

Time Series Analysis and Forecasting of Pfizer Stock Prices

Arvind SY

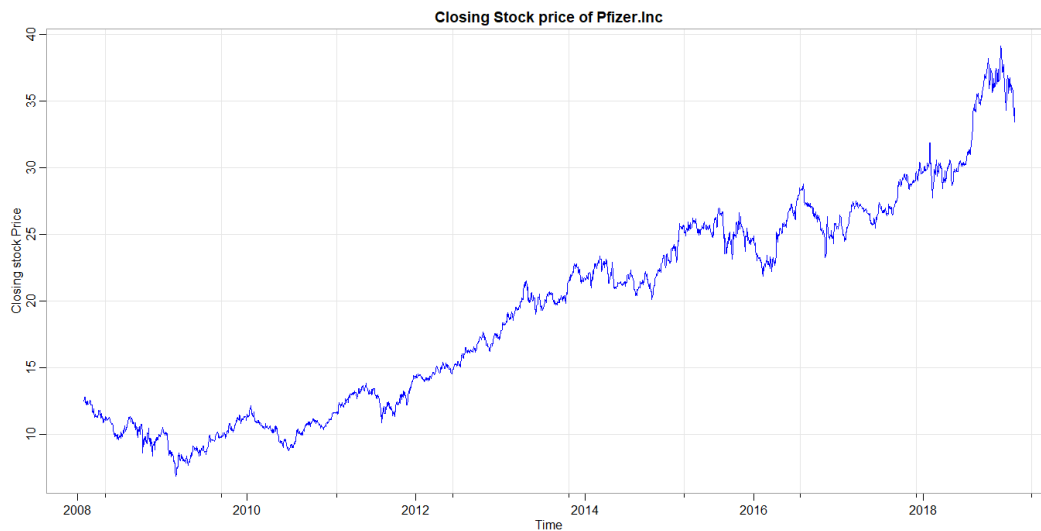
1 Introduction

1.1 Dataset

The dataset was taken from **quantmod** package in R. Data is from 2008-01-30 to 2019-01-30 of daily closing stock price of Pfizer Inc having symbol "PFE". The data structure for the analysis have 2769 observations of two variables i.e time and closing stock price. Let X_t be our closing stock price value and t is the index of time ranging from 1 to 2769.

1.2 Plot of the Time Series Data

Following is the time series plot of the data.



1.3 Motivation

Following are the important questions concerning the data you wish to address:

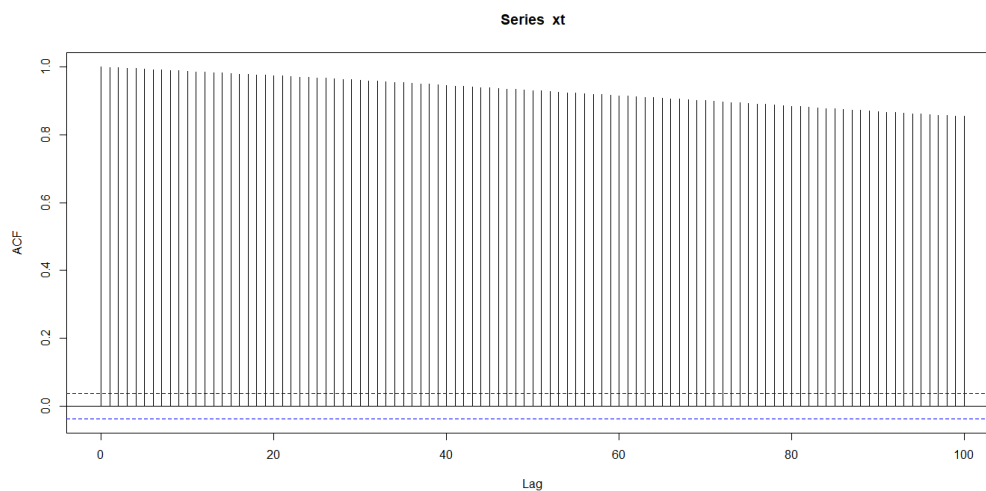
- Model identification.
- Model diagnostics.
- Forecasting

2 Statistical Analysis

From the above plot of the data we can observe that there is an increasing trend in closing stock prices of Pfizer.Inc.Now we will look at the acf and pacf of the data

