

# Forecasting Vietnam's Inflation Rate Based on Arima, Sarima, Scarima Models

Nghiem Van Tinh<sup>1\*</sup>, Nguyen Tien Duy<sup>1</sup>

\*Email: [nghiemvantinh@tnut.edu.vn](mailto:nghiemvantinh@tnut.edu.vn)

<sup>1</sup>Thai Nguyen University of Technology – Thai Nguyen University, Thai Nguyen, Viet Nam

## Abstract:

Time series analysis and forecasting is an efficient versatile tool in diverse applications such as in economics and finance, hydrology and environmental management fields just to mention a few. Among the most effective approaches for analyzing time series data, the method propounded by Box and Jenkins, the Autoregressive Integrated Moving Average (ARIMA) was employed in this study. The paper outlines the practical steps which need to be undertaken to use Autoregressive integrated moving average (ARIMA) time series models for forecasting the monthly inflation in Vietnam. A framework for ARIMA forecasting is drawn up. It considers two alternative approaches to the issue of identifying ARIMA models - the Box Jenkins approach and the objective penalty function methods. To investigate the performance of the forecasting model, a numerical dataset of inflation was collected monthly from the General Statistics Office of Vietnam (GSO) for the period from January 2004 to July 2014 are examined. The examination result shows that average inflation rate in the second half of 2014 tends to be higher than in the first half of 2014 and the average annual inflation is 5.98% based on the mean absolute percentage error (MAPE) criteria.

## Keywords

Forecasting, inflation rate, ARIMA model, SCARIMA model.

## 1. Introduction

There is no official notification specifying the time when an inflation-targeting framework will be conducted. However, in recent years the State Bank of Vietnam (SBV) has been successfully controlling inflation via monetary policy, which focuses not only on promoting economic growth but also on stabilizing the macro economy. This realization has been supported significantly by forecasting inflation. The SBV has incentives to perform inflation forecasting because one of its main tasks, as specified by the State

Bank Law of 2010, is to stabilize price levels to stabilize the domestic currency (VND). The Forecasting and Statistics Department, called FSD was established in December 2008 to implement the task of performing statistical and macroeconomic forecasts to support monetary policy, in which forecasting inflation is a core function. The FSD has been submitting its monthly inflation reports to the SBV Management Board since middle 2012. Additionally, forecasting inflation is indispensable work to prepare for an inflation-targeting framework, which will hopefully be conducted by the SBV in the near future. The problem with inflation is one that affects all the countries in the world. As such, it has become a primary objective in monetary policy many countries, in there has Vietnam. The adverse effects of the high inflation levels in investment and on the level of consumption become key essential problems today. Therefore, forecasting for this inflation problem is an issue that economists and policy-makers consider when planning their unit. The more accurate the forecast is, the more likely the plan will be. Stock & Watson (2007) argue that inflation is increasingly unpredictable as more and more information is available to planners. Nowadays, there are many different models used in forecasting. Each forecasting model has its own advantages and disadvantages (Khashei & Bijari, 2011). Muhammad Abdus Salam, Shazia Salam and Mete Feridun (2006) used the ARIMA model with Box-Jenkins (1976) to predict short-term inflation in Pakistan. M.Z. Babai, M.M. Ali and J.E. Boylan (2011) find forecasting model for automotive part and they found Regression suitable for automotive part data. Chi-Chen Wang (2011) found the results indicate strong potential for the use of seasonal ARIMA modeling and the extension of Holt-Winters for predicting up to about two to three days ahead and that, for longer lead times, a simplistic historical average is difficult to beat. Recently, Ezekiel N.N. Nortey et al. sought to model rates of inflation in Ghana using the Autoregressive Conditional Heteroscedastic model. For this purpose, two authors have found the ARIMA model (6,1,6) to be appropriate to predict the next 12 months inflation in Ghana with a 95% confidence interval. According

to Khashei & Bijari (2011), the ARIMA model is very suitable for linear relations between existing and past data. Due to the high applicability of the Box-Jenkins method and the effectiveness of the ARIMA model in short-term forecasting. This paper still uses that approach to establish model for forecasting inflation of Vietnam. With the data collected extensively and updated (from January 2004 to August 2014) and rigorous testing, the expected results will contribute to the provision of useful information to the Government in Effort to manage annual inflation at a one-digit level based on reliable and scientific basis. At the same time, it also provides investors with a basis for measuring and assessing the potential impact of inflation. Thereby, providing solutions to mitigate the negative effects of inflation. The main objective of the study is to compare both in sample and out of sample forecasting performance between three univariate models (ARIMA, SARIMA, SCARIMA) used to model the inflation rates in VietNam. This paper includes three sections in addition to the introduction. Section 2 presents step-wise of the forecasting model. Section 3 gives forecasting results and evaluates the forecasting performance of the forecasted models. The last section summarizes the conclusions derived from the empirical results

