14th National Convention on Statistics (NCS)
Crowne Plaza Manila Galleria, Quezon City
October 1-3, 2019

COMPARISON OF ARIMA AND SINGULAR SPECTRUM ANALYSIS IN FORECASTING THE PHILIPPINE INFLATION RATE

by

Arniel A. Alderite and Anthony F. Capili

For additional information, please contact:

Author's name	Arniel A. Alderite
Designation	Instructor I
Affiliation	Davao Oriental State College of Science and Technology
Address	Pagsilaan, Don Enrique Lopez, City of Mati, Davao Oriental 8200
E-mail	arniel.alderite@doscst.edu.ph

Author's name	Anthony F. Capili
Designation	Instructor I
Affiliation	University of Southeastern Philippines

COMPARISON OF ARIMA AND SINGULAR SPECTRUM ANALYSIS IN FORECASTING THE PHILIPPINE INFLATION RATE

by

Arniel A. Alderite and Anthony F. Capili

ABSTRACT

In this study, an appropriate forecasting method for the Philippine inflation rate was determined. Specifically, the forecasting performance of the Autoregressive Integrated Moving Average (ARIMA) model and Singular Spectrum Analysis (SSA) were compared. The dataset on monthly inflation rate was divided into two samples, the in-sample data and out-of-sample data. These samples were analyzed and after application of the Box-Jenkins approach, the resulting seasonal model is ARIMA$(1,1,0)\times(0,0,1)_{12}$. This model has the least AIC among all tentative models and the behavior of its residuals and forecast errors are satisfactory. SSA was also applied on the same samples and the most appropriate window length for the trajectory matrix is 25. In reconstructing the series, 10 eigentriples under three groups were used. The first group contains the trend component while the second and third groups contain the oscillatory components. Based on the computed in-sample and out-of-sample RMSE, SSA outperforms the ARIMA model. Thus, SSA is better than ARIMA in forecasting the Philippine inflation rate based on the sample data set.

1. Introduction

Studying and analyzing the inflation rate is of economic importance because low inflation rate is indicative of the drop in consumer demand, a sign of a broader financial problem, and harming sales figures for it encourages individuals and business to postpone purchases. A high inflation rate impairs individuals and companies purchasing power and harms economies (Adam, 2016). Doloriel, Salvaleon, Ronquillo, and Estal, (2014) signified the forecasting of inflation rate for many reasons: (a) it guides economic policymakers in their policy-decision making; (b) it gives baseline reference for owners and business administrators in their financial projections; and (c) workers impute inflation forecasts help determine wages that they ask from their employers.

In the Philippines, several studies have been conducted on the inflation rate. Mariano (1985) used regression analysis in forecasting monthly inflation as measured by the changes in the Consumer Price Index (CPI). Mariano, Dakila, and Claveria (2003) also constructed a structural long-term annual macroeconomic model for inflation. A generalized autoregressive conditionally heteroskedastic (GARCH) models for forecasting the volatility of Philippine inflation was developed by Ramon (2008). Medalla and Ferno (2013) utilized an autoregressive-moving-average and generalized autoregressive conditional heteroskedastic (ARMA-GARCH) models for describing the inflation rate. Doloriel et al. (2014) investigated the trend and projected the volatility of Philippine inflation using fractal analysis. Among several methods performed on Philippine inflation rate, one approach that has not been utilized, at least based on the current studies considered, is Singular Spectrum Analysis (SSA).

Golyandina, Nekrutkin, and Zhigljavsky (2001) defined Singular Spectrum Analysis (SSA) as "a relatively new non-parametric approach for analyzing time series data which incorporates elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems and signal processing and is very useful in the prediction of non-stationary time series." Methods for modeling and forecasting time series such as auto-regressive moving average (ARMA) models suffer from parametric restrictions (stationarity and normality). Although the transformation of a non-stationary series by differencing or detrending is possible, a large amount of information is lost by such data transformation. SSA does not depend on any parametric model for the trend or oscillations and does not make any assumptions about the signal or the noise component of the data (Hassani, Soofi, & Zhigljavsky, 2013).

