Comparison of ARIMA, SSA, and ARIMA – SSA hybrid model performance in Indonesian economic growth forecasting

Fajar, Muhammad and Hartini, Sri

Badan Pusat Statistik-Statistics Indonesia

16 June 2020

Comparison of ARIMA, SSA, and ARIMA – SSA Hybrid Model Performance in Indonesian Economic Growth Forecasting

Muhammad Fajar[1]; Sri Hartini Rachmad[2]

[1]BPS-Statistics Indonesia, National Office,Statistician mfajar@bps.go.id
[2]BPS-Statistics Indonesia, National Office,Reseacher shrachmat2@gmail.com

Abstract:
The aim of this research is to compare among the performance of ARIMA, Singular Spectrum Analysis (SSA), and ARIMA-SSA hybrid model which is applied to Indonesian economic growth forecasting. Data used in this research is economic growth (quarter to quarter, q to q) 1983 Q2 – 2018 Q2 taken from Badan Pusat Statistik (BPS). The result of this research concludes that ARIMA-SSA hybrid method shows a better performance in economic growth forecasting compared to ARIMA and SSA based on the RMSE results.

Keywords: Hybrid model; ARIMA-SSA; Forecasting; Growth

1. Introduction:

One indication of a successful national development is when it is supported by enstablished economic sector. Economic development is a series of efforts to improve the people's living standards, provide employment opportunities, and decrease the inequality of prosperity and income. To develop an economic system that could support the community development, data are needed as the indicators for economic planning. Thus, the available data could ease the policy making so that the economic development reaches the target efficiently. An indicator to evaluate and plan the future national development is economic growth.

The occurance of an event sometimes does not correspond with the expected time for it to happen. Usually, there is a time gap (time lag) between the expectation and reality. This is the main reason of why forecasting is needed. When the time interval is zero or really short, then the forecasting would not be necessary. Otherwise, when the time interval is relatively long, then performing the forecasting would be important. In such case, forecasting is needed to determine when an event would or be expected to occur, thus the right decisions or policies could be made. Therefore, the ideal national economic structure development could be achieved with the planning mechanism in determining the strategy and policies so that the decisions taken would reach the optimal target and be on time by using the available resources.

Time series forecasting is a quantitative method used to analyze a series of data collected in time order using the right technique. The result could be used as a reference to forecast the value of the series in the future (Makridakis & MacGee, 1999). The development of forecasting methods is increasingly rapid and complex as advances in the development of computing technology. The interesting thing from the time series method development is the reconstruction of hybrid time series forecasting method, a time series constructed from two different types of forecasting method (Aladag *et al.*, 2012; Fajar, 2016; Fajar, 2018; Zhang, 2003; Zhang, 2011). This research introduces ARIMA – Singular Spectrum Analysis (SSA) hybrid forecasting method based on an analogy from the method proposed by Zhang (2003). Then, the method's performance is compared with both of the methods contstructing the hybrid.

2. Methodology:

2.1 Data Source

The data used in this research is economic growth (quarter to quarter, q to q) 1983 Q2 (quarter 2) – 2018 Q2 taken from Badan Pusat Statistik-Statistics Indonesia (BPS). The data for testing is divided into 20% observations (28 forecast ahead), 10% observations (14 forecast ahead), 5% observations (7 forecast ahead), and 3% observations (4 forecast ahead).

2.2 Analysis Method

2.2.1 ARIMA (Autoregressive-Moving Average)

In general, ARIMA $(p, d, q)(P, D, Q)^S$ model for x_t time series is:

$$\Phi_P B^S \phi_p(B)(1-B)^d(1-B^S)^D x_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t$$

with:

B	: lag operator.
p, q	: nonseasonal autoregressive order and nonseasonal moving average order.
P, Q	: seasonal autoregressive order and seasonal moving average order.
d	: nonseasonal differencing order.
D	: seasonal differencing order.
S	: seasonal period, for monthly data (S = 12), quarter data (S = 4).
$\phi_p(B)$	: nonseasonal autoregressive component.
$\Phi_P B^S$	: seasonal autoregressive component.
$\theta_q(B)$	: nonseasonal moving average component.
$\Theta_Q(B^S)$	: seasonal moving average component.
$(1-B)^d$	: nonseasonal differencing.
$(1-B^S)^D$	: seasonal differencing.
ε_t	: error term.

