Time Series

Stock Price Analysis of

EFU LIFE ASSURANCE LIMITED

STUDENT PROJECT

Prepared By

Usman Azher

Table of Contents

1. INTRODUCTION:

EFU Life is the leading life insurer in the private sector of Pakistan. Having started operations in 1992 as the first private sector life insurance company, over a span of 22 years EFU Life has established itself as a trusted brand name in providing all types of financial planning solutions. The Company markets its business through three main distribution channels - Sales Force, Bancassurance and Group Benefits. A comprehensive range of retail products are available targeting low income persons up to high net worth individuals. In addition, tailor made solutions are offered to the corporate sector through group life schemes.

The two other companies under the EFU brand name are:

- EFU General Insurance Ltd which is the leading general insurer in the country, and
- Allianz-EFU Health Insurance Limited, which is the first and one of the leading health insurance providers in the country

Key Aspects about EFU Life Assurance Limited are illustrated below:

- Paid-Up Capital of PKR 1 Billion; highest in the life insurance private sector
- Strong Insurer Financial Strength Rating of AA Outlook: Stable by JCR-VIS Credit Rating Agency
- Diversified portfolio with most comprehensive product range
- Best Performing Investment Funds
- Largest Agency Sales Force in the private sector
- Caring Customer Support Services with dedicated help-line
- The first life insurance company in Pakistan to have ISO 9001:2008 certification
- Company Assets worth more than Rs 55 Billion
- One of the largest fund managers of Pakistan; with funds in excess of Rs. 50 Billion

As I am working in EFU Life as an Actuarial Analyst, this analysis has helped me a lot. I have extracted daily stock price of my company for the year 2012. This data was extracted from the website of Karachi Stock Exchange. As we know that stock prices are very volatile and many stochastic time series analysis could be done on this data.

2. DATA AND METHODOLOGY:

Time series models have been the basis for any study of a behavior of process or metrics over a period of time. The applications of time series models are manifold, including sales forecasting, weather forecasting, inventory studies etc. In decisions that involve factor of uncertainty of the future, time series models have been found one of the most effective methods of forecasting. Most often, future course of actions and decisions for such processes will depend on what would be an anticipated result. The need for these anticipated results has encouraged organizations to develop forecasting techniques to be better prepared to face the seemingly uncertain future. Also, these models can be combined with other data mining techniques to help understand the behavior of the data and to be able to predict future trends and patterns in the data behavior.

The development of new techniques and ideas in econometrics has been rapid in recent years and these developments are now being applied to a wide range of areas and markets.

Especially the area of forecasting and control is a hot issue these days since a lot of companies try to optimize their business processes and want to have a good estimate of production planning throughout a large time period. Therefore, better ways of data analysis are being developed to ensure promising forecasting methods.

Though time series analysis is a broad area of research it is mostly used to optimize planning and consists of two primary goals: identifying the nature of the phenomenon represented by the sequence of observations and forecasting (predicting future values of the time series variables). Both of these goals require that the pattern of observed time series data is identified and more or less formally described. Once that pattern is established, it can be interpreted and integrated into other data. Regardless of the depth of our understanding and the validity of our interpretation of the phenomenon, one can extrapolate the identified pattern to predict future events.

There are a lot of theories known that can make good trend estimations, however there are still a lot of problems that cannot be fully resolved when it comes to trends. It is often unclear where the trend ends and the fluctuations begin, and the desiderata for separating the two, if possible, have remained in dispute. Secondly, it is still hard to extract the trend, even when it is a clearly defined entity.

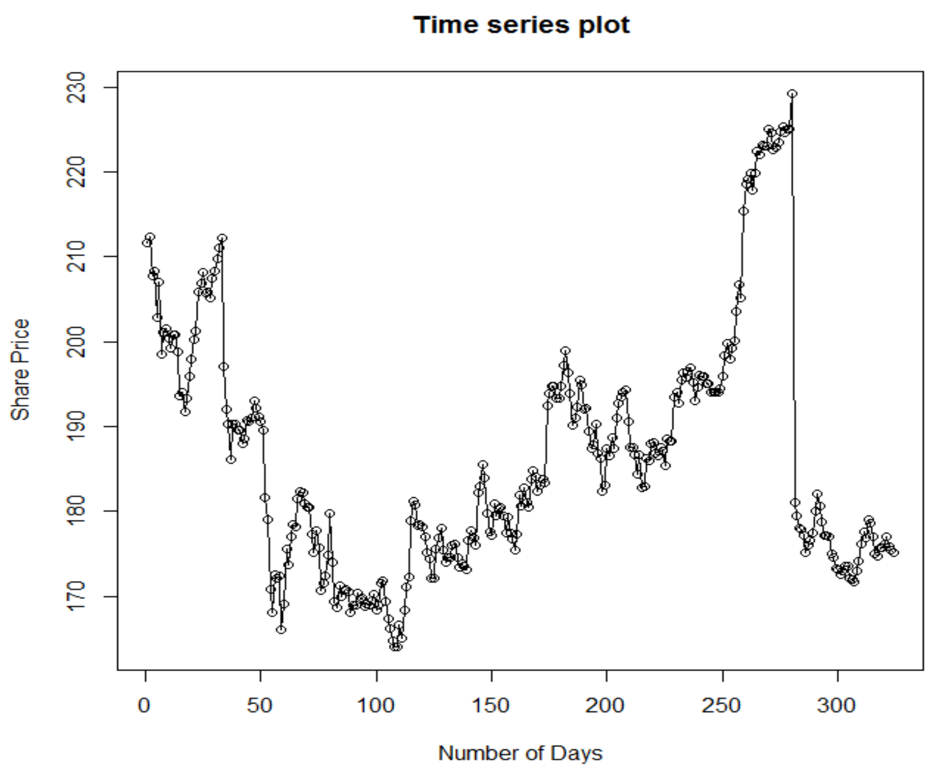
The purpose of this analysis therefore is to make a review of some methods which are available for obtaining estimates of the trend. Using an example some techniques will be discussed and compared to each other.