Review on the literature of SSA showed that there are more than a hundred papers on the application of SSA in different areas and, in most of them, SSA is superior compared to other time series analysis techniques (Mahmoudvand, Alehosseini, & Rodrigues, 2015). Also, recent developments in the theoretical and methodological aspects of SSA from the perspective of analyzing and forecasting economic and financial time series showed that SSA could be considered a suitable technique for predicting inflation rate (Hassani & Thomakos, 2010).

The application of the Box-Jenkins approach and SSA methods to the Philippine Inflation rate are compared and evaluated in this study and the most accurate method was used to forecast monthly inflation rate for the years 2019 to 2020.

2. Methods

The data in this study contains the monthly inflation rate of the Philippines using the 2012-based Consumer Price Index (CPI). The data on monthly inflation rate was retrieved from the PSA's website. The monthly inflation rate data ranges from January 2013 to December 2018. A total of 72 observations were used in the analysis. An in-sample dataset of 58 observations (January 2013 to October 2017) was used for Box-Jenkins approach in model development. The same set of observations was used for the decomposition and reconstruction stages of Singular Spectrum Analysis (SSA). The remaining 14 observations (November 2017 to December 2018) served as the out-of-sample dataset. This sample was used for forecast evaluation.

To find an appropriate ARIMA model, the Box-Jenkins approach was used. The Box-Jenkins approach are divided in three stages, namely, model identification, model estimation, and diagnostic checking (Box and Jenkins, 1976). In addition, forecast evaluation was also performed to check if forecast errors are satisfactory.

In model identification stage, the stationarity of the series is checked by examining the time series plot. A formal test (Augmented Dickey-Fuller test) is also performed on the data. Differencing is performed if the series is not stationary. Given a stationary time series data, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots are obtained. These plots are used as basis in the identification of tentative models. Among the tentative models obtained, the model that has the smallest Akaike Information Criterion (AIC) value is selected for estimation.

In the model estimation stage, the unknown parameters of the selected model are estimated. The combination of conditional sum of squares and maximum likelihood estimation is employed to estimate the parameters of the chosen model.

After the estimation process, the adequacy of the selected model is assessed and evaluated. Residual analysis is performed. Ideally, the residuals of an adequate model should behave like a Gaussian white noise process. In performing residual analysis, no patterns should be observed. Ljung-Box test is also performed, and the p-value should be large enough to conclude that the model is adequate. ACF and PACF plots of the residuals are also constructed and there must be no spikes in these values.

In addition, out-of-sample forecasts is considered in evaluating the performance of the model. The difference between the actual values and the forecast values at time t are called forecast errors. The ACF and PACF plots of these forecast errors are examined for significant values. A normality test is performed on the forecast errors. This test determines if the assumption of normally distributed forecast errors is reasonable.

The procedure for Box-Jenkins approach is summarized in Figure 1.

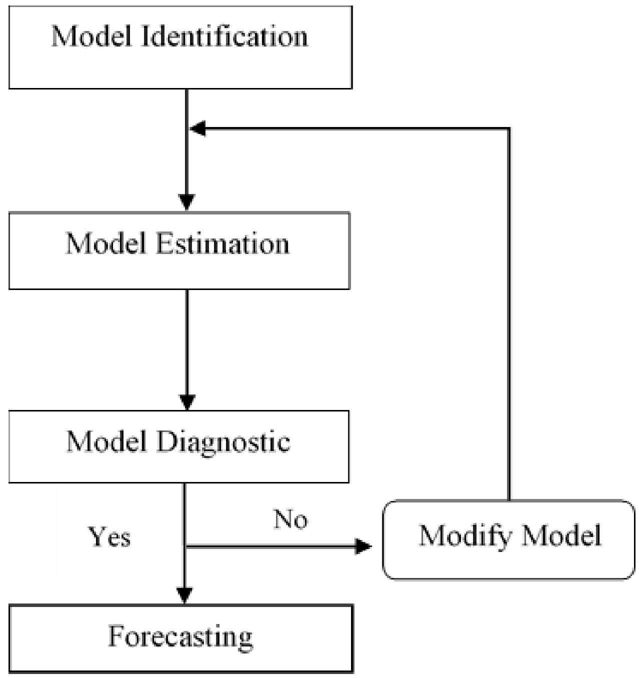

Figure 1. Box-Jenkins Flowchart

Singular Spectrum Analysis is composed of two stages. Specifically, the procedure involves decomposition and reconstruction stages (Golyandina et al., 2001).

