Univariate Time Series Analysis of Forecasting Asset Prices

Tanu Shivnani
Research Scholar,
Jawaharlal Nehru University, Delhi.

Received Aug 17, 2016 Accepted Sept. 15, 2016

ABSTRACT *This paper deals with forecasting of asset prices through Box Jenkins Methodology. It first checks the stationarity of the series and after finding it to be non stationary we either obtain log transformation of prices or percentage change in prices. Thereafter by autocorrelation and partial autocorrelation method we decide the model based on the significance. The AIC-BIC criteria along with least mean square error help to decide the model and forecast the future values of prices of the copper.*

KEY WORDS: *Stationary, Box Jenkins methodology, Autocorrelation, Partial Autocorrelation, AIC, BIC, forecast*

1. INTRODUCTION

"Anyone who thinks there's safety in numbers hasn't looked at the stock market pages". - Peter

Securities market is a market which brings together buyers and sellers, who trade stocks of publicly held companies listed on stock exchange. This trade takes place through exchange or over the counter markets. The buyer by purchasing the stocks of a company holds a share of ownership rights of that company by the proportion of stocks held by him. Investor can then participate in the financial and management activities of the company. There is certain risk involved in these markets. The investor may get dividends and capital gains (by selling the share at a higher price) if the company's performance is good but on the contrary the investor may have to go through losses if the company's performance or health is not good. (Example: during Satyam's scam when Ramalinga Raju admitted on 7th January, 2014 that he had manipulated the accounts of the company by $1.47mn, the stock price had plummeted by a large amount). Stock market is usually found to be volatile.

It is of two types, namely: primary and secondary. Primary market deals with creation and sale of new securities often called initial public offering (IPO), while secondary market deals with sale and purchase of securities already issued.

It has been observed that they are affected by external shocks like change in political party or change in economic conditions of other countries. It has been observed that there was a positive sentiment in the stock market in India when there was a rumour that right wing party is coming in power. There was huge inflow of foreign institutional investors, who invest in the country for a short period (around one year) in order to reap gains and there was capital outflow when it was head that a left wing party is coming to power.

During the sub-prime crisis of 2007-08, when the housing bubble had burst due to high rate of default because 90% of mortgages were adjustable rate of mortgages. This low quality of lending led to not only higher risk of default rate but there were job losses up to nine million, high debts, trade imbalance, around 30% fall in housing price and fall in GDP. The stock market plummeted up to 50% in USA in 2009 and India also witnessed a fall in stock market during that period on this account. It goes with the saying that "When USA sneezes whole world gets affected".

The Euro-zone crisis of 2009 where various European banks and central banks had asked for a bailout fund, as it had become difficult for them to refinance their government debt, also brought a negative sentiment in the stock market.

During the Harshad Mehta scam of 1992, when Mehta had manipulated stocks which were financed by worthless bank receipts, there had been huge stock market crash. It was a financial scandal of Rs. 4999 crores on Bombay Stock Exchange on account which stock prices fell drastically.

As wisely said by George Soros, *"Stock market bubbles don't grow out of thin air. They have a solid basis in reality, but reality as distorted by a misconception".*

2. DESCRIPTION OF DATA

The trading in India takes place through Bombay Stock Exchange (BSE) and National Stock Exchange (NSE).BSE has been established since 1875 while NSE was found in 1992 and started trading operations in 1994.There are about 4700 listed companies firms on BSE while there are about 1200 listed firms on NSE. Due to arbitrageurs the prices of both the markets are within certain bands.

Two prominent market indexes are: Sensex and NIFTY. There are 30 listed firms on BSE which account for 45% of index's market capitalisation. While, S&P CNX NIFTY has 50 firms listed on NSE which account for 62% of free float market capitalisation.

Here in this case we are taking analysing the asset price of Tata Motors. The data has been extracted from BSE, Bombay Stock Exchange. The face value of share of Tata Motors is Rs.2. It is a daily price series ranging from 1st January, 2010 to 3rd February, 2014. It is from group A and S&P BSE Sensex. The unit of data is in rupee terms.

Time Series plot of daily values of Tata Motors

Asset Price Series

Figure 1

This series shows an upward trend (except during the period of 2007-08 during which there was subprime lending crisis, which has not been taken into data under consideration). It can be observed that there will be no constant mean for various time periods.

The autocorrelation function also shows that series is non stationary. With reference to the figure 2, it can be observed that all the autocorrelations lie outside the 95% confidence interval.