3. ANAYSIS OF THE DATA:

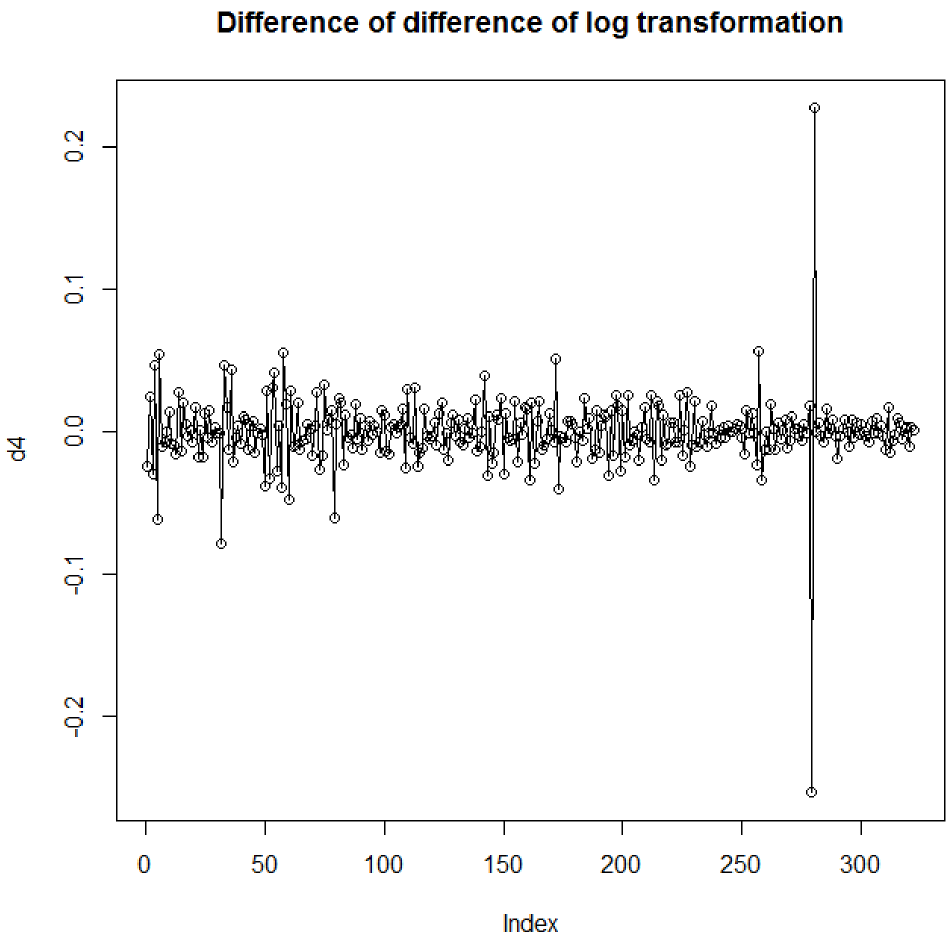
Macroeconomic and microeconomic time series often have an upward drift or trend which makes them non-stationary. Since many statistical procedures assume stationarity, it is often necessary to transform data before beginning analysis. There are a number of familiar transformations, including differencing.

