



Forecasting Inflation Rate in the United States of America

NagaShrikanth Ammanabrolu

Nikhil Soman

Chandan Ramakrishna

November 29, 2015

Forecasting Inflation Rate in the United States of America

November 29, 2011

Introduction

1.1 Abstract

Inflation rate in the United States and many other countries worldwide is a measurement of the increase or decrease in overall consumer good prices. The inflation rate is influenced by different factors including Consumer Price Index (CPI), Producer Price Index (PPI), Tax rate, GDP, and many other factors. The inflation rate in the United States, as defined by the Bureau of Labor Statistics of the United States Department of Labor, is '*a measure of a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services*'.

Our approach to the analysis of predicting Inflation rate in the United States is based on the ARMA modeling of the CPI index, collected for a span of 25 years. The modeling process allows us to closely model the past data and adequately represent the future inflation rate. We are able to achieve a decent level of predicting future inflation rates and given the numerous other factors influencing Inflation rate as discussed before, an exact forecasting model is unfathomable.

1.2 Inflation Rate and CPI

Inflation Rate is primarily defined as a sustained increase in the general level of prices for goods and services. It is measured as an annual percentage increase. The value of a dollar is observed in terms of purchasing power. When inflation rises, there is a decline in the purchasing power of money.

Although there happens to be no singular reason that's universally agreed upon, but at least two theories are generally accepted:

Demand-Pull Inflation - This theory can be summarized as "too much money chasing too few goods". In other words, if demand is growing faster than supply, prices will increase. This usually occurs in growing economies.

Cost-Push Inflation - When companies' costs go up, they need to increase prices to maintain their profit margins. Increased costs can include things such as wages, taxes, or increased costs of imports.

In the case of the United States of America, there are two main price indexes that measure inflation:

Consumer Price Index (CPI) - A measure of price changes in consumer goods and services such as gasoline, food, clothing and automobiles. The CPI measures price change from the perspective of the purchaser. U.S. CPI data can be found at the Bureau of Labor Statistics.

Producer Price Indexes (PPI) - A family of indexes that measure the average change over time in selling prices by domestic producers of goods and services. PPIs measure price change from the perspective of the seller. U.S. PPI data can be found at the Bureau of Labor Statistics.

1.3 Project Objectives

The primary objective of this project was to make use of the readily available Inflation rate in the United States, given in terms of CPI and use it to forecast the Inflation rate of the next 12 months. Another objective of this project was to understand how CPI influences the inflation rate and by what factor. The correlation between CPI and Inflation rate and its modeling helps us develop a better model to predict the values of the Inflation rate in the future.

1.4 Data Collection and Cleaning

The primary source of data for this project was obtained from the US Inflation Calculator, who lists the official historical Inflation Rate in the United States data set from the Bureau of Labor Statistics. This data set is readily available for public access and is listed from the onset of 1914. For our part of the project, we have extracted the Inflation rate in the United States, starting from the January of 1990. Since the Inflation rates back in the early 1900's does not adequately represent the inflation rate for predicting the inflation rate in the year 2015, we've cleaned the data and extracted parts of it required.

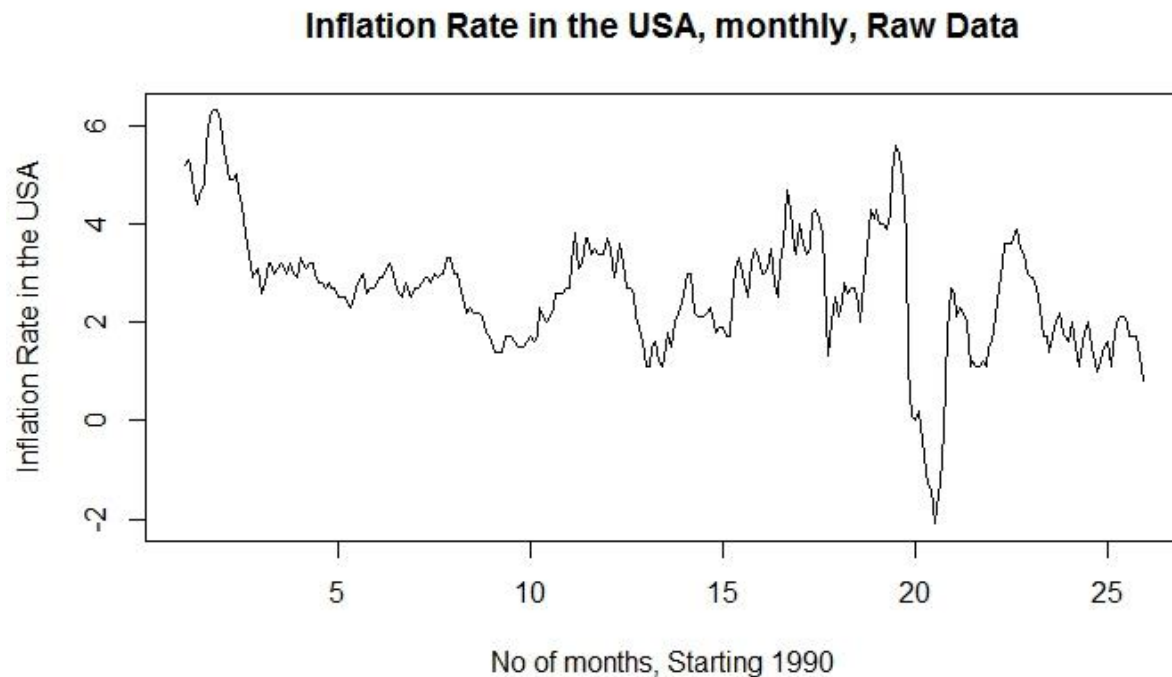
The data set used for this project is collected on a monthly basis, between the periods of January 1990 to December 2014. The entire data set is a time series of the Inflation rate in the United States and for de-trending the series, the mean subtracted data has been obtained and utilized over the entire course of the project.

