12 rows).

4. Result and Discussion

4.1 Data exploration

First replicate plot of generated data with σ_e^2 = 0.5 and 3 AR(1) parameter showed in Fig 1, the generated data with parameter ϕ_1 = 0.7 shows a much longer range of fluctuation compared with parameter ϕ_1 = 0.5 or ϕ_1 = 0.3.

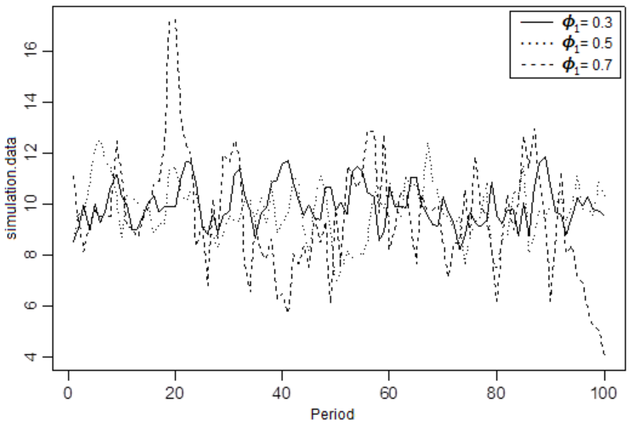


Fig. 1 First replicate plot of generated data with $\phi_1 = 0.5$ and 3 types of errors

First replicate plot of generated data with AR(1) $\phi_1 = 0.5$ parameter and 3 type variance of error as showed in Fig 2. Fig 2 showed the generated data with $\sigma_\varepsilon^2 = 3$ has a larger deviation compared to others. As for generated data from $\sigma_\varepsilon^2 = 0.5$ and $\sigma_\varepsilon^2 = 1$, despite some of the data showed a large deviation, in overall, the deviation is still smaller than deviation of $\sigma_\varepsilon^2 = 3$.

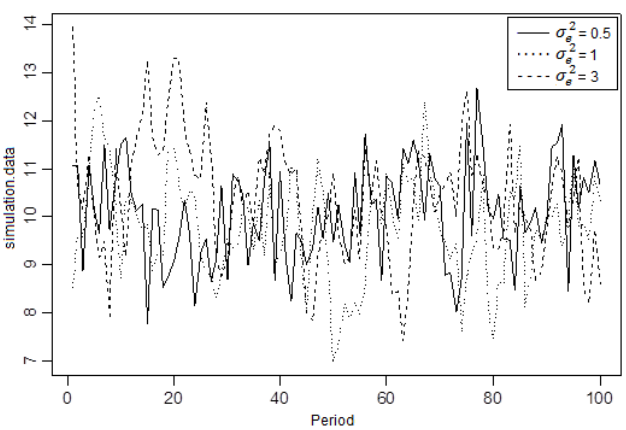


Fig. 2 First replicate plot of generated data with $\phi_1 = 0.5$ and 3 types of error variance

4.2 Comparison of MSE between ARIMA, Cheng FTS (C), Cheng2 FTS (C2), SP Cheng FTS (SPC), and SP Cheng2 FTS (SPC2 FTS)

MSE from simulation testing data set in Fig 3 for σ_ε^2= 0.5 and 3 types of AR(1) parameter (0.5, 1, and 3) showed the distribution center of MSE forecasting results with Cheng2 FTS method is a little bellow the center of MSE produced Cheng FTS method. Among the 4 FTS methods, Cheng2 FTS resulting a smaller average MSE. The average of MSE from Cheng2 FTS is in the range of 0.5568 – 0.5828 and variance range of 0.0712 – 0.0899.

The average of MSE from ARIMA is in the range of 0.5492 – 1.0152 with variance range of 0.0634 – 0.4516. On the data with σ_ε^2 = 0.5, the MSE variance of ARIMA increased as the AR(1) increased. While Cheng and Cheng2 FTS resulting a much stable MSE and not affected with the change of AR(1) parameter. Other results for SP Cheng2 FTS method, the center of MSE distribution is lower than SP Cheng FTS method with the average of MSE in the range of 0.5759 – 0.7800 with variance range of 0.0759 – 0.1532.

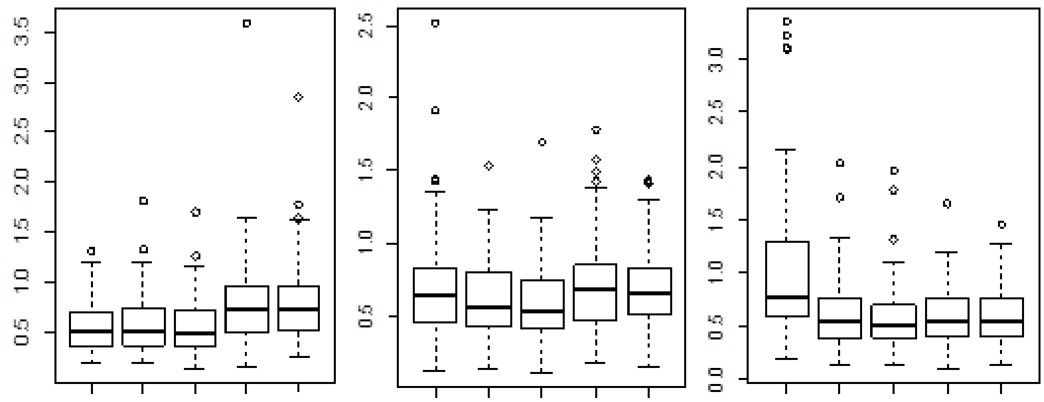
In Figure 4 for simulation with σ_ε^2= 1 and 3 AR(1) parameter, the center of MSE distribution from Cheng2 FTS is lower bellow the center of Cheng FTS. Among the 4 FTS methods, Cheng2 FTS still resulting the lowest average of MSE. The average MSE of Cheng2 FTS is in range of 1.0817 – 1.1109 with variance range of 0.2160 – 0.2871.