ARIMA model parameter estimation could be done by using maximum likelihood method. The best ARIMA model is determined by choosing a model with the lowest AIC (Akaike Information Criterion). AIC is calculated with the following formula:

$$AIC = -2 \times \log \text{maximum likelihood of ARIMA} + 2 \times \text{the number of parameter in ARIMA}$$

2.2.2 Singular Spectrum Analysis (SSA)

SSA is a non-parametric time series method based on multivariate statistics principle. SSA decomposes time series additively into several independent components. These components are identified as trend, periodic, quasi-periodic, and noise component. SSA procedure consists of four steps, they are:

- **Step 1. Embedding**

Given a $x_1, x_2, \ldots, x_T$ time series, choose an even number L, where L parameter is the window length defined as $2 < L < T/2$, and $K = T - L + 1$. The cross matrix is:

$$\boldsymbol{X} = (X_1, \ldots, X_T) = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_T \end{pmatrix}$$

The cross matrix proves to be a Hankel matrix, which means every element in the main anti diagonal has the same value. Thus, $\boldsymbol{X}$ could be assumed as multivariate data with L characteristic and K observations so that the covariance matrix is $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}'$ with dimension of $L \times L$.

- **Step 2. Singular Value Decomposition (SVD)**

Suppose that $\boldsymbol{S}$ has eigen value and eigen vector λ_i and U_i, respectively. Where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$ and $U_1, \ldots, U_L$. Thus, obtained SVD from $\boldsymbol{X}$ as follows:

$$\boldsymbol{X} = E_1 + E_2 + \cdots + E_d \qquad (1)$$

where $E_i = \sqrt{\lambda_i} U_i V_i'$, $i = 1,2, \dots, d$, E_i is the main component, d is the number of eigen value λ_i, and $V_i = \boldsymbol{X}' \, U_i / \sqrt{\lambda_i}$.

- **Step 3. Grouping**

In this step, $\boldsymbol{X}$ is additively grouped into subgroups based on patterns that form a time series, they are trend, periodic, quasi-periodic, and noise component. Partition the index set $\{1,2, \dots, d\}$ into several groups $I_1, I_2, \dots, I_n$, then correspond $\boldsymbol{X}_I$ matrix into group $I = \{i_1, i_2, \dots, i_b\}$ which is defined as:

$$\boldsymbol{X}_I = E_{i_1} + E_{i_2} + \dots + E_{i_b} \tag{2}$$

Thus, the decomposition represents as:

$$\boldsymbol{X} = \boldsymbol{X}_{I_1} + \boldsymbol{X}_{I_2} + \dots + \boldsymbol{X}_{I_n} \tag{3}$$

with $\boldsymbol{X}_{I_j} (j = 1,2, \dots, n)$ is reconstructed component (RC). $\boldsymbol{X}_I$ component contribution measured with corresponding eigen value contribution: $\sum_{i \in I} \lambda_i / \sum_{i=1}^{d} \lambda_i$. Using the close frequency range from the main components is based on the study of grouping process using auto grouping (Alexandrov & Golyandina, 2005). Main components with relatively close frequency ranges are grouped into one reconstructed component. So on, until several reconstructed components are formed.

- **Step 4. Reconstruction**

In this last step, $\boldsymbol{X}_{I_j}$ is transformed into a new time series with T observations obtained from diagonal averaging or Hankelization. Suppose that $\boldsymbol{Y}$ is a matrix with $L \times K$ dimensions and has $y_{ij}, 1 \le i \le L, 1 \le j \le K$ elements. Then, $L^* = \min(L, K), K^* = \max(L, K)$, and $T = L + K - 1$. Then, $y_{ij}^* = y_{ij}$ if $L < K$ and $y_{ij}^* = y_{ji}$ if $L > K$. $\boldsymbol{Y}$ matrix transferred into $y_1, y_2, \dots, y_T$ series with using the following formula:

$$y_k = \begin{cases} \dfrac{1}{k} \displaystyle\sum_{m=1}^{k} y_{m,k-m+1}^*, 1 \le k \le L^* \\ \dfrac{1}{L^*} \displaystyle\sum_{m=1}^{L} y_{m,k-m+1}^*, L^* \le k \le K^* \\ \dfrac{1}{T - k + 1} \displaystyle\sum_{m=k-K^*+1}^{T-K^*+1} y_{m,k-m+1}^*, K^* \le k \le T \end{cases} \tag{4}$$