System Analysis

The time series data obtained as mentioned earlier is a monthly data of the Inflation rate in the United States for a period of 25 years. De-trending of this time series data is achieved by subtracting the mean from the entire data set. This mean subtracted, $X = (\hat{X}(t) - \bar{X})$ data is used for further analysis of the time series dataset of the Inflation rate in the United States.

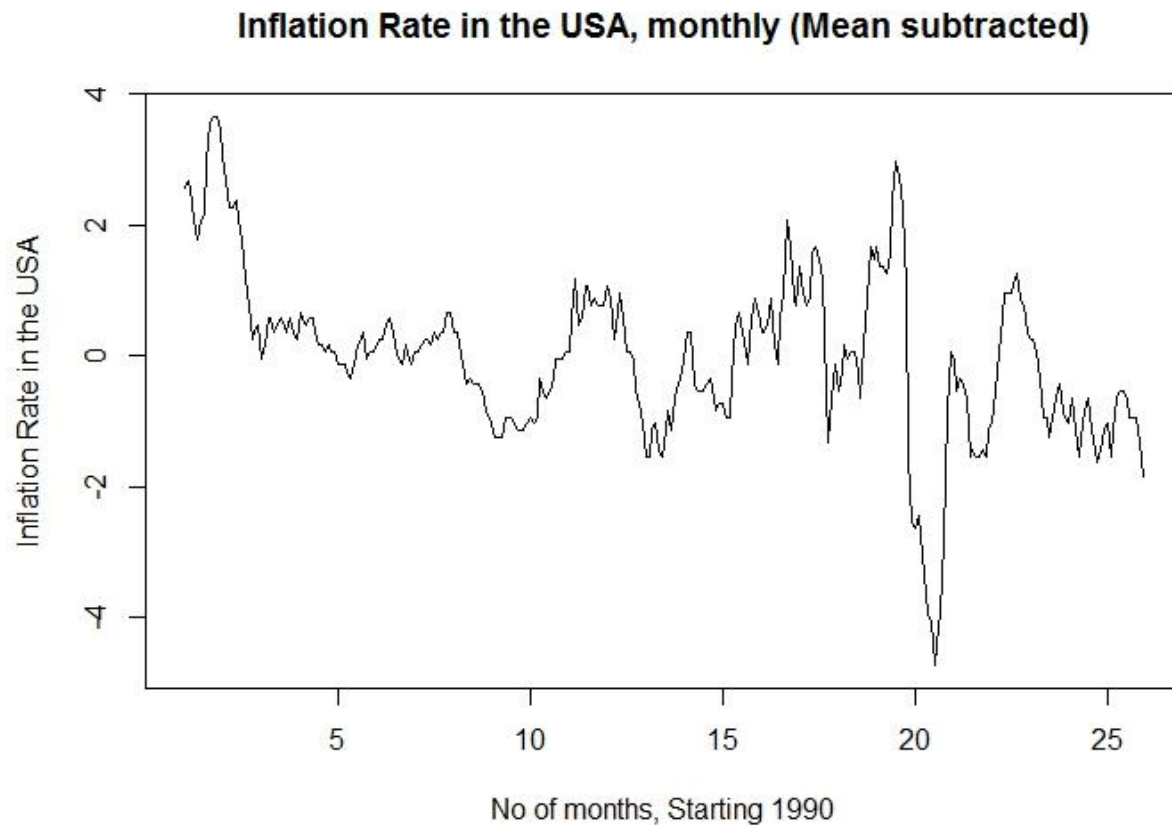
The following equation pertaining to forecasting a time series data has been used over the entire course of this project,

$$Y(t) = F(t) + X(t)$$



We begin the system analysis of this particular time series dataset by obtaining an adequate trend model by trend analysis. Later, an optimal trend equation is reached at through the implementation of F-Test between different trend models. After the trend analysis, the residuals obtained from such an analysis are used for ARMA modeling and again F-Test is used for adequacy checks of the model. Once the time series data is adequately represented by the $F(t)$ or trend parameter and the $X(t)$ or the residuals parameter, one-step ahead forecasting is used to predict the inflation rate in the United States for the subsequent 12 months.