## 2. Forecasting models

Time series forecasting is the use of a model to forecast future events based on known past events: to predict data points before they are measured. An example of time series forecasting in econometrics is predicting the opening price of a stock based on its past performance. Models for time series data can have many forms and represent different stochastic processes. When modeling variations in the level of a process, three broad classes of practical importance are: (1) the autoregressive (AR) models, (2) the integrated models, and (3) the moving average (MA) models. These components are combined which named autoregressive integrated moving average (ARIMA). This model was first introduced by box & jenkins (1976). According to gujarati (2004), estimating of the ARIMA model includes four main steps as follows:

### Step 1: Recognizing model

As introduced above, Arima model includes 3 components. Let's denote AR is p; I is d; MA is q, the new have Arima (p, d, q). To apply the ARIMA model (p, d, q) for the prediction, we must first identify the three components p, d, and q of the model.

Integrated part of the model can be defined by stationary test of a time-series. If the time-series integrate at level 0, we have I(d=0). If the time-series integrate at level 1, we have I(d=1). or if the time-

series integrate at level 2, we have I(d=2). The popular method is used for stationary test is Dickey-fuller.

After stationary test, we define p and d by using autocorrelation function (ACF) and partial correlation function (PACF). Where:

The AR (1) model uses only the first-order term, but in general, one may use additional higher-order AR terms. Each AR term corresponds to the use of a lagged value of the residual in the forecasting equation for the unconditional residual. An autoregressive model of order p, AR (p) has the form according to equation (1) as follows:

$$Y_t = \beta_1 X_t + \beta_2 X_{t-1} + \dots + \beta_p X_{t-p} + \varepsilon_t \quad (1)$$

Each integration order corresponds to differencing the series being forecast. A first-order integrated component means that the forecasting model is designed for the first difference of the original series. A second order component corresponds to using second differences, and so on

$$Y_t = \varepsilon_t + \alpha_1 X_t + \alpha_2 X_{t-1} + \dots + \alpha_q X_{t-q}$$

A moving average model uses lagged values of the forecast error to improve the current forecast. A first-order moving average term uses the most recent forecast error; a second-order term uses the forecast error from the two most recent periods, and so on. An MA (q) has the form according to equation (2) as follows:

$$Y_t = \varepsilon_t + \alpha_1 X_t + \alpha_2 X_{t-1} + \dots + \alpha_q X_{t-q} \quad (2)$$

Where;  $Y_t$  is the time series to predict,  $u_t$  is the error of the model.

If time-series belongs to type MA (q), the coefficients of ACF have statistical meaning from 1 to q and the values decrement rapidly to 0. As for PACF, partial coefficients will also decrement to 0.

P values are identified through ACF and PACF charts. If the time series belongs to type AR (p), the PACF diagram will have a statistically significant correlation coefficient of 1 to p and the values will decrease rapidly to 0. With ACFs partial coefficients will also decrement to 0. The combination (1) and (2) have the ARMA (p, q) model as follows.

$$Y_t = \beta_1 X_t + \beta_2 u_{t-1} + \dots + \beta_p u_{t-p} + \varepsilon_t + \alpha_1 X_t + \alpha_1 X_{t-1} + \dots + \alpha_q X_{t-q} \quad (3)$$

### Step 2: Estimating parameters and selecting model for prediction

All parameters of the model will be estimated using the Eview software. The process of choosing model is the process of comparing adjusted  $R^2$ , AIC criteria and p-value of models until having the best model with p-value of each parameters are smaller than 0.05. This model is the best model for forecasting.

### Step 3: Verify the model

In order to ensure the coherence of Arima model, the error of this model have to be white noise. We can use autocorrelation function ACF or Breusch-Godfrey test to check autocorrelation of errors. We use White ARCH test to check whether the residual variance of errors in this model is constant. If any result is not meet the requirement, such as there is autocorrelation between errors or error of this model is not white noise, then we have to do step 2 again.