The average of MSE from ARIMA is in the range of 1.1131–1.8158 with variance range of 0.2041 – 1.5344. On the data with σ_ε^2 = 1, MSE variance of ARIMA forecasting result increased as the AR(1) parameter increased. While Cheng and Cheng2 FTS resulting a much stable MSE and not affected with the change of AR(1) parameter. The difference between two scenarios for σ_ε^2= 1 and ϕ_1= 0.5 as showed in Fig 6.b, the center of MSE distribution for Cheng2 FTS still lower than MSE Cheng FTS although with a larger variance.

For the last scenario, the simulation showed in Fig 5 with σ_ε^2= 3 and 3 AR(1) parameter 0.5, 1, and 3. The results showed that the distribution center of MSE forecasting results with Cheng2 FTS method still lower than distribution center of Cheng FTS for every AR(1) parameter used in data generation. The average MSE from Cheng2 FTS is the lowest compared with ARIMA and 3 others FTS methods. The average of MSE from Cheng2 FTS method is in range of 3.2636 – 3.4812 with variance of 2.2077 – 2.7157.

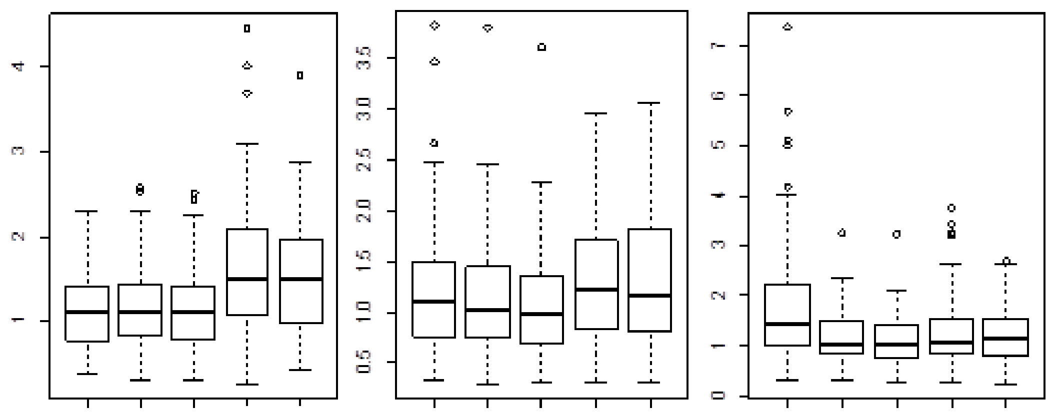
Meanwhile, the average of MSE from ARIMA is in the range of 3.4264 – 6.1310 and variance range of 2.2664 – 15.7633. On the data with σ_ε^2 = 3, MSE variance of ARIMA forecasting results increased as the AR(1) increased. While MSE from Cheng and Cheng2 FTS showed much stable and not affected with the change of AR(1) parameter.

IJCSN - International Journal of Computer Science and Network, Volume 6, Issue 6, December 2017
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5



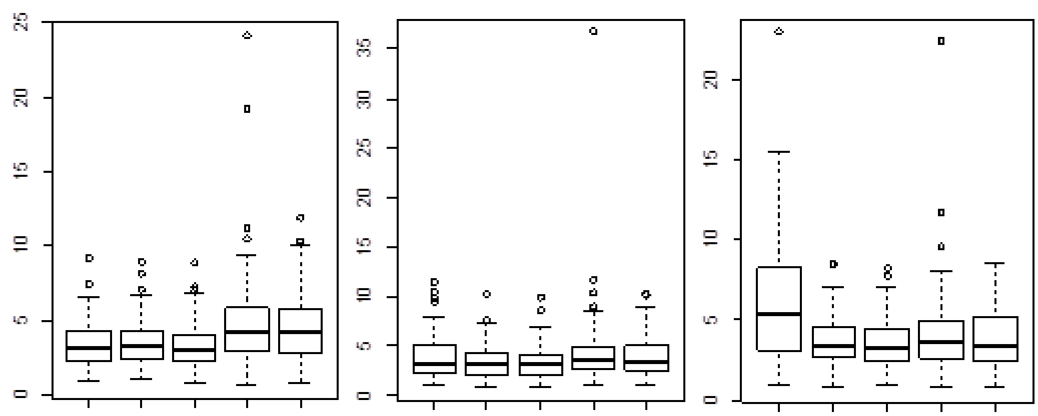
AR(1) C C2 SPC SPC2 AR(1) C C2 SPC SPC2 AR(1) C CST SPC SPC2
(a) (b) (c)