2.1 Trend Analysis and F-Test



The above seen plot is the mean subtracted Inflation rate in the USA dataset. Once the dataset is ready to be analyzed, the first step is to obtain an adequate trend equation which closely represents the raw data. The two primary trend analysis plots we considered for this dataset was the Linear and Polynomial trends of different orders. On comparing the data plot with different other trends like Quadratic and Exponential, only Linear and Polynomial trend lines were able to adequately represent the raw dataset.

Through the use of R programming language, Excel and F-Test calculations an adequate trend equation was developed for further analysis of the dataset.

Adequacy check for Trend fitting

	Linear	Polynomial order 2	Polynomial Order 3	Polynomial Order 4	Polynomial Order 5
RSS	388.381	378.047	321.659	308.254	308.161
F ~ (1, ∞)	3.84	3.84	3.84	3.84	3.84

F-Test between Linear and Polynomial of Order 2

$$F = (A_1 - A_0 / S) / (A_0 / N - r)$$

F = 8.11, which is greater than F (1, ∞) = 3.84.

Therefore, Polynomial of Order 2 is significant and we continue further.

F-Test between Polynomial of Order 2 and 3

$$F = (A_1 - A_0 / S) / (A_0 / N - r)$$

F = 51.88, which is greater than F (1, ∞) = 3.84.

Therefore, Polynomial of Order 3 is significant and we continue further.

F-Test between Polynomial of Order 3 and 4

$$F = (A_1 - A_0 / S) / (A_0 / N - r)$$

F = 12.82, which is greater than F (1, ∞) = 3.84.

Therefore, Polynomial of Order 4 is significant and we continue further.

F-Test between Polynomial of Order 4 and 5

$$F = (A_1 - A_0 / S) / (A_0 / N - r)$$

F = 0.0887, which is less than F (1, ∞) = 3.84.

Therefore, Polynomial of Order 4 is significant and we conclude the process.

Output of R for Polynomial of Order 4

```
lm(formula = inflation ~ poly(months, 4, raw = TRUE))
```

Residuals:

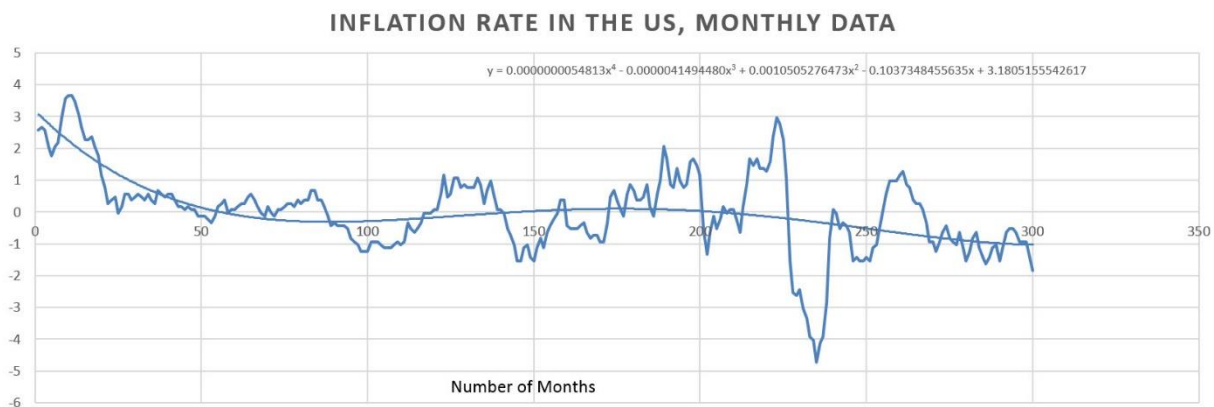
Min	1Q	Median	3Q	Max
-4.4156	-0.5268	0.0000	0.5642	3.1396

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.181e+00	3.011e-01	10.563	< 2e-16	***
poly(months, 4, raw = TRUE)1	-1.037e-01	1.381e-02	-7.511	7.04e-13	***
poly(months, 4, raw = TRUE)2	1.051e-03	1.862e-04	5.643	3.92e-08	***
poly(months, 4, raw = TRUE)3	-4.149e-06	9.285e-07	-4.469	1.12e-05	***
poly(months, 4, raw = TRUE)4	5.481e-09	1.530e-09	3.582	0.000399	***

Residual standard error: 1.022 on 295 degrees of freedom
Multiple R-squared: 0.3515, Adjusted R-squared: 0.3427
F-statistic: 39.97 on 4 and 295 DF, p-value: < 2.2e-16

The next step is to fit the Polynomial of Order 4 to our mean subtracted data and this is achieved using Excel and seen below.