Under the decomposition stage, a series N was embedded into a trajectory matrix by choosing an appropriate window length L where $1 < L < N/2$ and Singular Value Decomposition (SVD) was performed on the trajectory matrix. This process results into L eigentriples and elementary matrices.

In the reconstruction stage, the elementary matrices were summed into a smaller number of groups called the resultant matrices that represent different underlying aspects of the time series and combine the various components. These components are typically the 'trend' components, which are smooth parts of the series that show long-term variance, the various 'oscillatory' components that indicate periodicities in the time series and finally the largely undesired 'noise' components. The process of combining the matrices into different groups or subsets is known as the eigentriple grouping. Groups were determined through the help of analyzing the scree plot of the singular values and the percent of contribution of each eigentriple. An eigentriple is considered as noise if the percent of contribution is low (e.g., less than 0.05%). Also, the separability between noise and signal was determined via the w-correlation plot. Once the individual components of the time series have been separated into the relevant groups, the original time series was reconstructed with a smaller number of components. This was done by performing diagonal averaging on the matrices resulting from the summation of the relevant products in each of the groups.

The steps under Singular Spectrum Analysis are summarized in Figure 2.

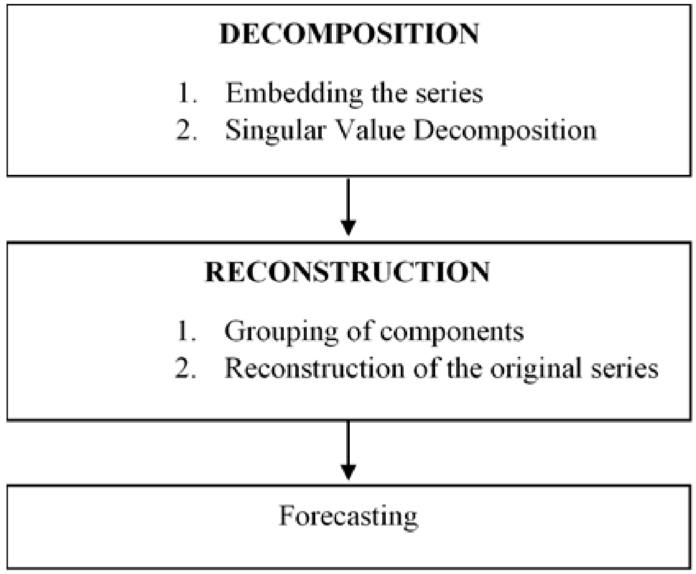

Figure 2. Singular Spectrum Analysis Flowchart

In determining the most accurate method, the root mean square error (RMSE) was utilized. RMSE was computed for both Box-Jenkins approach and Singular Spectrum Analysis and was the basis for comparison. The method with the least RMSE was considered as best in terms of forecasting performance.

3. Results and Discussion

Figure 3 shows the time series plot of the monthly inflation rate in the Philippines from January 2013 until December 2018. During the years 2013 to 2014, inflation rate is within the government's average inflation target of 2% to 4%. Low inflation can be seen during the first quarter of 2015 until the last quarter of the same year. There is a gradual increase from 2016 until the first quarter of 2018 but it is still within the government's average inflation target. High inflation started in the second quarter of 2018 until the middle of quarter three and a gradual decrease started in the 4th quarter of 2018.

The series seems to be nonstationary. There are irregular peaks and troughs with no consistent trend. Furthermore, its variance appears to be constant.