**Step 4:** Forecasting and forecast evaluation

After testing the errors, if the model is consistent, the model will be used in forecasting. The forecast consists of two main parts: forecasting inside of sample and forecasting outside of sample. To evaluate the performance of the forecasted model, the root mean square error (RMSE), the mean absolute percentage error (MAPE) and  $R^2$  (coefficient of determination) are employed as evaluation criterions to represent the forecasted accuracy. The RMSE, MAPE,  $R^2$  value are calculated as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (4)$$

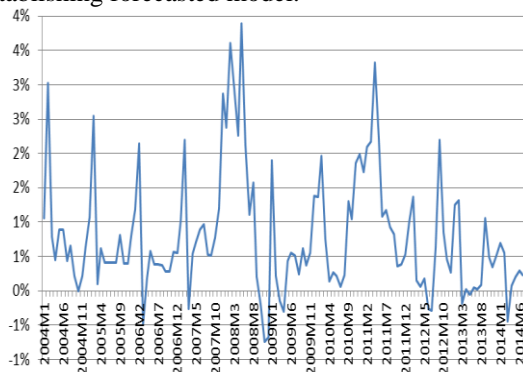
$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| * 100\% \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2} \quad (6)$$

Where,  $Y_i$  denotes actual data at month  $i$ ,  $\hat{Y}_i$  is forecasted value at month  $i$ ,  $\bar{Y}_i$  is the average value of the  $i$ th sample,  $n$  is number of the forecasted data.

**3. Experimental Results Based on Forecasted Model Selection**

In this study, the monthly inflation data of VietNam from January 2004 to July 2014 is selected for forecasting model and shown in Figure 1, in which it taken from the site [www.gso.gov.vn](http://www.gso.gov.vn). A total of 127 observations. All these observations are used for establishing forecasted model.



**Figure 1:** Source GSO: Inflation rate data of VietNam

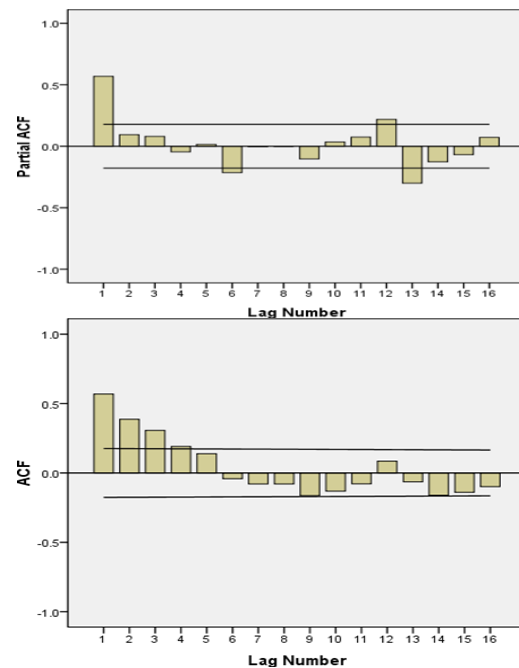
**3.1. Building ARIMA model**

In order to build the ARIMA model, we first have to test the continuity of the inflation chain. The results of Dickey-Fuller (ADF) and Phillips-Perron (PP) tests show that the t-stat value is smaller than the critical value, so we can conclude that the inflation chain stops at level 0 or I ( $d = 0$ ).

**Table 1:** Check for stationarity of inflation data

Test	t-stat value	Probability
ADF	-5,803	0,000
PP	-5,836	0,000

To determine the  $p, q$  value of the ARIMA model, we must rely on the ACF and PACF histograms as shown in Figure 2. From the PACF diagram shown that the correlation coefficients are not in the lags of 1,6, 12 and 13. As for the ACF chart, we have different partial correlation coefficients Latency 1,2,3 and 4. Temporarily, we determine the maximum latency of  $p$  is 13 and the highest latency of  $q$  is 4.