Diagonal averaging on equation (4) is applied to every matrix component $\boldsymbol{X}_{I_j}$ on equation (3) resulting a $\widetilde{\boldsymbol{X}}^{(k)} = \left(\breve{x}_1^{(k)}, \breve{x}_2^{(k)}, \dots, \breve{x}_T^{(k)} \right)$ series. Thus, $x_1, x_2, \dots, x_T$ series is decomposed into an addition of reconstructed m series:

$$x_t = \sum_{k=1}^{m} \breve{x}_t^{(k)}, t = 1,2, \dots, T \tag{5}$$

2.2.3 SSA Forecasting

SSA forecasting used in this research is SSA recurrent, with estimating min-norm LRR (Linear Recurrence Relationship) coefficient. The LRR coefficient is calculated with the following algorithm:

1. Input: $\mathbf{P} = [P_1 : \dots : P_r]$ matrix, $\mathbf{P}$ is a matrix composed of U_i eigen vector from SVD step. Suppose that $\underline{\mathbf{P}}$ is a $\mathbf{P}$ that the last row is removed, and $\overline{\mathbf{P}}$ is a $\mathbf{P}$ that the first row is removed.

2. For every P_i vector column from **P**, calculate π_i, where π_i is a the last component from P_i , and $\underline{P_i}$ is a P_i that the last component is removed.
3. Calculate: $v^2 = \sum_{i=1}^{r} \pi_i^2$. If $v^2 = 1$, then STOP with a warning message "Verticality coefficient equals 1."
4. Calculate the min-norm LRR coefficient ($\mathcal{R}$):

$$\mathcal{R} = \frac{1}{1 - v^2} \sum_{i=1}^{r} \pi_i \underline{P_i}$$

5. From point (4) obtained: $\mathcal{R} = (\alpha_{L-1} \dots \alpha_1)'$.
6. Then, calculate the forecasting value with:

$$\hat{x}_n = \sum_{i=1}^{L-1} \alpha_i \tilde{x}_{n-1}, \qquad n = T + 1, \dots, T + h$$

2.2.3 ARIMA-SSA Hybrid

ARIMA – SSA hybrid method is a combination of ARIMA and Singular Spectrum Analysis (SSA) method. Time series data is assumed to consist of linear and nonlinear components, thus could be represented as:

$$x_t = P_t + N_t \tag{6}$$

with P_t is a linear component and N_t is a nonlinear component. ARIMA is used to forecast on linear component, then the residue from the linear component is the nonlinear component. Then, SSA is used to forecast the nonlinear component.

$$\hat{x}_{T+h} = \hat{P}_{T+h} + \hat{N}_{T+h} \tag{7}$$

with $\hat{x}_{T+h}$ is the x forecasting result on the $T + h$ period, $\hat{P}_{T+h}$ is the P forecasting result on the $T + h$ period, $\hat{N}_{T+h}$ N forecasting result on the $T + h$ period, and h is the ahead period.

2.2.4 Forecasting Performance Measurement

In the forecasting performance measurement, this research uses RMSE (Root Mean Square Error), because the result of Chai & Draxler (2014) research shows that RMSE fulfills the triangle inequality criterias for metric distance. RMSE formula is expressed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{h} \sum_{t=T+1}^{h} (\hat{x}_t - x_t)^2} \tag{8}$$

RMSE calculation is done according to the number of data used, RMSE from test data of 20% observations, 10% observations, 5% observations, and 3% observations.

3. Result:

Selected ARIMA model based on the smallest AIC is ARIMA (0,0,0) $(1,0,1)^4$, and the nonseasonal component in ARIMA is constant. The ARIMA (0,0,0) $(1,0,1)^4$ and SSA method directly forecasts the economic growth on a few periods ahead. While the ARIMA (0,0,0) $(1,0,1)^4$ – SSA in the linear component forecasting uses ARIMA, and the nonlinear component forecasting uses SSA. Then, the results of both components are aggregated thus ARIMA (0,0,0) $(1,0,1)^4$ – SSA forecasting result is obtained.