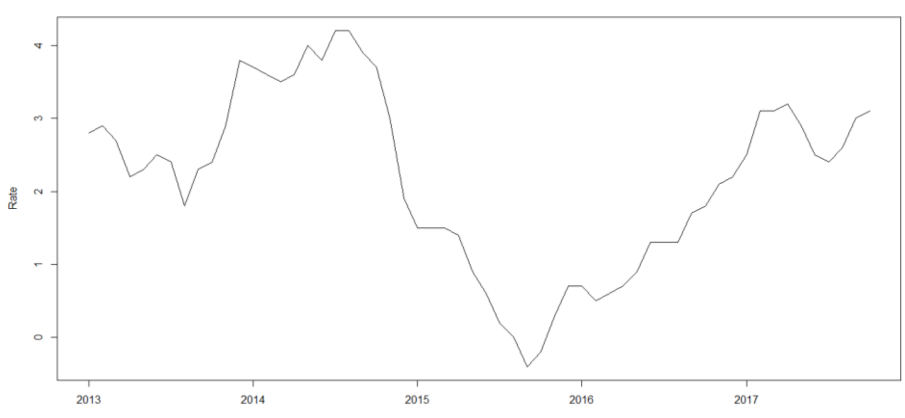

Figure 3. Time Series Plot of Monthly Inflation Rate

Upon application of the Box-Jenkins approach to the data. The data was transformed through first order differencing to achieve stationarity. Stationarity was determined using the Augmented Dickey-Fuller test. Tentative models were identified and considered upon careful examination of the ACF and PACF plots. Also, seasonal components for the model were considered since the ACF and PACF plots showed significant spikes at seasonal lags.

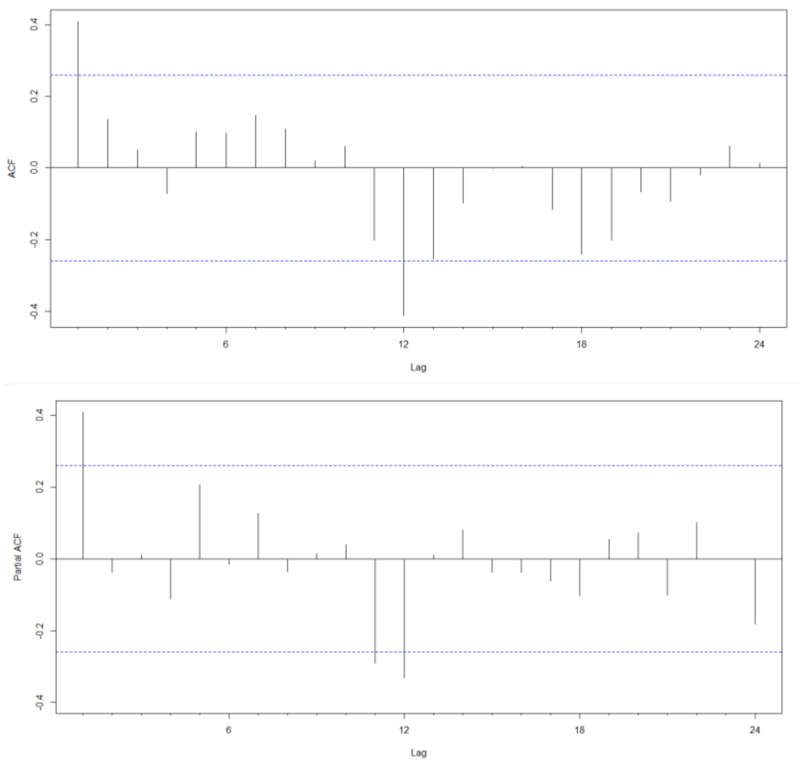

Figure 4. ACF and PACF Plots the First Difference

Among the tentative models in table 1, ARIMA(1,1,0)×(0,0,1)$_{12}$ was chosen since it has the least AIC value and its parameters are significantly different from zero.

Table 1. Tentative ARIMA Models

Model	Model Term	Estimate	Standard error	z-value	p-value	AIC Value
ARIMA(0,1,1)×(0,0,1)$_{12}$	MA(1)	0.309	0.117	2.639	0.008	21.16
	SMA(1)	-0.999	0.469	-2.132	0.033	
ARIMA(1,1,0)×(0,0,1)$_{12}$	AR(1)	0.339	0.124	2.735	0.006	20.43
	SMA(1)	-0.999	0.509	-1.961	0.049	
ARIMA(1,1,1)×(0,0,1)$_{12}$	AR(1)	0.329	0.335	0.986	0.324	22.43
	MA(1)	0.010	0.348	0.029	0.976	
	SMA(1)	-0.999	0.508	-1.968	0.049	

Residual plots of the model are shown in figure 5. The residuals showed that there is no observable structure in the plot of residuals versus order. The plot of residuals versus fitted values is also structureless. Some of the residuals are far from the theoretical line. This is an indication that the residuals are not normally distributed. However, in general, the behavior of the residuals is still satisfactory.