Fig 3. MSE of testing data set on simulation with $\sigma_\varepsilon^2 = 0.5$
and $\phi_1 = 0.3$ (b) $\phi_1 = 0.5$ (c) $\phi_1 = 0.7$



AR(1) C C2 SPC SPC2 AR(1) C C2 SPC SPC2 AR(1) C CST SPC SPC2
(a) (b) (c)

Fig 4. MSE of testing data set on simulation with $\sigma_\varepsilon^2 = 1$
and $\phi_1 = 0.3$ (b) $\phi_1 = 0.5$ (c) $\phi_1 = 0.7$



AR(1) C C2 SPC SPC2 AR(1) C C2 SPC SPC2 AR(1) C CST SPC SPC2
(a) (b) (c)

Fig 5. MSE of testing data set on simulation with $\sigma_\varepsilon^2 = 3$
and $\phi_1 = 0.3$ (b) $\phi_1 = 0.5$ (c) $\phi_1 = 0.7$

5. Conclusions

The results of forecasting with simulation of generated data with ARIMA(1,0,0) or AR(1) with non zero means using 9 scenarios of different σ_ε^2 (0.5, 1, 3) and 3 parameters ϕ_1 (0.3, 0.5, 0.7) showed that the increased σ_ε^2 and AR(1) parameter, then the variance of MSE with ARIMA also increased in the range of 0.0634 – 15.7633.

MSE from Cheng and Cheng2 FTS showed a much stable variance of MSE for each combination σ_ε^2 and ϕ_1. The variance MSE from Cheng FTS is in a range of 0.0712 – 2.9648 and variance MSE from Cheng2 FTS is in a range of 0.0640 – 2.7157.

The variance MSE from SP Cheng FTS is in a range of 0.0722 – 14.7045 and the variance MSE from SP Cheng2 FTS is in range of 0.0759 – 4.6803.

Despite SP Cheng and SP Cheng2 FTS method seemed look not better than Cheng and Cheng2 FTS, the more increased in AR(1) parameter showed MSE both methods are better than ARIMA even with a closer result with Cheng and Cheng2 FTS.

Cheng2 FTS method without used of sub-interval splitting had distribution center of MSE forecasting result which much lesser than Cheng FTS method. The use of historical data difference as universe of discourse in SP Cheng2 FTS at least able to make a stable MSE.

Acknowledgments

The authors wish to thank BPS-Statistics of Indonesia for allowing author to be part of 2nd Batch of BPS-APBN Master Scholarship Programme.

References

[1] Cheng CH, Chen TL, Teoh HJ, Chiang CH.. Fuzzy Time Series based on Adaptive Expectation Model for TAIEX Forecasting". Expert System with Application. 34: 1126-1132. 2008.
[2] Dan J, Dong F, Hirota K.. Fuzzy Local Trend Transform based Fuzzy Time Series Forecasting Model. International Journal of Computers, Communication & Control. Vol. VI , No. 4, pp. 603-614. 2011.
[3] Hasbiollah M, Hakim RB. F. Peramalan Konsumsi Gas Indonesia Menggunakan Algoritma Fuzzy Time Series Stevenson Porter. Disajikan pada Seminar Nasional Matematika dan Pendidikan Matematika. 2015. Jakarta. Universitas Islam Indonesia. 2015.
[4] Montgomery DC, et al. Introduction to Time Series Analysis and Forecasting. Canada: John Wiley and Sons, Inc. 2008.
[5] Song Q, Chissom BS. Forecasting enrollments with fuzzy time series-Part I. Fuzzy Sets and Systems. 54:1-9. 1993a.
[6] Stevenson M, J. E. Porter. Fuzzy time series forescasting using percentage change as the universe of discourse. Proceedings of World Academy of Science, Engineering and Technology. 55: 154-157. 2009.
[7] Tauryawati ML, Irawan MI. Perbandingan Metode Fuzzy Time Series dan Metode Box-Jenkins untuk Memprediksi IHSG. Jurnal Sains dan Seni POMITS Vol. 3, No. 2. 2014.
[8] Zhiqiang Z, Qiong Z. Fuzzy time series forescasting based on k-means clustering. Open Journal of Applied Sciences. Supplement: 2012 World Academy of Science, Engineering and Technology. 2012.

Wahyu Dwi Sugianto currently pursuing masters degree program in Applied Statistics in Bogor Agricultural University, Indonesia. Attained bachelor degree from STIS Jakarta in 2006.

Agus Mohamad Soleh is lecturer at Department of Statistics, Bogor Agricultural University, Indonesia. His main interest is in Data Mining, Statistical Computation and Statistical Modelling.

Farit Mochamad Afendi is lecturer at Department of Statistics, Bogor Agricultural University, Indonesia. Her main interest is in Geoinformatics.