Table 3.1 presents RMSE according to the number of test data used from the observed method. In general, when the test data is smaller, the RMSE from ARIMA (0,0,0) $(1,0,1)^4$ and ARIMA (0,0,0)

$(1,0,1)^4$ – SSA hybrid is decreasing, whereas the RMSE result of SSA is unstable. ARIMA-SSA hybrid method gives a minimum RMSE compared to the other two methods. This shows that forecasting performance of ARIMA-SSA hybrid method is better than ARIMA and SSA.

Table 2.1 RMSE of ARIMA, SSA, and ARIMA-SSA Hybrid Method.

Method	Forecast Ahead			
	28	**14**	**7**	**4**
ARIMA (0,0,0) $(1,0,1)^4$	**1.764**	1.691	1.118	0.843
SSA	2.207	2.374	2.523	2.507
ARIMA (0,0,0) $(1,0,1)^4$-SSA hybrid	1.861	**1.674**	**1.092**	**0.813**

source: author.

4. Discussion, Conclusion and Recommendations:

Based on the previous discussion, it could be concluded that the forecasting performance of ARIMA – SSA hybrid is better than ARIMA and SSA in Indonesian economic growth forecasting (q to q), based on the result of ARIMA – SSA hybrid RMSE is smaller than ARIMA and SSA method.

References:

1. Aladag. C.H., Egrioglu. E. & Kadilar. C. (2012). Improvement in Forecasting Accuracy Using the Hybrid Model of ARFIMA and Feed Forward Neural Network American. *Journal of Intelligent Systems* **2(2)**, pp 12-17.
2. Alexandrov. Th. & Golyandina. N. (2005). Automatic extraction and forecast of time series cyclic components within the framework of SSA. *In Proceedings of the 5th St.Petersburg Workshop on Simulation*, June 26 – July 2, 2005, St.Petersburg State University,St.Petersburg, pp 45–50.
3. Chai. T. & Draxler. R.R. (2014). Root mean square error or mean absolute error (MAE)?- Arguments against avoiding RMSE in the literature. Geosci. *Model Dev* **7**, pp 1247- 1250.
4. Chang. P.C. Wang. Y.W. & Liu. C.H. (2007). The development of a weighted evolving fuzzyneural network for PCB sales forecasting. *Expert Systems with Applications* **32**, pp 86–96.
5. Cryer. J.D. & Chan. K.S. (2008). Time series Analysis: With Application in R, Second Edition. USA: Spinger Science and Businiess Media, LLC.
6. Fajar. M. (2016). Perbandingan Kinerja Peramalan Pertumbuhan Ekonomi Indonesia antara ARMA, FFNN dan Hybrid ARMA-FFNN.DOI: 10.13140/RG.2.2.34924.36483.
7. Fajar. M. (2018).Meningkatkan Akurasi Peramalan dengan Menggunakan Metode Hybrid Singular Spectrum Analysis-Multilayer Perceptron Neural Networks.DOI:10.13140/RG.2.2.32839.60320.
8. Makridakis. W. & MacGee. (1999). Metode dan Aplikasi Peramalan. Binarupa Aksara.
9. Rahmani. D. (2014). A forecasting algorithm for Singular Spectrum Analysis based on bootstrap Linear Recurrent Formula coefficients. *International Journal of Energy and Statistics* **2 (4)**, pp 287–299.
10. Wei. W.W.S. (2006). Time series Analysis: Univariate and Multivariate Methods. Pearson Education, Inc.
11. Zhang. G.P. (2003). Time series Forecasting using a Hybrid ARIMA and Neural networks Model. *Neurocomputing* **50**, pp 159-175.
12. Zhang. Q. Wang. B.D., He. B., Peng. Y. & Ren. M.L. (2011). Singular Spectrum Analysis and ARIMA Hybrid Model for Annual Runoff Forecasting. *Water Resources Management* 25, pp 2683-2703.