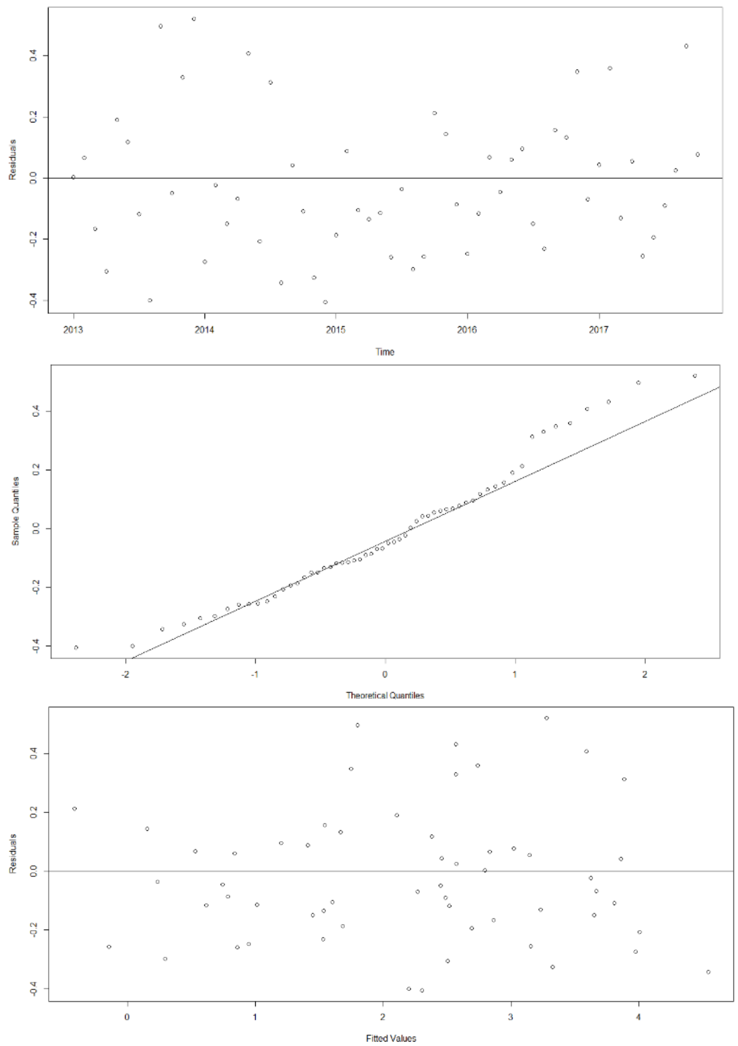

Figure 5. Residual Plots

The model's performance was also evaluated by examining the forecast errors and its forecast errors behave like a Gaussian white noise process and implies that the model is adequate.

Under the Singular Spectrum Analysis, the window length L is the only parameter in the decomposition stage. The choice of the window length L is based on the following: L should be large enough but not greater than half of a series or $1 < L < N/2$.

The out-of-sample root mean squared error (RMSE) were computed for different values of L (see Table 2). For each L, the number of eigentriples r was determined based on its percent of contribution. That is, an eigentriple was dropped if the percent of contribution is lower than 0.05%. Among different values of L and r, a window length L of 25 with r of 10 reported the least RMSE value.