**Figure 2.** Chart PACF and ACF of inflation series

**Table 2:** The regression results of the ARIMA (1.0.3) model

Variab le	coefficient	Standar d data	t-Stat value	Probabili ty
C	0,007934	0,001623	4,894209	0,0000
AR(1)	0,526510	0,078454	6,610916	0,0000
MA(3)	0,191265	0,090037	2,124258	0,0362

To find out the most proper forecasted model, we must use the experimental method by comparing the correlated  $R^2$ , AIC and Schwarz. Comparison results show that the ARIMA (1,0,3) is the best suitable for the dataset given. The regression results of the selected model are shown in Table 2. Furthermore, by considering the ARIMA (3,0,1) model has error of the

assumptions of the regression model, we try on handle some more examines as follows:

The White test shows that the model has no variance in variance. The Breusch-Godfrey test show that has no autocorrelation error. We can conclude that the above model is very proper for forecasting. Table 3 shown that forecasting result of models from February 2004 to July 2014 and get the RMSE value is 0.0068 and the MAPE value of 0.49%.

### 3.2. Developing the SARIMA model

The Box-Jenkins ARIMA model is generalized into a Seasonal Autoregressive Integrated Moving Average (SARIMA) model that accounts for both seasonal and non-seasonal characterized data. The SARIMA model is derived from the ARIMA model described above and also uses information on past observations and past errors of the series. Since the ARIMA model is inefficient for those series with both seasonal and non-seasonal behavior for example in terms of wrong order selection, the SARIMA model is preferred when any seasonal behavior is suspected in the series.

Therefore, we will add seasonal factors when making predictions using the ARIMA model and called SARIMA. The SARIMA model also sometimes referred to as the Multiplicative Seasonal Autoregressive Integrated Moving Average model, is denoted as ARIMA(p,d,q) (P,D,Q)S.

In order to construct the SARIMA model, we first need to create 11 dummy variables (named D1 to D11, respectively) to represent 12 months of the year. Next, we will regress inflation by a constant and 11 dummy variables, then save the residual number in Table 4. This residual part will be further processed by the ARIMA model procedures. Then, the ARIMA model will be used to forecast the surplus. Finally, the residual part is included in the initial regression equation for predicting inflation. The data processing results show that the ARIMA model (1,0,5) corresponds to the existing residual part shown in Table 5. Combined with the seasonality of inflation, this model has named SARIMA (1,0,5). The detailed forecast of the SARIMA model (1,0,5) is shown in Table 3.

**Table 3:** The inflation forecasting results of three forecasting models

Date	Inflation rate	ARIMA	SARIMA	SCARIMA
2004M1	1.06%			
2004M2	3.02%	1.01%	1.90%	1.79%
2004M3	0.79%	1.92%	1.11%	0.76%
2004M4	0.45%	0.78%	1.05%	0.86%

**Table 4:** The regression of inflation by a constant

Dependent Variable: RESID1
Method: Least Squares

Sample (adjusted): 2004M02 2014M07				
Included observations: 126 after adjustments				
Convergence achieved after 8 iterations				
MA Back cast: 2003M09 2004M01				
Date	Inflation rate	ARIMA	SARIMA	SCARIMA
C	0.008870	0.002571	3.450429	0.0008
D1	0.004441	0.003552	1.250366	0.2137
D2	0.010240	0.003552	2.882943	0.0047
D3	-0.003981	0.003552	-1.120907	0.2647
D4	-0.002172	0.003552	-0.611622	0.5420
D5	-2.95E-05	0.003552	-0.008292	0.9934
D6	-0.002845	0.003552	-0.801062	0.4247
D7	-0.004317	0.003552	-1.215425	0.2267
D8	-0.003220	0.003635	-0.885802	0.3776
D9	-0.000827	0.003635	-0.227384	0.8205
D10	-0.004495	0.003635	-1.236387	0.2188
D11	-0.003634	0.003635	-0.999502	0.3196
R-S	0.228049	Mean dependent var		0.008027
AR	0.154211	S.D. dependent var		0.008839
SER	0.008129	Akaike info criterion		-6.697018
SS	0.007599	Schwarz criterion		-6.428276
LL	437.2607	Hannan-Quinn criter.		-6.587832
F-S	3.088479	Durbin-Watson stat		0.668798
Prob	0.001145			

Note: R-s(R-squared), AR (Adjusted R-squared), SER (S.E. of regression), SS (Sum squared residue), LL (Log likelihood), FS (F-statistic).