2.1 ACF

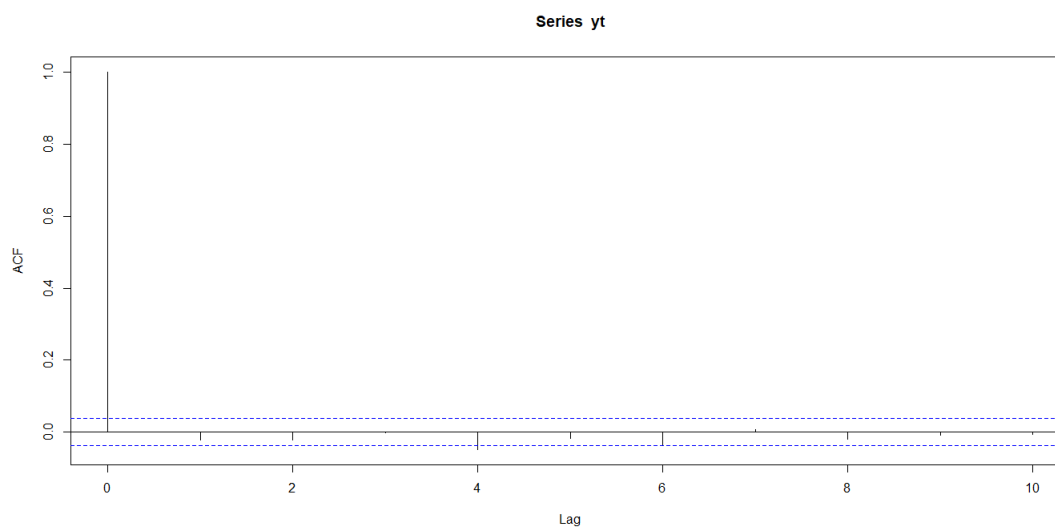
Following is the acf of the time series data X_t :



Observe that the values does not cutoff even after the 100 lags Now we will look at the acf of the first differenced data i.e

$$Y_t = X_t - X_{t-1}.$$

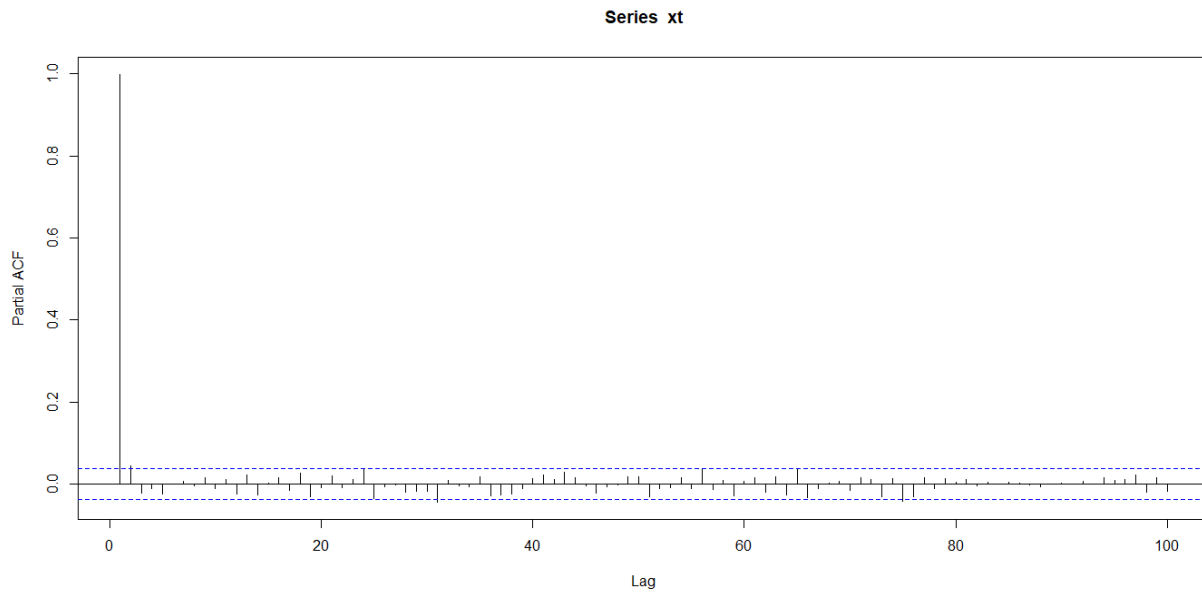
Following is the acf of the time series data Y_t



Observe that there is no significant lags.