Table 2. SSA Parameter Combinations

L	r	RMSE (out-of-sample)
28	11	5.12333
27	11	3.69589
26	10	1.02821
25	10	1.01322
24	10	1.08574
23	10	1.28077
22	10	1.49280
21	9	4.05266
20	9	4.15623
19	9	3.92895
18	9	3.67928
17	8	1.55733
16	8	1.71723
15	8	3.69013
14	7	3.91772
13	7	3.88818
12	7	3.28714
11	6	3.35588
10	6	3.88910
9	7	2.63037
8	7	2.60364
7	6	4.81758
6	5	2.56361
5	4	3.11883
4	3	2.72732
3	2	2.53457
2	1	2.15485

The monthly inflation rate data was embedded into an $(L \times K)$ trajectory matrix with window length $L = 25$ and $K = 34$.This trajectory matrix was decomposed using Singular Value Decomposition (SVD). This process results in 25 eigentriples and 25 elementary matrices $X_1, X_2, \dots, X_{25}$.

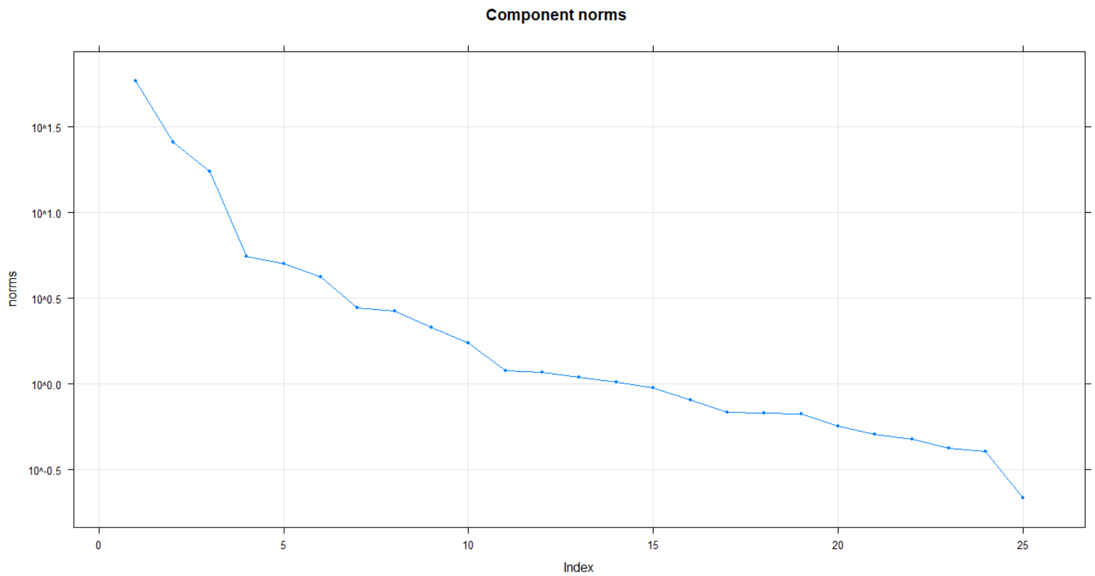

Figure 6: Scree Plot of Singular values

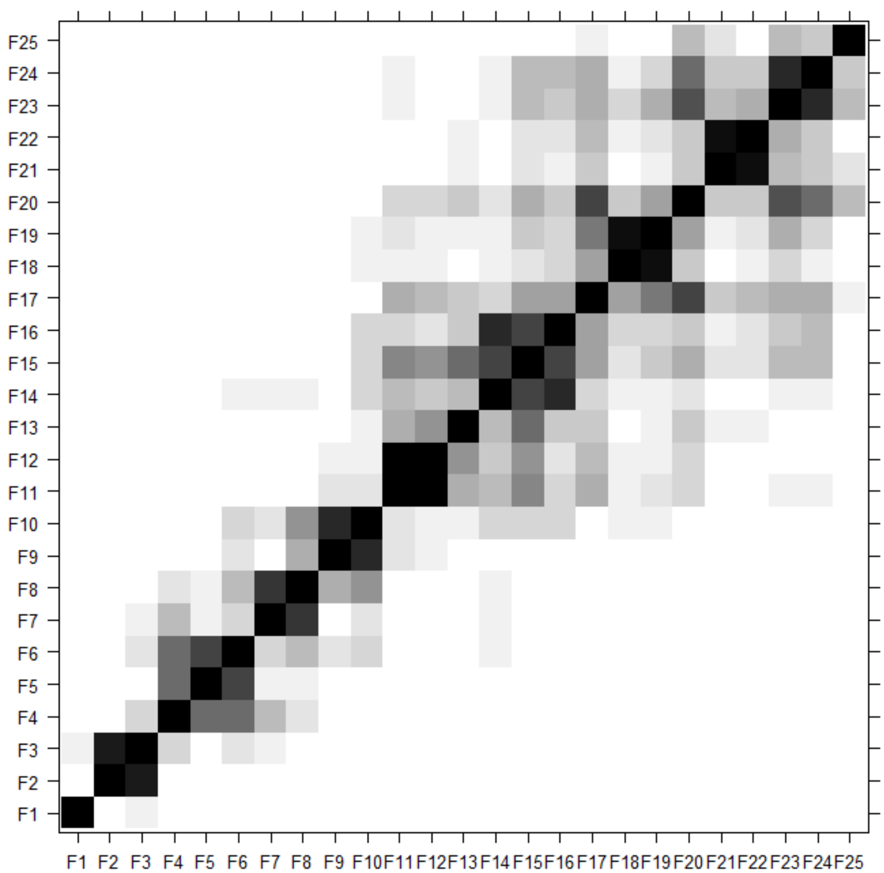

Figure 7. W-correlation Plot

The scree plot of the singular values or components of the decomposed trajectory matrix is shown in Figure 6. The “elbow” or “shark break” is the key in the scree plot. The components before the “elbow” are the important ones. In the figure, an “elbow” or “shark break”, can be seen at component 11. This indicates that the significant components are from 1 to 10 only.

The w-correlation plot was also used to determine groupings as shown in Figure 7. The w-correlation plot suggest that eigentriples 1 until 10 are the only significant eigentriples. This means that these eigentriples are enough to reconstruct the decomposed series. The shade of each square represents the strength of the w-correlation between two eigentriples.