ACF of Price Series

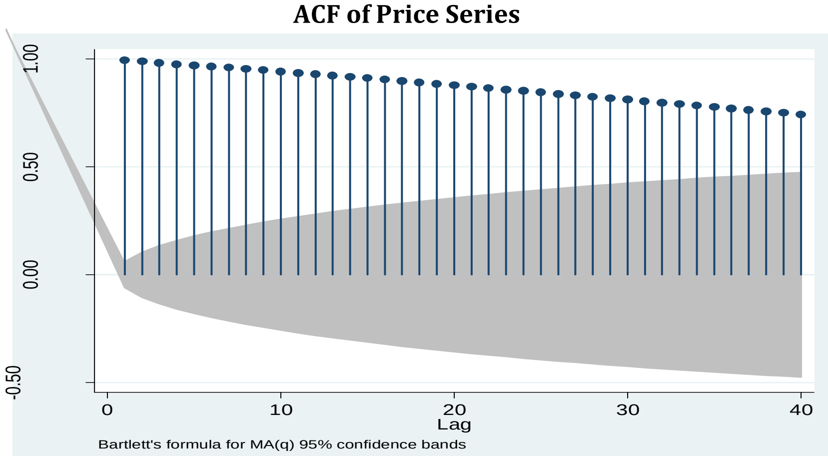

Figure 2

As observed in Partial Autocorrelation lies within in 95% confidence interval except the first lag.

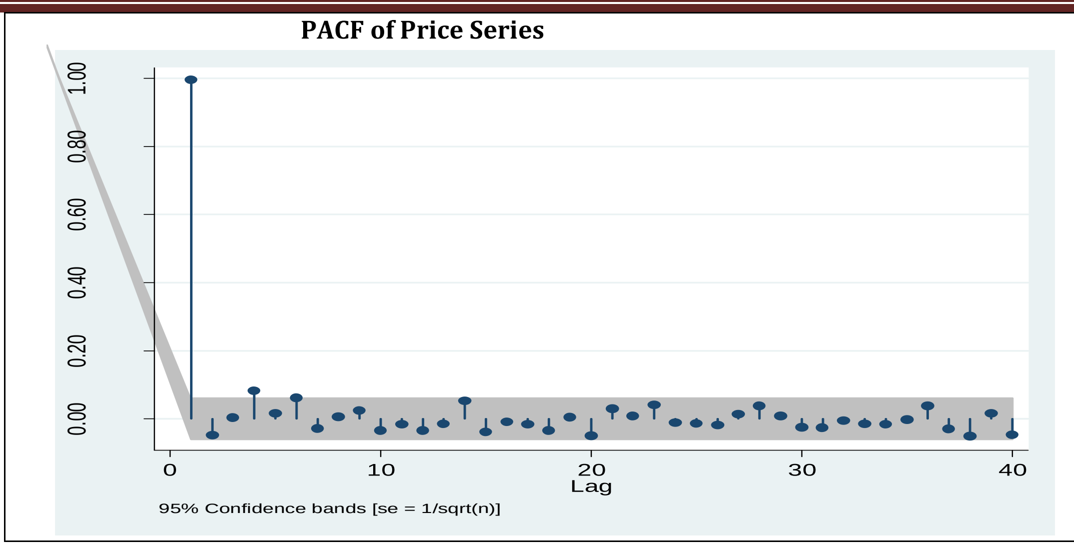

Figure 3

From the fig2&3 it can be seen that ACF decaying slowly while PACF has cut off at lag one. Therefore the price series is non stationary, and in order to make it stationary we take the first difference of price series (following difference stochastic process).

3. STATIONARITY

In order to confirm whether the given price series is stationary or not we check it using Dickey Fuller test.

Dickey Fuller Test

Y_t = □ Y_{t-1} + u_t (also called random walk model without drift)

$Y_t - Y_{t-1}$ = □Y_t = □ □ Y_{t-1} + u_t

Dickey-Fuller test for unit root				No. of Observations=1048
	Interpolated Dickey-Fuller			
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	0.565	-2.580	-1.950	-1.620
MacKinnon approximate p-value for Z(t) = 0.5811				

Table 1

	Coefficient	Std. Error	t	P>|t|	[95% Conf. Interval]	
Lag 1	.0003964	.0007022	0.56	0.573	-.0009815	.0017744

Y_t = □$_1$ + □ Y_{t-1} + u_t (also called random walk model with drift)