3a. To check Stationarity:

Time series plot of the data is as follows:



Since, our data is not following stationarity, we took difference of log transformation. Difference of log transformation is shown as follows:

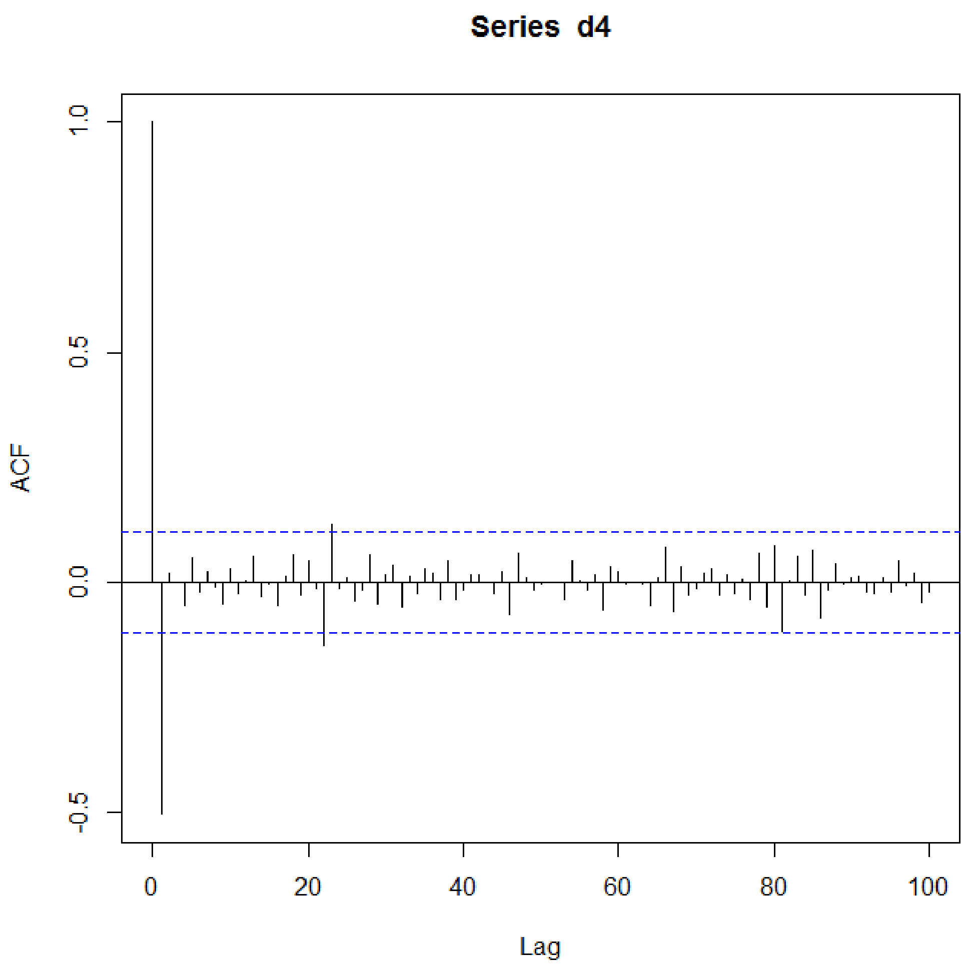


By taking difference of difference of log transformation, our series seems to be much more stationary than the original data. A big reason for using a stationary data sequence instead of a non-stationary sequence is that non-stationary sequences, usually, are more complex and take more calculations when forecasting is applied to a data series.

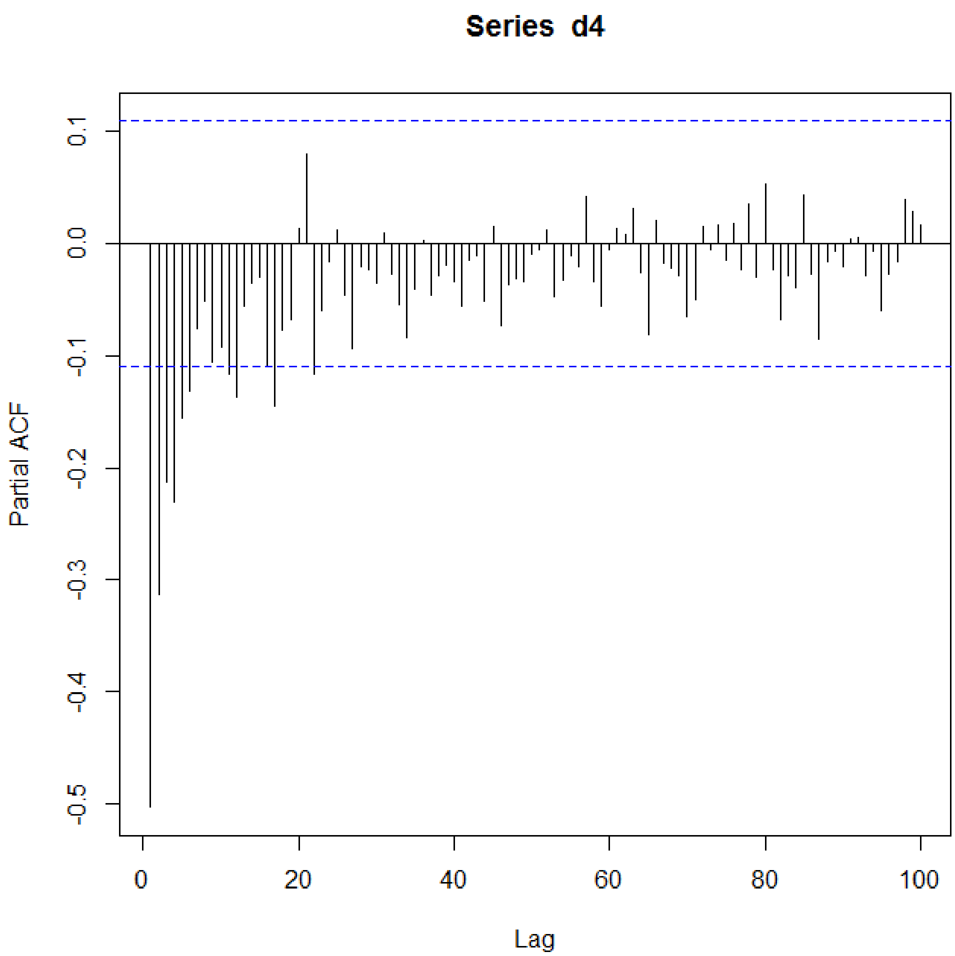
3b. Plotting the Autocorrelation Functions:

When modeling, an advantage of using the time series model is to fit a random walk model as it easily allows adding terms to correct the model for autocorrelation in the residuals, if this should be necessary. In particular, if the time series model has significant positive autocorrelation in the residuals at lag 1, one should try and use a so called AR (1) model. On the other hand, if the random walk model has significant negative autocorrelation in the residuals at lag 1, one should try and fit a MA (1) model.

AUTOCORRELATION FUNCTION OF THE TRANSFORMED DATA



PARTIAL AUTOCORRELATION FUNCTION OF THE TRANSFORMED DATA



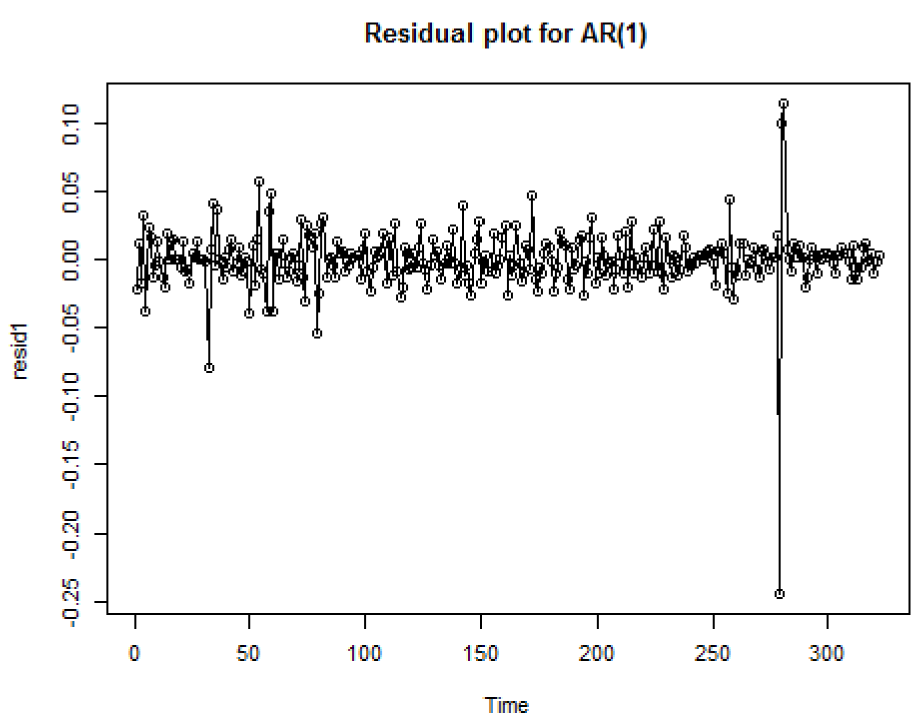
Normally, the correct amount of differencing is the lowest order of differencing that yields a time series which fluctuates around a well-defined mean value and whose auto-correlation function plot decays rapidly to zero, either from above or below. If the series still exhibits a log-term trend, or otherwise lacks a tendency to return to its mean value, or if its autocorrelations are positive out to a high number of lags (10 or more), then it needs a higher order of differencing. Now, if the lag-1 autocorrelation is zero or even negative, then the series does not need further differencing. In this case it is not recommended to difference another time, because this will only result in "over differencing" the series and end up adding extra AR or MA terms to undo the damage.