The w-correlation plot showed for this case that the first eigentriple is uncorrelated with the other eigentriples. This means that the first eigentriple (F1) is considered as the first group. The second (F2) and third (F3) eigentriples are highly correlated and are considered as the second group. The fourth (F4) until the tenth (F10) eigentriples are correlated and is considered as the last group. The eigentriples (from 11 to 25) can be associated to the inherent noise of the series.

Each group is a resultant matrix X_I. Group 1 (F1) is the resultant matrix X_{I_1}, group 2 (F2 and F3) is X_{I_2} and group 3 (F4 to F10) is X_{I_3}. Diagonal averaging was performed on each resultant matrix. The first group represents the trend component and the second and third group represent the oscillatory components. Combining all the groups, the reconstructed series was used to forecast. In this study, recurrent forecasting algorithm was used.

The two methods were compared based on their in-sample RMSE and out-of-sample RMSE. For both the in-sample and out-of-sample forecasts, the method with the smallest overall RMSE is Singular Spectrum Analysis (see Table 3). Hence, SSA outperformed ARIMA in terms of forecasting performance.

Table 3. RMSE of ARIMA and SSA

Model	RMSE (In-sample)	RMSE (Out-of-sample)
ARIMA$(1,1,0)\times(0,0,1)_{12}$	0.22631	2.61352
SSA (L=25, r =10)	0.09736	1.01322

Table 4 shows the forecasted values of SSA. The 95% confidence intervals are also included in the said table.

The forecasted values as shown in Table 4 exhibits fluctuating characteristics. This is an indication that the monthly inflation rate will continue to fluctuate in the future. Most of the forecasted values are outside the government’s target of 2% to 4% inflation. However, upon considering the confidence interval, there is still a possibility that future values will be within the target. That is, the lower limits meet the government target.