$Y_t - Y_{t-1}$ = □Y_t = □$_1$ + □ □ Y_{t-1} + u_t Where □ □ □ □ □ □ □ 1

H_o: □ = 0, implies there is unit root i.e. series is non stationary

H_a: □ □ □ □ □ implies series is stationary

Dickey-Fuller test for unit root				No. of Observations=1048
	Interpolated Dickey-Fuller			
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-1.402	-3.430	-2.860	-2.570
MacKinnon approximate p-value for Z(t) = 0.8500				

Table 2

	Coefficient	Std. Error	t	P>|t|	[95% Conf. Interval]	
Lag 1	-.0038435	.0027407	-1.40	0.161	-0.0092215	0.0015345
Constant	1.113313	0.6956738	1.60	0.110	-0.2517649	2.478391

Since the absolute value of calculated Z(t) is less than the critical value at 1%, 5% and 10% level of significance and the coefficient of lag(1) term is close to zero(implying □=1), in both Random Walk model with drift and without drift, therefore we don't reject the null hypothesis; i.e. series is not stationary.

In order to make it stationary we take the first difference of the series (or we can even take simple returns series which also a form of first difference series).

Now, we check for stationarity of returns series (or alternatively first difference series).

Dickey-Fuller test for unit root			No. of Observations=1046	
	Interpolated Dickey-Fuller			
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-30.409	-2.580	-1.950	-1.620
MacKinnon approximate p-value for Z(t) = 0.8500				

Table 3

	Coefficient	Std. Error	t	P>|t|	[95% Conf. Interval]	
Lag 1	-.9400946	0.0309154	-30.41	0.000	-1.000758	-0.8794313

Table 3 shows the unit root test or random walk test for without drift model.

Dickey-Fuller test for unit root			No. of Observations=2558	
	Interpolated Dickey-Fuller			
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-30.444	-3.430	-2.860	-2.570
MacKinnon approximate p-value for Z(t) = 0.0000				

Table 4

	Coefficient	Std. Error	t	P>|t|	[95% Conf. Interval]	
Lag 1	-.9418037	0.0309354	-30.44	0.000	-1.002506	-.8811011
Constant	0.0009492	0.0007457	1.27	0.203	-.0005141	.0024124

Table 4 shows the unit root test or random walk test for with drift model.

Both the models, random walk model with drift and without drift, show that absolute value of calculated Z(t) more than the critical value at 1%, 5% and 10% level of significance and value of [illegible] therefore we reject the null hypothesis that return series is non stationary.

As seen from figure 4&5, the return series for logarithmic returns and simple returns is stationary.