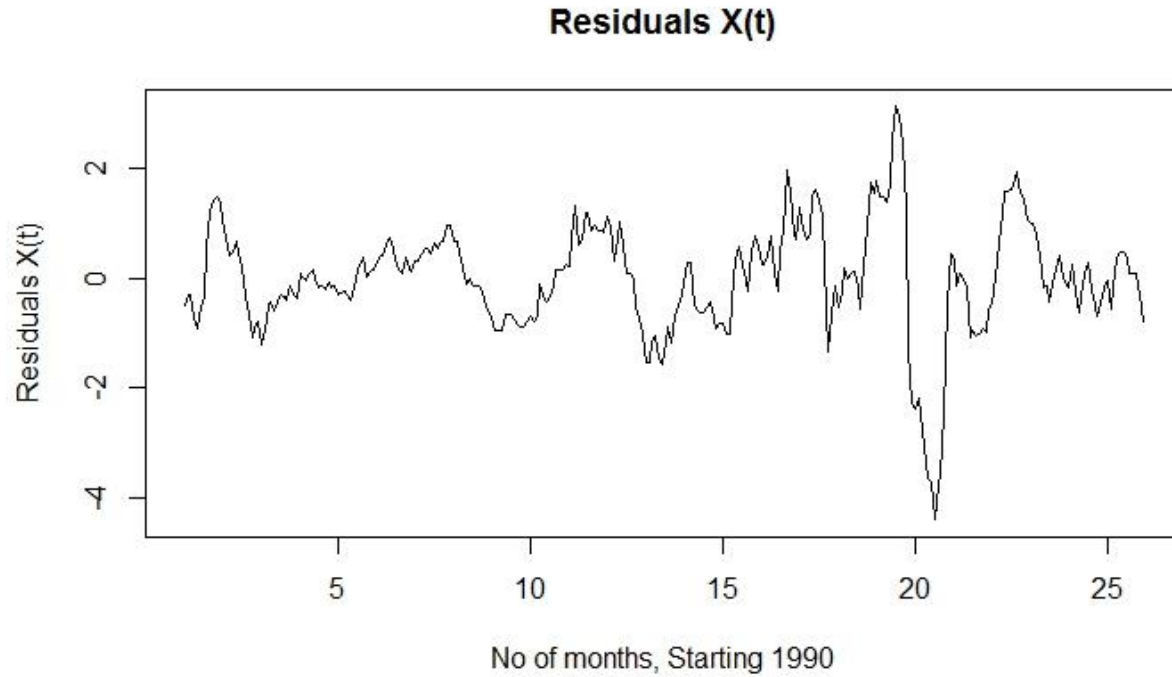
The adequate trend equation for this raw dataset is extracted from Excel and is described by the equation as follows:

$$y = 0.0000000054813x^4 - 0.0000041494480x^3 + 0.0010505276473x^2 - 0.1037348455635x + 3.1805155542617$$

This polynomial equation of Order 4 is fitted to the mean subtracted dataset and thus the values of $F(t)$ are obtained. As seen earlier, since $Y(t) = F(t) + X(t)$, the values of the residuals are obtained from this equation and the values of $F(t)$ obtained in this part of the process.

The values obtained from the trend equation are subtracted from the mean subtracted dataset and thus we obtain the residuals $X(t)$ needed for further analysis and forecasting of the data.

2.2 Residuals and ARMA Modeling



The residuals $X(t)$ obtained from the previous step is now used for ARMA modeling and analysis. The ARMA modeling approach used in this project is that of an ARMA $(n, n-1)$ nature. Since the observations are not in large numbers, the ARMA $(2n, 2n-1)$ approach is not called upon.

According to the ARMA $(n, n-1)$ modeling process, the different Φ , Θ and the Residual Sum of Squares (RSS) are calculated using R and are tabulated below.

Parameters	ARMA Order					
	(1,0)	(2,1)	(3,2)	(4,3)	(5,4)	(6,5)
Φ_1	0.9179	1.1117	-0.2721	0.4264	-0.1858	0.3089
Φ_2		-0.2433	0.5862	-0.209	-0.1927	0.7116
Φ_3			0.3843	0.7727	0.1465	0.5201
Φ_4				-0.2954	0.3567	-0.0584
Φ_5					0.2777	-0.6679
Φ_6						0.1087
Θ_1		0.2486	1.6425	0.9756	1.5911	1.0633
Θ_2			0.9180	1.0036	1.8426	0.0989
Θ_3				0.0037	1.5948	-0.9739
Θ_4					0.9242	-1.0870
Θ_5						-0.1012
RSS	47.577	38.365	36.193	34.856	31.465	31.66

Now, from the tabulated ARMA Modeling process data, using F-Test, an adequate ARMA model for the data is reached upon. The calculation process for the F-test process of ARMA modeling is illustrated below.

F-Test between AR (1) and ARMA (2,1)

$$F = (A_1 - A_0 / S) / (A_0 / N - r)$$

$F = 35.536$, which is greater than $F(2, \infty) = 3$.

Therefore, ARMA (2,1) is significant and we continue further.