Table 4. Forecasted Values

Year	Month	Forecast	Lower 95% Confidence Interval	Upper 95% Confidence Interval
2019	January	5.27	4.68824	6.18835
2019	February	5.48	4.31763	7.09613
2019	March	5.61	4.16414	7.51100
2019	April	5.31	3.58165	7.08463
2019	May	4.60	3.02226	6.52903
2019	June	3.91	2.34205	5.68684
2019	July	3.68	1.57154	6.44166
2019	August	4.05	1.08013	7.88897
2019	September	4.62	0.88252	8.55115
2019	October	4.92	0.87329	8.16493
2019	November	4.80	1.37914	8.07280
2019	December	4.59	1.96272	7.70301
2020	January	4.76	2.41468	7.79095
2020	February	5.38	2.66729	9.33829
2020	March	6.07	3.07584	10.22512
2020	April	6.31	3.43859	9.78401
2020	May	5.94	3.71362	9.23585
2020	June	5.41	3.72147	9.11229
2020	July	5.27	3.87464	9.48572
2020	August	5.69	4.14099	11.74887
2020	September	6.26	3.82312	12.86765
2020	October	6.41	3.47471	12.04029
2020	November	5.88	3.09269	11.12080
2020	December	5.09	2.81583	10.96086

Conclusion/s and Recommendations

The comparison of ARIMA and SSA in this study showed that SSA is superior in terms of forecasting performance. Hence, between these methods, SSA is a more appropriate forecasting method for Philippine inflation rate.

Based on the results and conclusions, the following recommendations are made.

1. The choice of SSA window length L in this study is very subjective. However, there are recent studies and new algorithms of SSA (e.g., Circulant SSA), that may improve the results of the analysis. Hence, this study also suggests utilizing them and compare the results herein.
2. Policy makers and economic managers can use the results of this study as basis for economic policies and fiscal programs that try to control inflation within normal levels.
3. Business owners and administrators can use the forecasted inflation rate as baseline reference for financial projections. This will help in their strategic plans and business operation.
4. Other forecasting methods may also be considered and compare the results in this study.

References

Adam, J. A. (2016, December 28). What is inflation, how does it work and why is it important? Retrieved from https://www.morganintl.com/blog/finance-treasury/what-is-inflation-how-does-it-work-and-why-is-it-important/

Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control, revised ed*. Holden-Day.

Doloriel, N.S., Salvaleon, R.G., Ronquillo, A.O., & Estal, B.R., (2014). Fractal analysis of Philippines inflation rate. *SDSSU Multidisciplinary Research Journal Vol. 2 No. 1 Jan.–June 2014,* 179-182. Retrieved from https://www.smrj.sdssu.edu.ph/index.php /SMRJ/article/view/57/55

Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. (2001). *Analysis of time series structure: SSA and related techniques*. Chapman & Hall/CRC, New York – London.

Hassani, H. (2007). Singular spectrum analysis: methodology and comparison. *Journal of Data Science , Vol. 5, Number 2 April 2007*, 239-257.

Hassani, H., & Thomakos, D. (2010). A review on singular spectrum analysis for economic and financial time series. *Statistics and its Interface, 3(3)*, 377-397. doi: 10.4310/SII.2010.v3.n3.a11

Hassani, H., Soofi A.S., & Zhigljavsky A. (2013). Predicting inflation dynamics with singular spectrum analysis. *Journal of the Royal Statistical Society. Series A (Statistics in Society), Vol. 176,* 743-760. doi:10.1111/j.1467-985X.2012.01061.x

Mahmoudvand, R., Alehosseini, F., & Rodrigues, P. C., (2015). Forecasting mortality rate by singular spectrum analysis. *REVSTAT – Statistical Journal Volume 13, Number 3, November 2015,* 193–206. Retrieved from https://www.ine.pt/revstat/pdf/rs150301.pdf

Mariano, R. S. (1985). Forecasting monthly inflation in the Philippines. The Philippine Institute for Development Studies. Retrieved from https://opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/3463/pidsms85-10.pdf?sequence=1

Mariano, R. S., Dakila, F. G., & Claveria, R. A., (2003). The Bangko Sentral's structural long-term inflation forecasting model for the Philippines. *Philippine Review of Economics, 40, issue 1,* 58-72. Retrieved from https://EconPapers.repec.org/RePEc: phs:prejrn:v:40:y:2003:i:1:p:58-72.

Ramon, Haydee L., (2008). *Forecasting the Volatility of Philippine Inflation using GARCH Models.* BSP Working Paper Series No. 2008-01. Retrieved from http://www.bsp.gov.ph/downloads/Publications/2008/WPS200801.pdf