Simple Returns

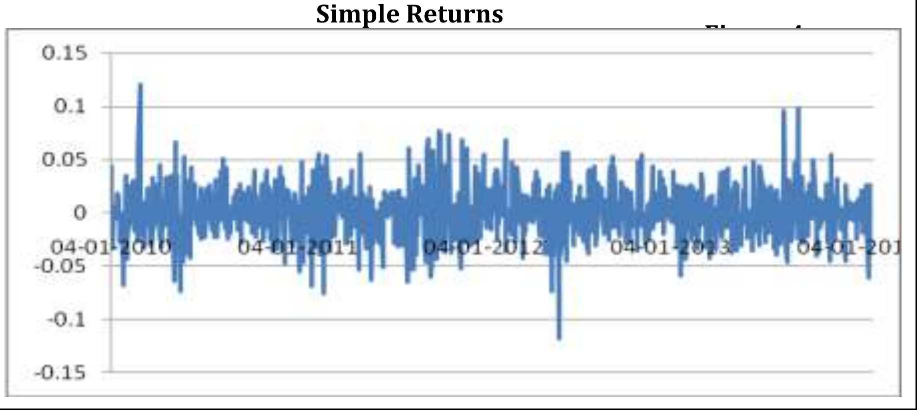

Figure 4

Logarithmic Returns

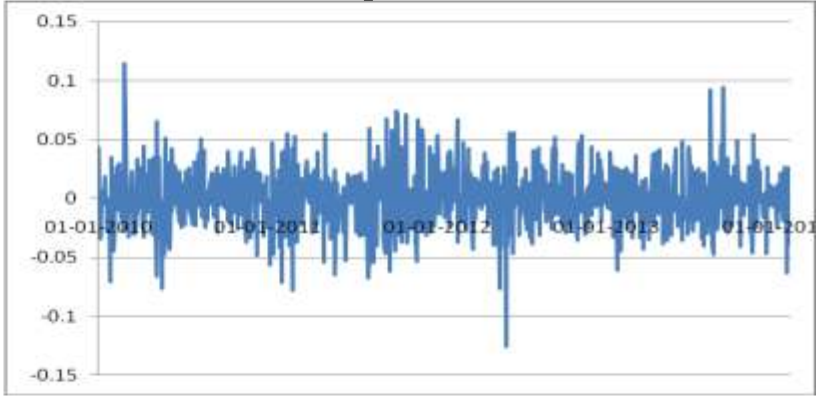

Figure 5

4. BOX JENKINS METHODOLOGY

The Box Jenkins methodology has four steps:

Step1. Identification: In this stage we identify(p,d,q) the cut-offs for the lags of Autoregressive Integrated Moving Average(ARIMA) where p gives cut off for AR process, q gives cut off for MA process and d gives order of integration.

Step2. Estimation: After identifying the cut-offs, we estimate the parameters of autoregressive and moving average terms in the model.

Step3. Diagnostic Checking: We check if the chosen model is best fit or not. Or if some other ARIMA model will do good job. In order to find whether the model is perfect fit or not we check if the residuals are white noise or not.

Step4. Forecasting: We forecast the future values based on the estimated model.

4.1 Identification

Autocorrelation function is a test for stationarity

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{covariance at lag } k}{\text{variance}}$$

ρ_k is population autocorrelation and $-1 \leq \rho_k \leq 1$.

Sample covariance at lag k is $\hat{\gamma}_k$, sample variance is $\hat{\gamma}_o$, sample autocorrelation $\hat{\rho}_k$

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$$

Where

$$\hat{\gamma}_k = \frac{\sum (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{n}$$

And

$$\hat{\gamma}_0 = \frac{\sum (Y_t - \bar{Y})^2}{n}$$

Partial autocorrelation (ϕ_{kk}) is correlation between Y_t and Y_{t-k} after removing the effect of intermediate Ys .

We plot the correlograms of ACF and PACF for return series.