**Table 5:** The forecasting result of ARIMA model (1,0,5) corresponds to the existing residual number

Dependent Variable: RESID1				
Method: Least Squares				
Sample (adjusted): 2004M02 2014M07				
Included observations: 126 after adjustments				
Convergence achieved after 8 iterations				
MA Back cast: 2003M09 2004M01				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.97E-05	0.001680	0.053420	0.9575
AR(1)	0.571480	0.079739	7.166841	0.0000
MA(2)	0.211210	0.091499	2.308340	0.0227
MA(5)	0.229332	0.086747	2.643688	0.0093
R-S	0.487654	Mean dependent var		2.18E-05
AR	0.475056	S.D. dependent var		0.007793
SER	0.005646	Akaike info criterion		-7.484355
SS	0.003890	Schwarz criterion		-7.394314
LL	475.5143	Hannan-Quinn criter.		-7.447774
F-S	38.70680	Durbin-Watson stat		2.013319
Prob	0.000000			

### 3.3. SCARIMA model

Apart from the seasonal factor, Vietnam's inflation has had two major fluctuations in 2008 and 2011. We will consider more the impact of these fluctuations on the results of the forecast. By considering the impact of these fluctuations, we will first regress inflation by 11 dummy variables which representing the crop and a new dummy variable (named as K variable, which



has a value of 1 for the months from December 2007 to August, 2008 and from March 2011 to September 2011, with a value of 0 in other months) and representing an abnormal fluctuation of inflation and a residual number are listed in Table 6. This residual number will be processed according to the procedures of the ARIMA model. The regression results show that the ARIMA model (1,0,2) is consistent with the residuals Table 7. Combined with seasonality and abnormal fluctuations of inflation, the author named.

**Table 6:** The extraordinary fluctuations of inflation for SCARIMA model

Dependent Variable: RESID1 Method: Least Squares Sample (adjusted): 2004M02 2014M07 Included observations: 126 after adjustments Convergence achieved after 8 iterations MA Back cast: 2003M09 2004M01				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006685	0.002247	2.975598	0.0036
D1	0.005633	0.003073	1.832861	0.0694
D2	0.011431	0.003073	3.719706	0.0003
D3	-0.003783	0.003068	-1.233103	0.2201
D4	-0.001974	0.003068	-0.643426	0.5212
D6	-0.002647	0.003068	-0.862769	0.3901
D7	-0.004118	0.003068	-1.342542	0.1821
D8	-0.003220	0.003140	-1.025683	0.3072
D9	-0.000827	0.003140	-0.263292	0.7928
D10	-0.003402	0.003144	-1.082082	0.2815
D11	-0.002541	0.003144	-0.808202	0.4207
K	0.010923	0.001723	5.989412	0.0000
R-S	0.429253	Mean dependent var	0.008027	
AR	0.369175	S.D. dependent var	0.008839	
SER	0.007020	Akaike info criterion	-	
SS	0.005619	Schwarz criterion	6.983245	
LL	456.4361	Hannan-Quinn criter.	-	
F-S	7.144864	Durbin-Watson stat	6.692108	
Prob	0.000000		6.864960	
			0.949976	

**Table 7:** The forecasting result of SCARIMA model (1,0,2) corresponds to the existing residual number

Dependent Variable: RESID1 Method: Least Squares Sample (adjusted): 2004M02 2014M07 Included observations: 126 after adjustments Convergence achieved after 8 iterations MA Back cast: 2003M09 2004M01				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.23E-05	0.000810	0.052191	0.9585
AR(1)	0.369473	0.088551	4.172420	0.0001

MA(2)	0.125747	0.094458	1.331249	0.1856
R-S	0.182194	Mean dependent var		9.85E-06
AR	0.168896	S.D. dependent var		0.005598
SER	0.005104	Akaike info criterion		-7.694182
SS	0.003204	Schwarz criterion		-7.626651
LL	487.7334	Hannan-Quinn criter.		-7.666746
F-S	13.70117	Durbin-Watson stat		1.949468
Prob	0.000004			

### 3.4. Comparison of forecasting efficiency of selected models.

In this section, it is assumed that all three models have passed the diagnostic checks leading to three models being considered fit to use for forecasting inflation. To know which model is better, we compare forecasting accuracy of models based on three RMSE, MAPE and  $R^2$  criteria (squared correlation coefficient between actual value and predicted value). A comparison of the forecasting results among three models is shown in Table 8.