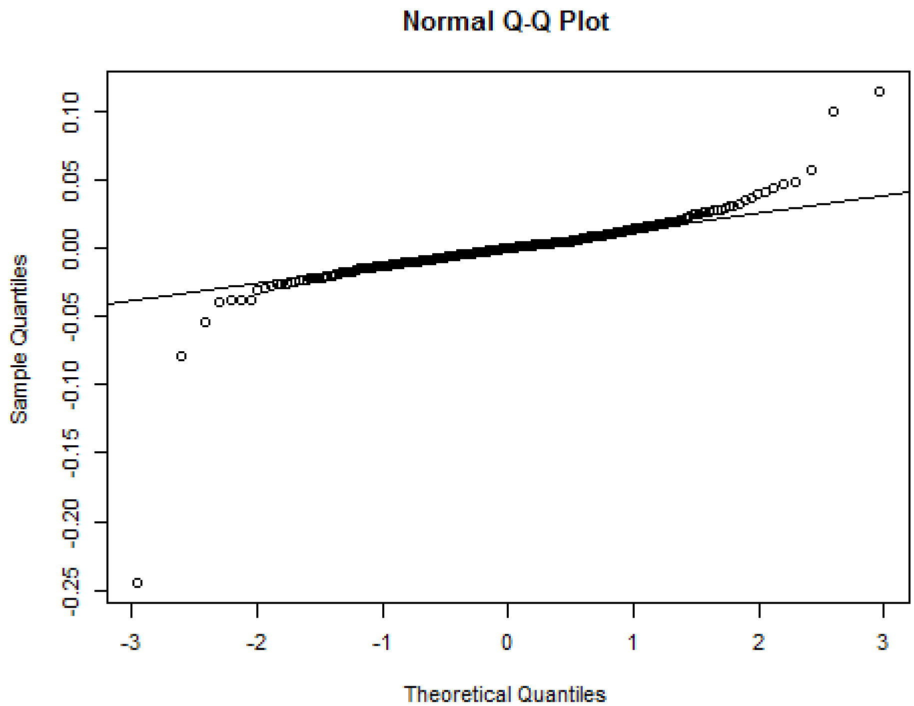
4. MODEL FITTING

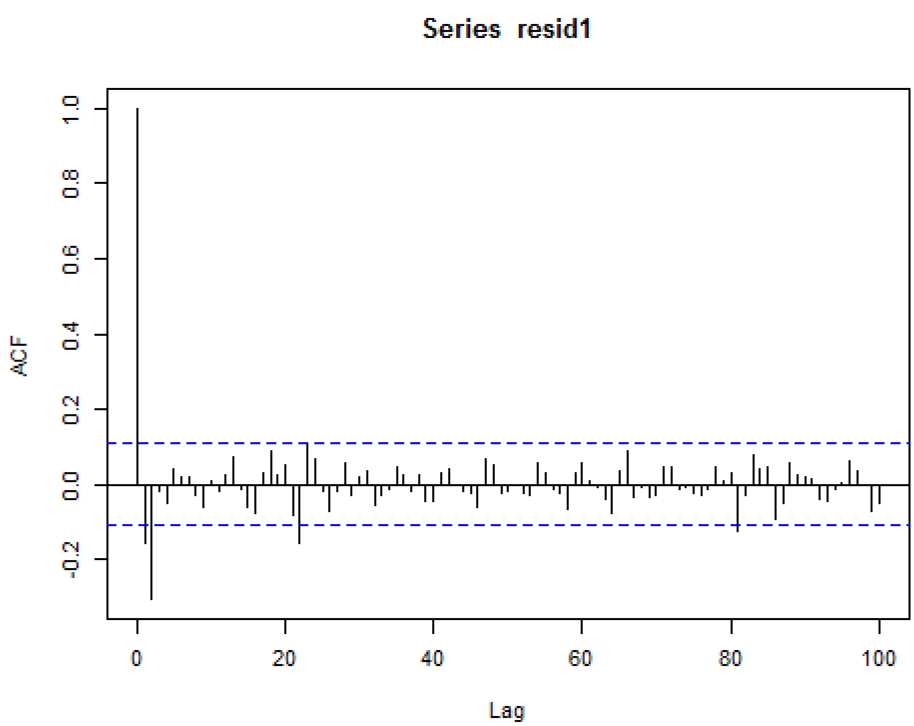
Two different models were fitted as follows:

4a. <u>Autoregressive model of order 1:</u>

```
Call:

arima(x = d4, order = c(1, 0, 0))



Coefficients:

        ar1    intercept

     -0.5032      0e+00

s.e.   0.0481      8e-04



sigma^2 estimated as 0.0005006:  log likelihood = 766.52,  aic = -1527.05
```




Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.20
-0.25
-3
-2
-1
0
1
2
3


Series resid1
ACF
Lag
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
20
40
60
80
100