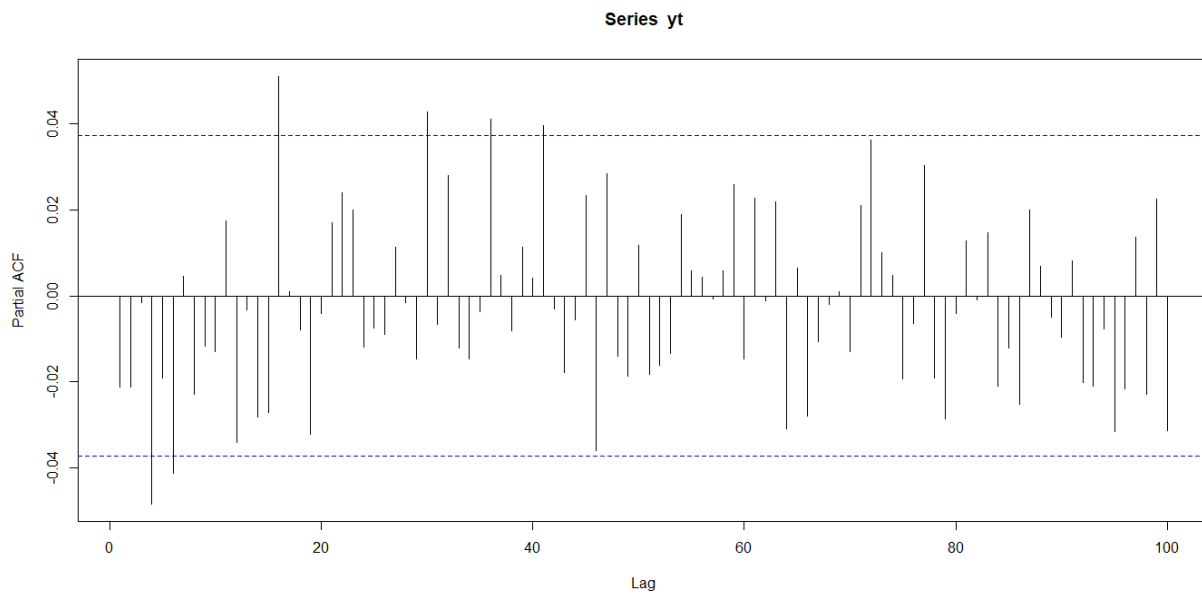
2.2 PACF

Following is the pacf of the time series data X_t



There is no significant lag in this .

Following is the acf of the Y_t .



Most of the lags are below the significance level .

We can say that we have a ARIMA(p,d,q) model with d=1. Now we will identify other parameters by using auto.arima and then will check its adequacy .

2.3 Model Identification

Now we will find the parameters of the our ARIMA model.Folowing are the parameters that have been found by the analysis of the data X_t .

```
Series: xt
ARIMA(1,1,1) with drift

Coefficients:
          ar1      ma1      drift
        0.8464  -0.8794   0.0080
s.e.    0.0567   0.0506   0.0037

sigma^2 estimated as 0.06014:  log likelihood=-35.55
AIC=79.09   AICC=79.11   BIC=102.8
```

Hence we have found the model to be **ARIMA(1,1,1)** with a drift value of 0.008. Now we will check the adequacy of our model.

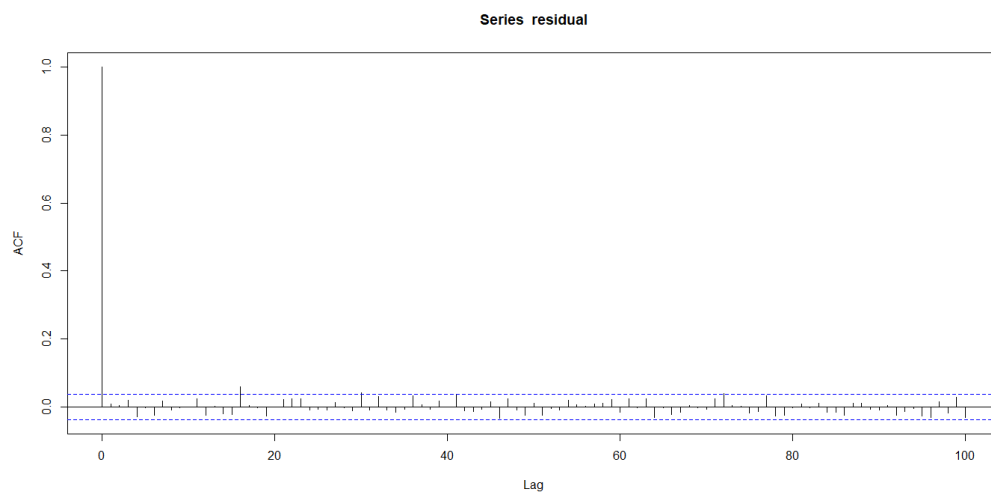
2.4 Model Diagnostics

For the diagnostic part we will check if the residuals of the model have a auto-correlation or not. Residual e_t is defined as

$$e_t = X_t - \hat{X}_t$$

where \hat{X}_t is the fitted model.