F-Test between ARMA (2,1) and ARMA (3,2)

$$F = (A_1 - A_0 / S) / (A_0 / N - r)$$

$F = 8.821$, which is greater than $F(2, \infty) = 3$.

Therefore, ARMA (3,2) is significant and we continue further.

F-Test between ARMA (3,2) and ARMA (4,3)

$$F = (A_1 - A_0 / S) / (A_0 / N - r)$$

$F = 5.6$, which is greater than $F(2, \infty) = 3$.

Therefore, ARMA (4,3) is significant and we continue further.

F-Test between ARMA (4,3) and ARMA (5,4)

$$F = (A_1 - A_0 / S) / (A_0 / N - r)$$

$F = 15.626$, which is greater than $F(2, \infty) = 3$.

Therefore, ARMA (5,4) is significant and we continue further.

F-Test between ARMA (5,4) and ARMA (6,5)

$$F = (A_1 - A_0 / S) / (A_0 / N - r)$$

$F = 0.924$, which is less than $F(2, \infty) = 3$.

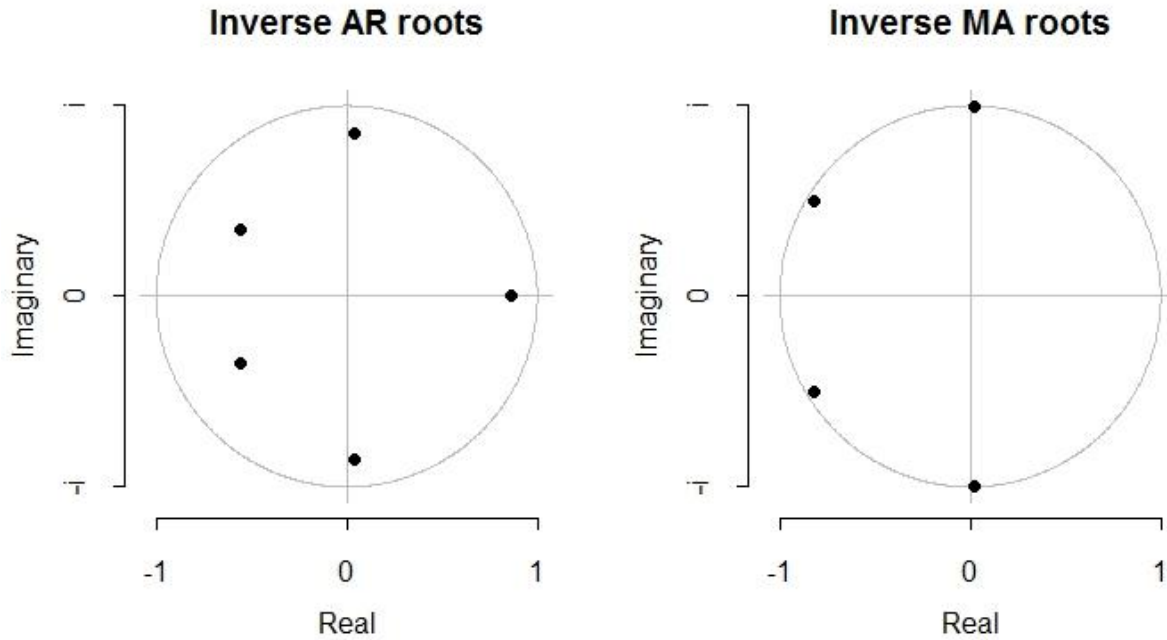
Therefore, ARMA (5,4) is significant and we conclude the process.

Characteristic Roots and checking for a Parsimonious ARMA model

Now that the adequate ARMA model of ARMA (5,4) nature is obtained, the next process is to obtain the characteristic roots described by the ARMA (5,4) equation given below.

$$\lambda^5 - \phi_1 \lambda^4 - \phi_2 \lambda^3 - \phi_3 \lambda^2 - \phi_4 \lambda^1 - \phi_5 = 0$$

The ϕ values are obtained from the ARMA modeling results tabulated above and the characteristic roots of this equation are obtained as described below.



The characteristic roots of the above mentioned ARMA model equation are calculated to be as follows:

$$\lambda_1 = 0.8593$$

$$\lambda_2, \lambda_3 = -0.560066 \pm 0.354404i$$

$$\lambda_4, \lambda_5 = 0.0374667 \pm 0.856854i$$

From the above Real roots of the characteristic equation, we see that the value of the Real roots do not tend to one and hence there is no possible Trend for obtaining a Parsimonious ARMA model.

Similarly, since the complex conjugate pairs of the roots of the characteristic equation do not tend to one, there is no possible Seasonality for obtaining a Parsimonious ARMA model.

Hence, we proceed with the ARMA (5,4) model as described before and now we obtain the one-step ahead forecasts of the obtained $X(t)$'s.