Shapiro-Wilk normality test

data: resid1

W = 0.7075, p-value < 2.2e-16

Box-Pierce test

data: resid1

X-squared = 7.97, df = 1, p-value = 0.4756

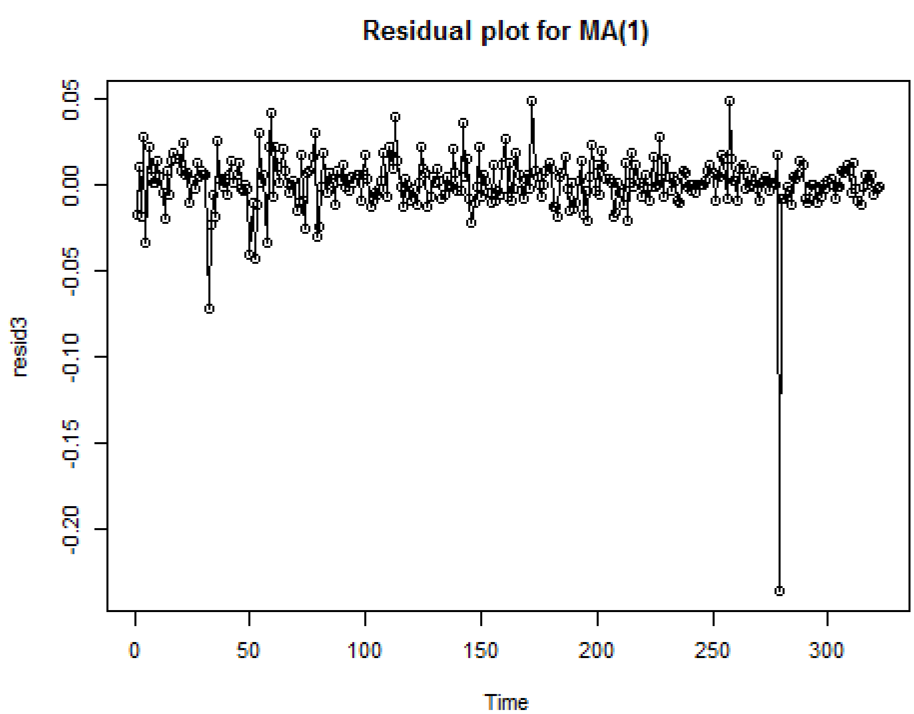
When faced with a time series that shows irregular growth, differencing can be seen as predicting the change that occurs from one period to the next in a time series *Y(t)*. In other words, it may be helpful to look at the first difference of the series, to see if a predictable pattern can be discerned there. For practical purposes, it is just as good to predict the next change as to predict the next level of the series, since the predicted change can always be added to the current level to yield a predicted level.

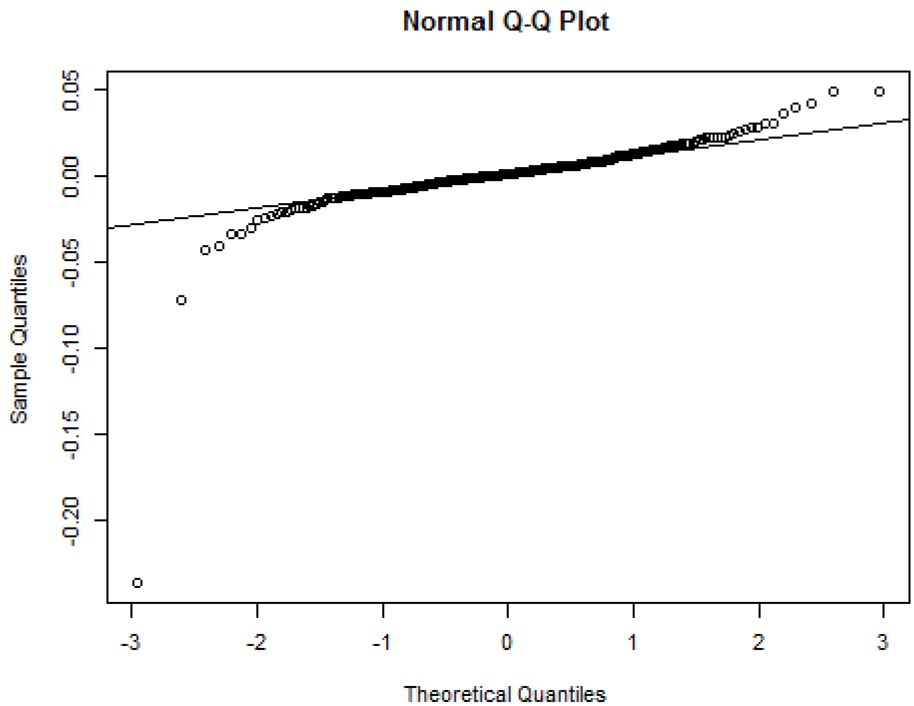
In this particular model intercept appears to be zero. Residuals are following a normal distribution and the residuals are identical and independently distributed as shown in the Shapiro test with small p-value. One of the methodologies that use ARIMA processes and differencing as basis is the Box method. In this model Box test also shows that there is no significant autocorrelation between the residuals.