First we will check the acf of the residual and then will perform Ljung–Box test .Following is the ACF of the residuals:



Residuals does not have that significant lags in ACF.

2.4.1 Ljung–Box test

The Ljung–Box test may be defined as:

- H_0 : Residuals are independently distributed and not have correlation
- H_a : Residuals are not independently distributed, they exhibit serial correlation.

We will calculate p-value for this test on the residuals with level of significance at 0.05. Following is the analysis:

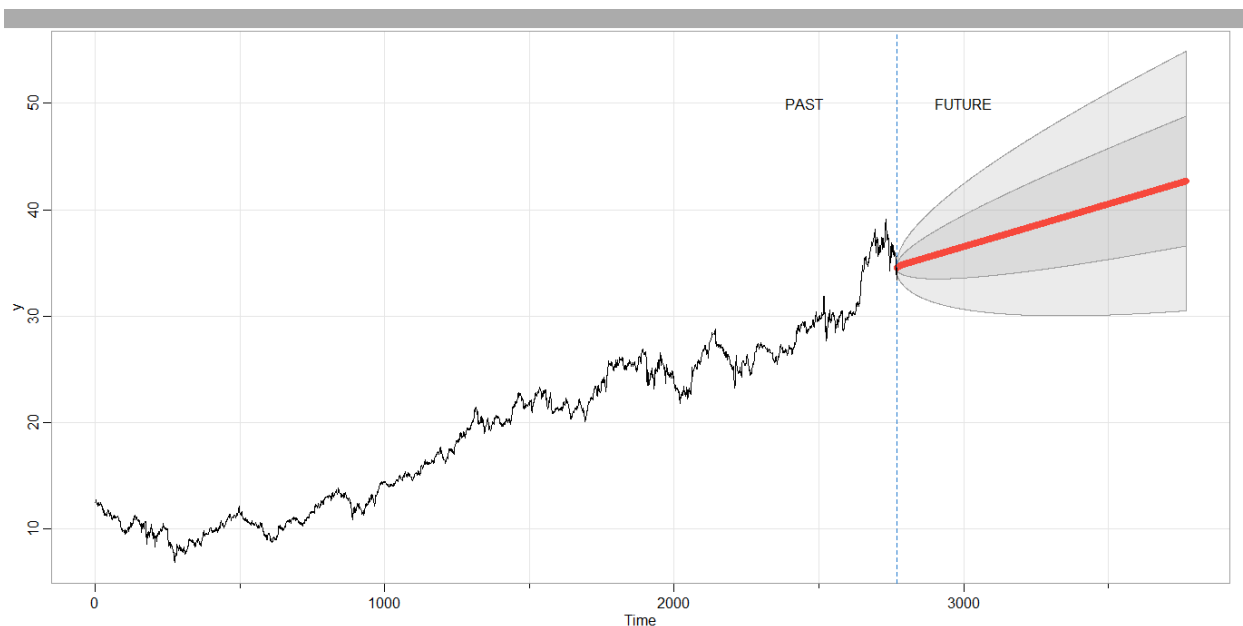
```
Box-Ljung test

data: residual
x-squared = 0.1932, df = 1, p-value = 0.6603
```

In the analysis p-value is found to be 0.6603 hence we accept H_0 at 5% level of significance and conclude that residuals are independently distributed and not have correlation. Now our model is good for forecasting the future values.

2.5 Forecasting

Now we will forecast the closing stocks price for the future 1000 days. Below is the plot.



3 Conclusion

Following are the conclusion drawn from our analysis:

- Stock price of the Pfizer has a increasing trend.
- Data was not stationary.
- After first order differencing data became stationary.
- Found that best suitable model for this data was ARIMA(1,1,1)
- Residuals was found to be uncorrelated using Ljung–Box test.

Our concern was to analyse the data and identify the model best suitable for forecasting. Also to estimate the parameters. After analysis we found that best suitable model for closing stock prices for Pfizer Inc data was ARIMA(1,1,1). Future forecasting has also been done.