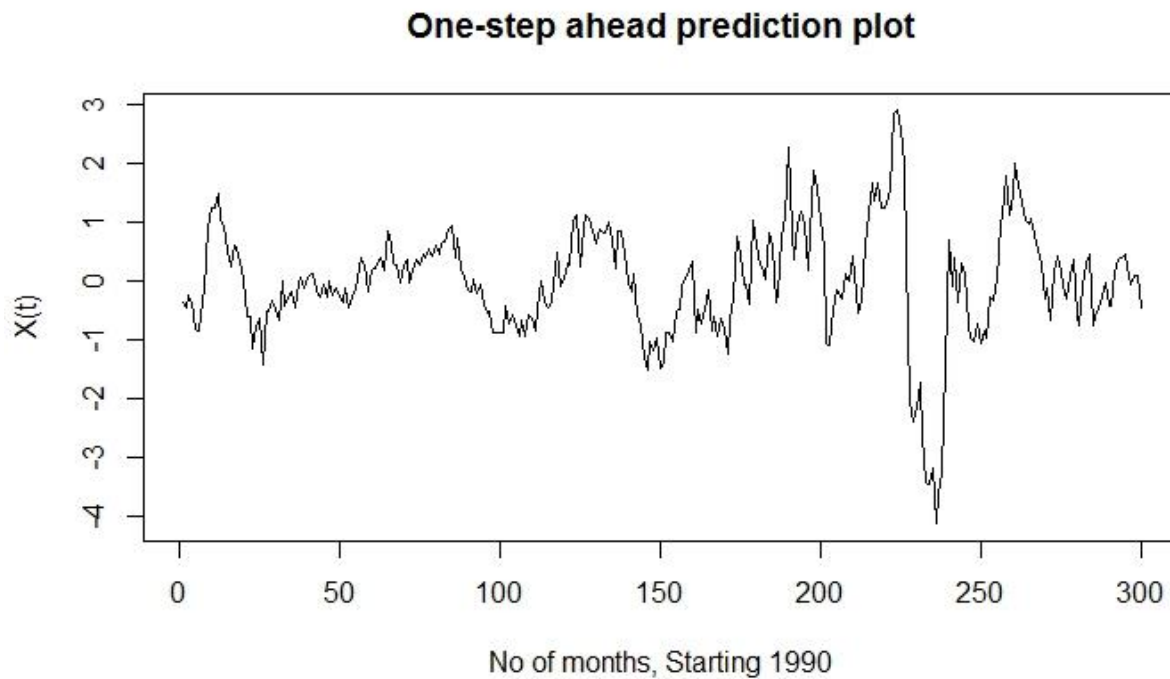
2.3 One-Step Ahead prediction

After the $X(t)$ values of the adequate ARMA model is obtained, the next process is to obtain a one-step ahead prediction of the dataset to forecast further values. The $X(t)$ at one-step ahead prediction or $\hat{X}_t(1)$ is obtained from the equation described below.

$$\hat{X}_t(1) = \phi_1 X_t + \phi_2 X_{t-1} + \phi_3 X_{t-2} + \phi_4 X_{t-3} + \phi_5 X_{t-4} - \theta_1 a_t - \theta_2 a_{t-1} - \theta_3 a_{t-2} - \theta_4 a_{t-3} + a_{t+1}$$

The one-step ahead predicted values for the values of $X(t)$ at time X_{t-1} are obtained from the above mentioned equation.

This one-step ahead prediction could similarly be adequately achieved by using the *fitted.Arima* function of the R's Forecast package. The *fitted.Arima* function is used to obtain one-step ahead forecasts for the data which was used in fitting the ARMA model.



The values obtained from the R's *fitted.Arima* function are plotted over time as seen above. These values are now exported and are to be used for further forecasts. The one-step ahead forecasts obtained in this step are used to obtain the values of the predicted $Y(t)$ values and for predicting the Inflation rate for the next 12 months.