**Table 8:** A comparison of the accuracy of the forecasted models based on different evaluation criterions

Forecasted models	Evaluation criterions		
	RMSE	MAPE	$R^2$
ARIMA(1,0,3)	0,0068	0,49%	0,3415
SARIMA(1,0,5)	0,0053	0,38%	0,6035
SCARIMA(1,0,2)	<b>0,0051</b>	<b>0,34%</b>	<b>0,6634</b>

From forecasted results in Table 3 and Table 8 shown that the SCARIMA (1,0,2) model provides the best forecasted result with the smallest RMSE, MAPE value and the largest  $R^2$  value among the three compared models. Therefore, we will apply this model to the forecast outside the sample.

### 3.5. Forecast the out of sample

For forecasting out of sample, the SCARIMA model (1,0,2) can predict a period ahead. For example, to forecast a new inflation value in August 2014, the inflation data in July 2014 used as the training data. Getting forecasted value in August 2014, we continue for forecasting inflation in September. Similarly, for forecasting inflation for October, 11 and 12. Detailed forecast results are given in the following Table 9.

**Table 9:** The forecasted results of outside sample of the SCARIMA model (1,0,2)

Months	8/2014	9/2014	10/2014	11/2014	12/2014
Forecasted inflation	0,37%	0,60%	0,36%	0,41%	0,65%

Based on forecasted outside sample of the model, the average inflation rate in the latter half of 2014 tends to be higher than in the former half of 2014. For the whole of 2014, the SCARIMA (1, 0, 2) model give average inflation of **5.98%**.

#### 4. Conclusions

This paper discussed the applicability of ARIMA, SARIMA, SCARIMA models in modeling to find the best model for forecasting inflation in Vietnam. All models were proven to be quite well-fitted according to the results of the diagnostic checks. More exactly, the experimental results show that the SCARIMA (1,0,2) model provides the best forecasting result among three forecasting models compared. In addition, forecasting total inflation in year 2014 shows that inflation index is 5,89% smaller than the National Assembly's target of 7%. However, inflation in the last six months of year 2014 tends to be higher than the inflation rate in the former six months of year 2014. Inflation rate in May and June which are all increased, the latter month higher than in the former month. So, the inflation rate does not exceed the 7% set by Congress for 2014, policy makers need to combine flexible monetary policy and tight fiscal policy. Through, this forecasted results will help policy makers gain insight into more appropriate economic and monetary policy in other to combat the predicted rise in inflation rates for following years.

#### 5. References

- [9] Ezekiel N.N. Nortey, Benedict Mbeah-Baiden, Julius B. Dasah and Felix O. Mettle (2014). *Modelling Rates of Inflation in Ghana: An Application of Arch Models*, *Current Research Journal of Economic Theory*, 6(2): 16-21.
- [1] Box, G. E. P and Jenkins, G.M., (1976). "Time series analysis: „Forecasting and control," Holden-Day, San Francisco.
- [2] Gujarati, (2004). *Basic Econometrics*. McGraw-Hill.
- [3] Khashei & Bijari, (2011). A novel hybridization of artificial neural networks and ARIMA
- [4] Stock & Watson, (2007). *Why Has U.S. Inflation Become Harder to Forecast?* *Journal of Money, Credit and Banking*, Vol 39, pp.3-33
- [5] Muhammad Abdus Salam, Shazia Salam and Mete Feridun (2006), *Forecasting inflation in developing nations: The case of Pakistan*, *International Research Journal of Finance and Economics*, issue 3, pp. 138 – 159.
- [6] Samuel Erasmus Alnaa and Ferdinand Ahiakpor, (2011), *ARIMA approach to predicting inflation in Ghana*, *Journal of Economics and International Finance*, vol. 3(5), pp. 328 – 336.
- [7] M.Z. Babai, M.M. Ali and J.E. Boylan, (2011). "Forecasting and inventory performance in a two-stage supply chain with ARIMA (0,1,1) demand: Theory and empirical analysis," *International Journal of Production Economics*.
- [8] Chi-Chen Wang ,(2011). *A comparison study between fuzzy time series model and ARIMA model*. Department of Financial Management National Defense University. *Expert Systems with Applications* 38, pp. 9296- 9304.