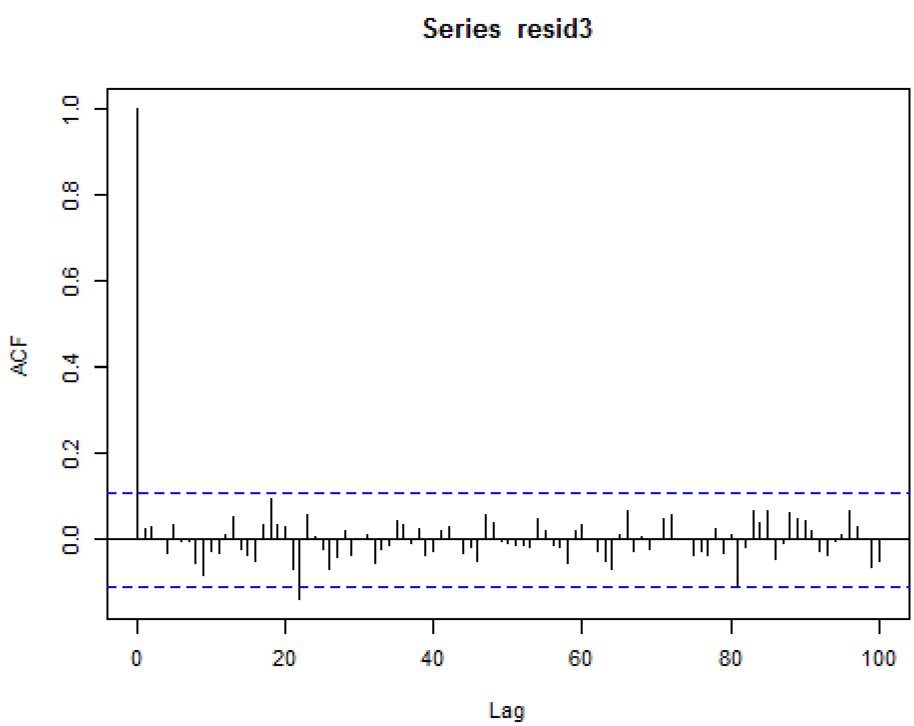
Hence we can say that as autoregressive variable has p-value above 2, therefore the assumed variable is significant as well.

4b. Moving average model of order 1:

```
Call:

arima(x = d4, order = c(0, 0, 1))



Coefficients:

        ma1  intercept

    -1.0000      0e+00

s.e.   0.0108      1e-04



sigma^2 estimated as 0.0003442:  log likelihood = 824.07,  aic = -1642.15
```




Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
0.05
0.00
-0.05
-0.10
-0.15
-0.20
-3
-2
-1
0
1
2
3


Series resid3
ACF
Lag
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100

Shapiro-Wilk normality test

data: resid3

W = 0.6264, p-value < 2.2e-16

Box-Pierce test

data: resid3

X-squared = 0.2018, df = 1, p-value = 0.6533

It has been argued that the best way of separating the trend from the fluctuations is to model both of them at the same time within the framework provided by a structural ARIMA model which assigns separate parameters to the components. There have been a lot of examples of detrending which follow this prescription. Two of them that will be discussed are the method described by Hillmer and Tiao, which are concerned with extracting the hidden components from existing seasonal ARIMA models, and a methodology which is concerned with filtering short sequences using a rational operator.

In this particular model, once again intercept appears to be zero. Residuals are following a normal distribution and the residuals are identical and independently distributed as shown in the Shapiro test with small p-value. Box test also shows that there is no significant autocorrelation between the residuals (large p-value).

Hence we can say that as autoregressive variable has p-value above 2, therefore the assumed variable is significant as well.

5. MODEL SELECTION

On the basis of our above constructed models and tests conducted on them, we have selected **AR(1) model** as the most appropriate as the included autoregressive variable is very significant with t value above 5 and furthermore residuals also follow a normal distribution as shown in the qq plot above.

Shapiro and Box test conducted on the residuals show that residuals are identical and independently distributed and residuals does not have significant autocorrelation between them.

AR(1) model with seasonal effects

```
Call:

arima(x = d4, order = c(1, 0, 0), seasonal = list(order = c(1, 0, 0), period = 12))



Coefficients:

         ar1     sar1    intercept

     -0.5040  0.0248      0e+00

s.e.   0.0481  0.0557      9e-04



sigma^2 estimated as 0.0005002:  log likelihood = 766.62,  aic = -1525.25
```

Since the t value associated with seasonal variable is not significant, we negate the presence of any seasonality in our data set. Box-Jerkins technique can be used for determining the most appropriate ARMA or ARIMA model for a given variable. The Box-Jenkins methodology is often referred to as more an art then a science, this lack of theory behind the models is one criticism of them, however they are used as an effective model for forecasting.