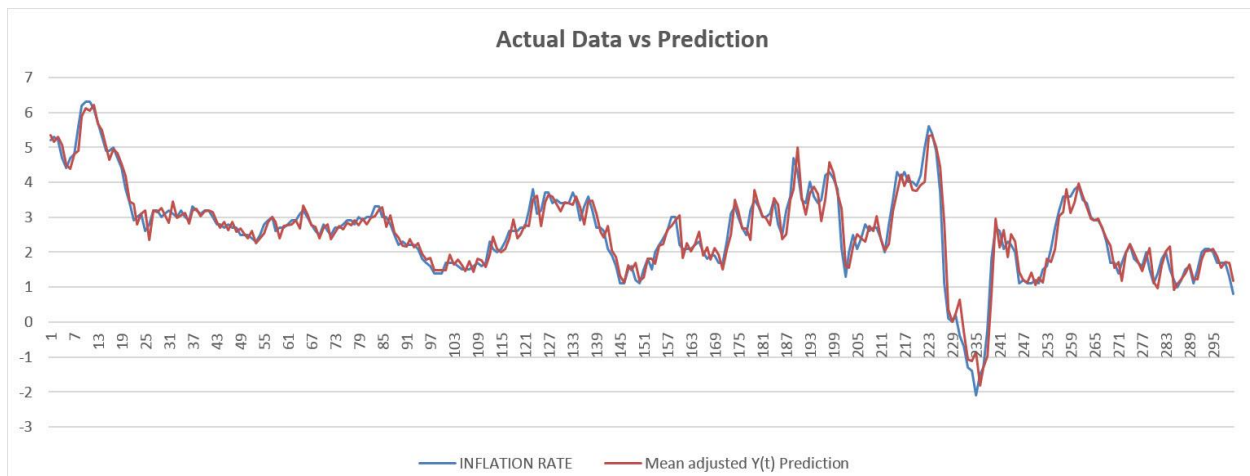
2.4 Forecasting and Plotting the Graphs

After obtaining the $\hat{X}_t(1)$ values, the next process is to plot the forecasted values of the Inflation rate for the next 12 months.

$$Y(t) = F(t) + \hat{X}_t(1)$$

Now from the above equation, we make use of the $\hat{X}_t(1)$ values and add the trend equation values to it for obtaining the mean adjusted $Y(t)$ predicted values. Here, the mean is added back to the predicted dataset for comparing it with the raw data.

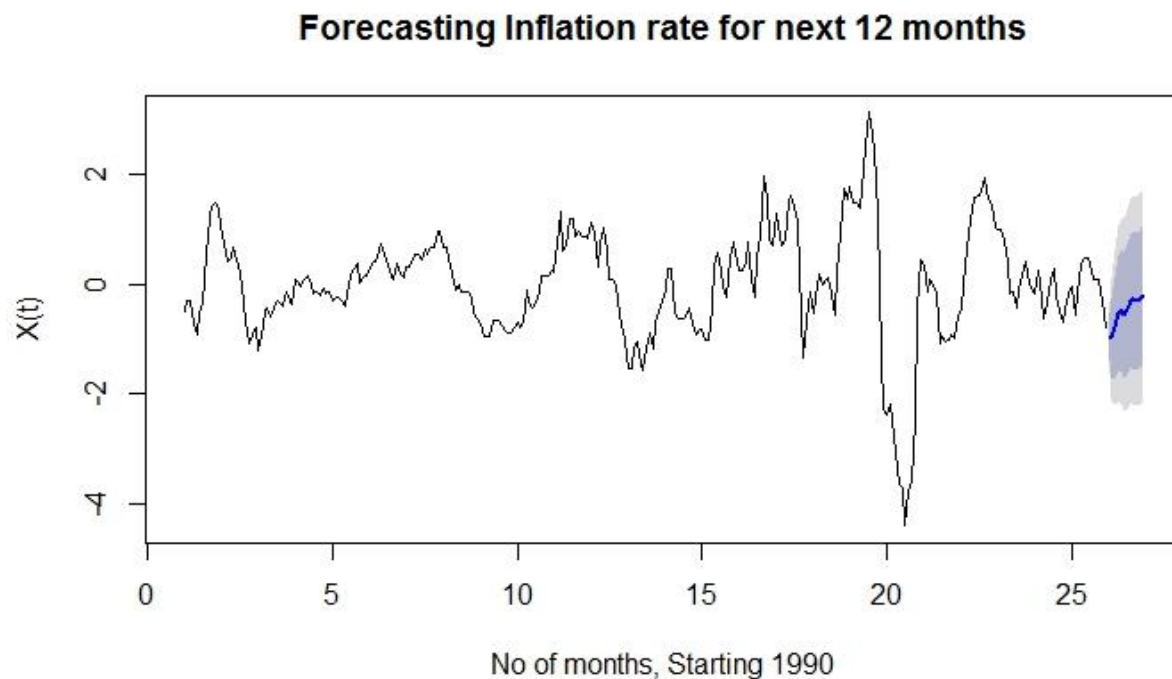
The $Y(t)$ predicted values thus obtained are now used for comparing with the actual raw data. The subsequent plot is shown below, which represents an adequate model fitting of the predicted data compared with the raw data.



The close representation of the predicted values of $Y(t)$ with the actual values of the raw data illustrate the adequacy of the ARMA model selected in the previous mentioned steps.

Now, using the R's Forecast package, we are able to forecast the values of the Inflation rate for the next 12 months. Based on the results predicted by the adequate ARMA modeling, R's Forecast function allows us to predict the Inflation rate of the next 12 months between the 95% and 65% confidence intervals, as seen below.

Inflation (Next 12 months)
0.645586163
0.60711369
0.800029497
1.070046521
1.12325667
1.035480845
1.142115452
1.300218529
1.335171203
1.301532768
1.3425636
1.437709248

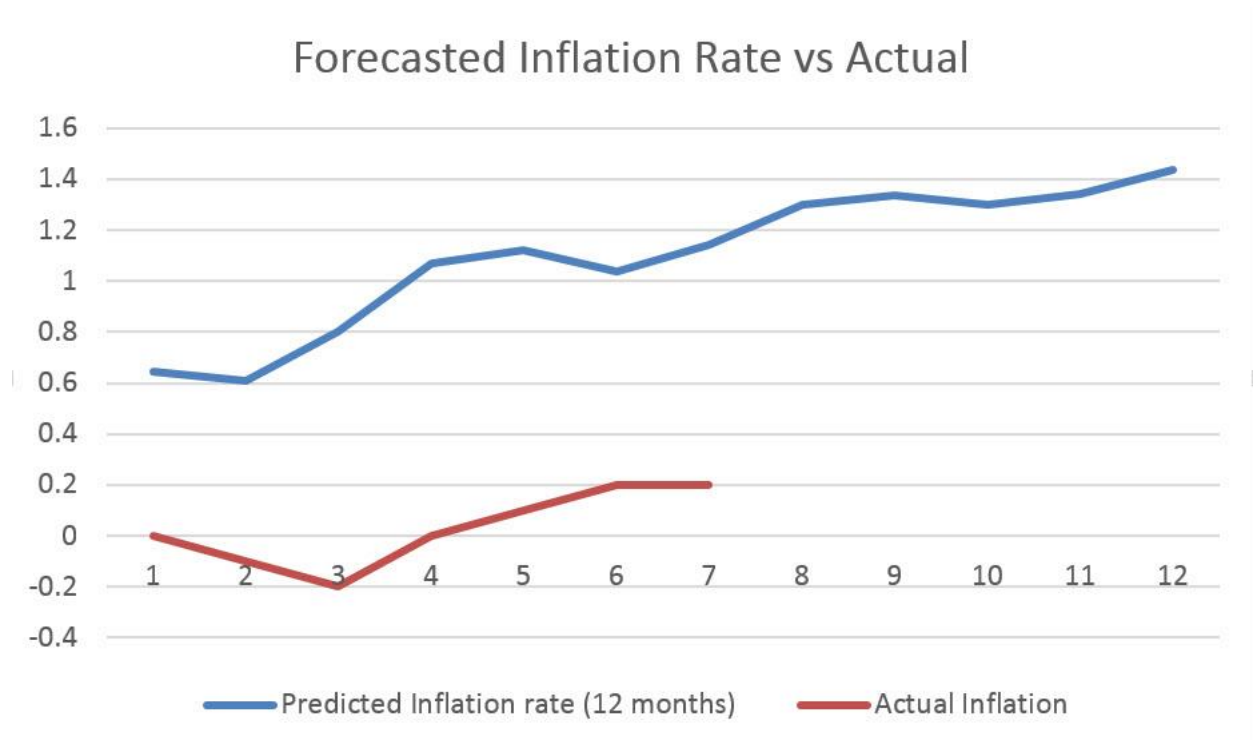


Interpreting the Results

From the trend fitting, ARMA modeling and the forecasting, we obtain the below set of Inflation rate values for the next 12 months. Now, this predicted data is compared with the actual Inflation rate obtained for the first quarter of 2015 by the Bureau of Labor Statistics.

Inflation (Next 12 months)	Actual Inflation Rate
0.645586163	-0.1
0.60711369	0
0.800029497	-0.1
1.070046521	-0.2
1.12325667	0.1
1.035480845	0.2
1.142115452	0.2
1.300218529	-
1.335171203	-
1.301532768	-
1.3425636	-
1.437709248	-

This predicted dataset for the next 12 months is now plotted against the actual data for the first quarter of 2015.



From the graph and the predicted values of the Inflation rate of the United States, we can illustrate the stochastic nature of this dataset. Given the number of parameters that the Inflation rate is correlated with, the CPI alone is just a decent indicator of the future Inflation rate. Also, since the present values of the Inflation rate are independent of the previous values of CPI, the stochastic nature of the dataset is bound for some uncertainties.

On examining the abrupt drop of the inflation rate in the United States in the first quarter of 2015, different other factors come into the picture. A slack economy, drop in oil prices and other energy products in past year, rise of the value of dollar over the past year and other reasons are some of the primary factors for the drastic drop in Inflation rate in the United States in the first quarter of 2015.

By examining our forecasts, we can observe the adequate representation of the trend line, but considering the above mentioned recent conditions, the stochastic nature of the dataset is visible in plain sight. The abrupt drop in inflation rate during the first half of 2015 have caused our predictions to be bit on the higher end from the actual observed data.

Conclusion

We can conclude that our stochastic modeling process has been able to adequately analyze and forecast the Inflation Rate in the United States. Considering the CPI as a factor for this change in the Inflation rate in the United States, following the process documented above yields the closest and best approximation for the given dataset.

The stochastic nature of the dataset and the number of factors influencing the Inflation rate in the United States is illustrated by the deviation of the predicted values from the actual values during the first quarter of 2015. Through the process of ARMA modeling, Fitting the trend and forecasting the Inflation rate for the next 12 months, the time series analysis of the raw dataset has been completed.