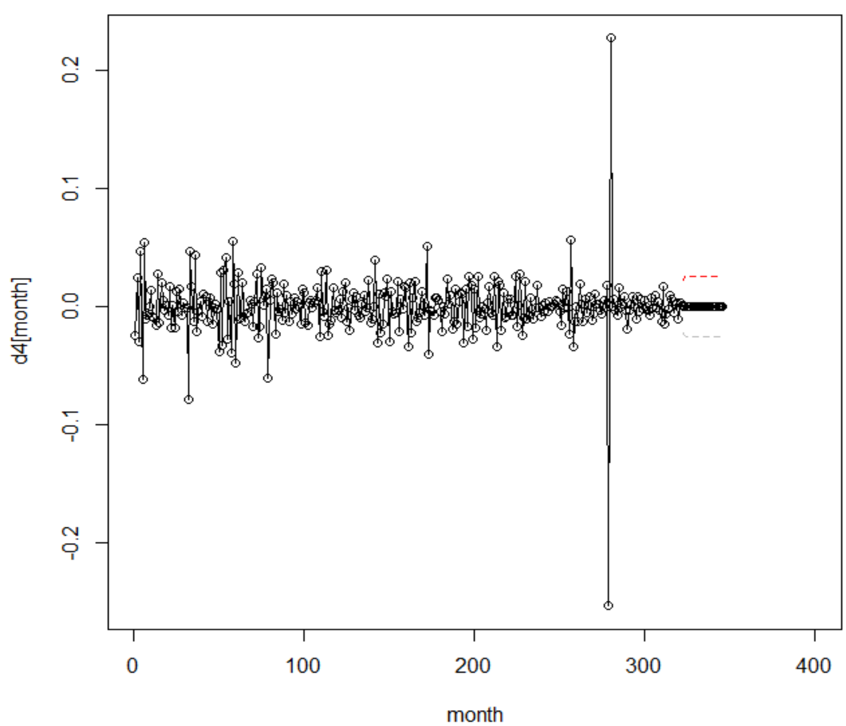
6. FORECASTING

One of the most important tests of any model is how well it forecasts. This can involve either in-sample or out-of-sample forecasts, usually the out-of-sample forecasts are viewed as the most informative, as the data used for the forecast is not included in the estimation of the model used for the forecast. When assessing how well a model forecasts, we need to compare it to the actual data, this then produces a forecast value, an actual value and a forecast error (difference between forecast and actual values) for each individual observation used for the forecast.

When forecasting future values of a variable, it is often important to have a benchmark model, such as the random walk to compare the forecasts of the model with, if it can not beat the random walk it can be argued to be a relatively poor forecaster, the random walk often wins.

Initially, we left out last two observations from our dataset to be predicted later on. On the basis of our yuke walker model, the predicted values are as follows:

Observation number	Predicted price	True price
325	175.6	174.96
326	176.02	174.76



7. CONCLUSION:

Though forecasting and prediction is not very accurate, it would be good if we could achieve higher percentage of accuracy. Many types of business data are observed at regular time intervals thus giving rise to time series data. We often use regression models to estimate the Trend component of a time series variable. However, regression models of time series may pose a special problem. Because business time series tend to follow economic trends and seasonal cycles, the value of a time series at time t is often indicative of its value at time (i + 1). That is, the value of a time series at time t is correlated with its value at times (i + 1). Consequently, we cannot apply the standard least squares inference making tools and have confidence in their validity. In this section, we present a method of testing for the presence of residual correlation. Time series analysis serves as a framework for making any business related decisions in the modern world. It is in fact a very popular tool for business forecasting. It serves as a basis of understanding past behavior and also a basis for comparison.

Our report also, establishes concrete evidence that, time series analysis is a reliable tool for business forecasting and decisions. Our forecasted values are also very close to the actual values in terms of stock price.

8. R COMMANDS:

PLOT

```
d1=read.table("E:/RAZA/Time Series/timeseries/share.dat",header=TRUE)
attach(d1)
d2=log(share)
d3=diff(d2)
d4=diff(d3)
plot(share,main=’Time series plot’)
plot(d4,main=’Difference of difference of log’)
```

Acf and pacf

```
acf(d4,100)
Pacf(d4,100)
```

AR(1)

```
d1=read.table("E:/RAZA/Time Series/timeseries/share.dat",header=TRUE)
attach(d1)
d2=log(share)
d3=diff(d2)
d4=diff(d3)
mod1=arima(d4,order=c(1,0,0))
mod1

resid1=resid(mod1)
plot(resid1)
plot(resid1,type="o")
qqnorm(resid1);qqline(resid1)

plot(resid1,type="o",main="residual plot for AR(1)")
acf(resid1,main="ACF for AR(1)")
shapiro.test(resid1)
Box.test(resid1)
```

MA(1)

```
d1=read.table("E:/RAZA/Time Series/timeseries/share.dat",header=TRUE)
attach(d1)
d2=log(share)
d3=diff(d2)
d4=diff(d3)
mod3=arima(d4,order=c(0,0,1))
mod3

resid3=resid(mod3)
plot(resid3)
plot(resid3,type="o")
qqnorm(resid3);qqline(resid3)

plot(resid3,type="o",main="residual plot for AR(1)")
acf(resid3,main="ACF for AR(1)")
shapiro.test(resid3)
Box.test(resid3)
```

FORECASTING

```
mod1=ar.yw(share,order=1)
 mod1.pr=predict(mod1,n.ahead=2)
 mod1.pr

U=mod1.pr$pred+mod1.pr$se
L=mod1.pr$pred-mod1.pr$se
month=1:326
plot(month,d4[month],type='o')
lines(mod1.pr$pred)
lines(U,lty='dashed',col='10')
lines(L,lty='dashed',col='1000')
```

SEASONALITY

```
Mod1=arima(d4,order=c(1,0,0),seasonality=c(list(order=c(1,0,0),period=12))
```