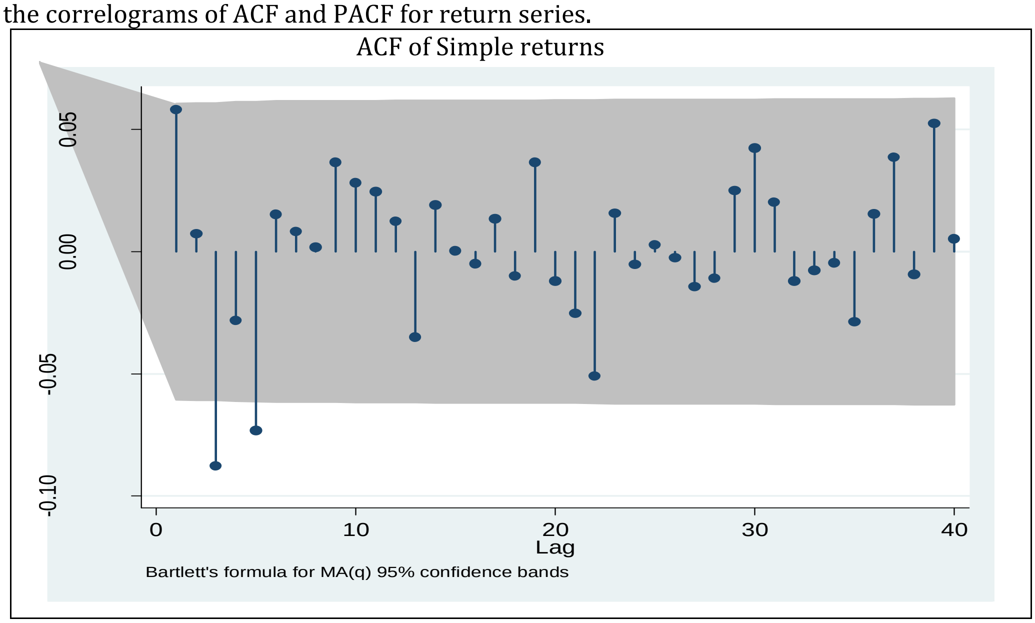

Figure 6

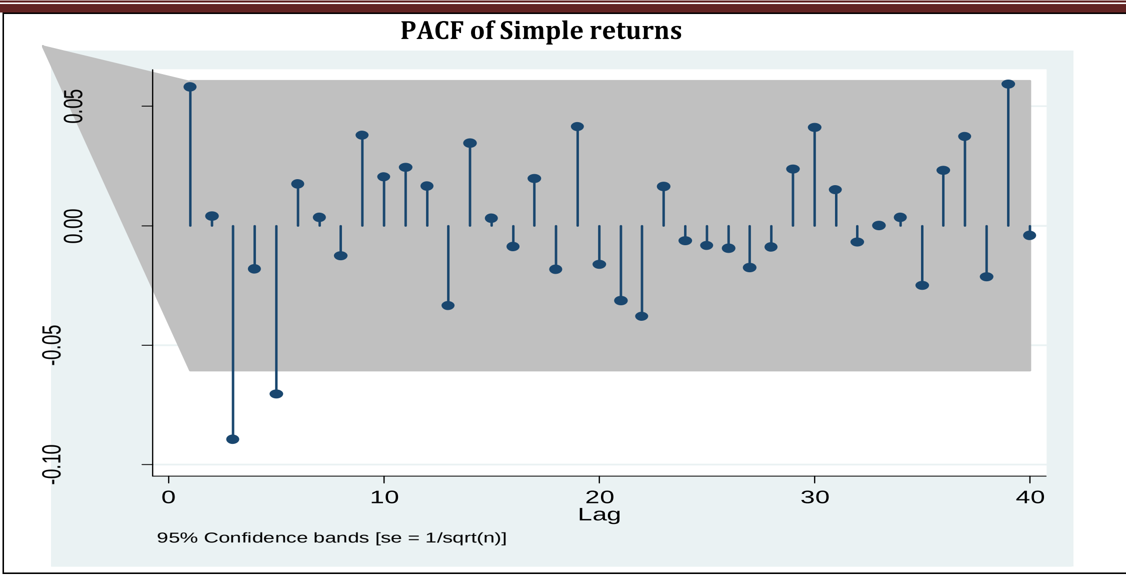

Figure 7

While ACF gives cut-offs for MA process and PACF gives cut-offs for AR process. ACF and PACF of simple returns both have cut offs at 3and 5. It is found that this series follows ARMA process.

The Box-Jenkins Model should satisfy the condition of Parsimony i.e. the least number of parameters should be estimated as incorporating additional coefficients will necessarily increase at a cost of reducing the degrees of freedom. Larger models tend to fit the data very well but perform very poorly when used for out-of-sample forecasts. Instead, smaller values of *p* and *q* in the case of univariate *ARMA* processes give better forecasts which are more robust.

4.2 Estimation

While working on various combinations of lags p & q of AR and MA respectively. We will choose the model which has minimum value of AIC and BIC criterion and correspondingly maximum value of log likelihood.

ARMA process may resemble the following pattern:

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots \dots \dots \dots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4} + \cdots \dots \dots \theta_q \varepsilon_{t-q} + \varepsilon_t$$

But since minimum value of AIC is for the model (5, 0, 0) and minimum value of BIC is for the model (3, 0, 0). If the minimum value of AIC and BIC gives different models then we will choose the one with lower minimum value for BIC.

This is shown in the table below:

ARIMA	AIC	BIC	Log Likelihood
(0,0,3)	-4831.07	-4806.31	2420.537
(0,0,5)	-4833.24	-4798.57	2423.622
(3,0,0)*	-4831.87	**-4807.1**	2420.936
(5,0,0)	**-4833.41**	-4798.73	2423.703
(3,0,3)	-4830.23	-4790.6	2423.117
(5,0,5)	-4827.525	-4768.081	2425.762
(3,0,5)	-4829.35	-4779.81	2424.674
(5,0,3)	-4830.01	-4780.47	2425.003

Table 5

After doing the analysis of minimum value of AIC and BIC criterion, we come to the conclusion that ARIMA model (3, 0, 0) fits the data best. It is represented by following equation:

$Y_t^{*} = 0.0010089 + 0.0584131Y_{t-1}^{*} + 0.0093046Y_{t-2}^{*} - 0.0894247\, Y_{t-3}^{*}$

Where Y_t^{*} represents the simple returns series.

4.3 Diagnostic Checking

In order to find whether the given model fits the data best or not we will have to undergo a diagnostic check to be doubly sure of whether the model fits the data well or is there some other model which can fit the data well.

For this we find the residuals series and see if it is white noise or not.We plot the ACF and PACF of the residual series below:

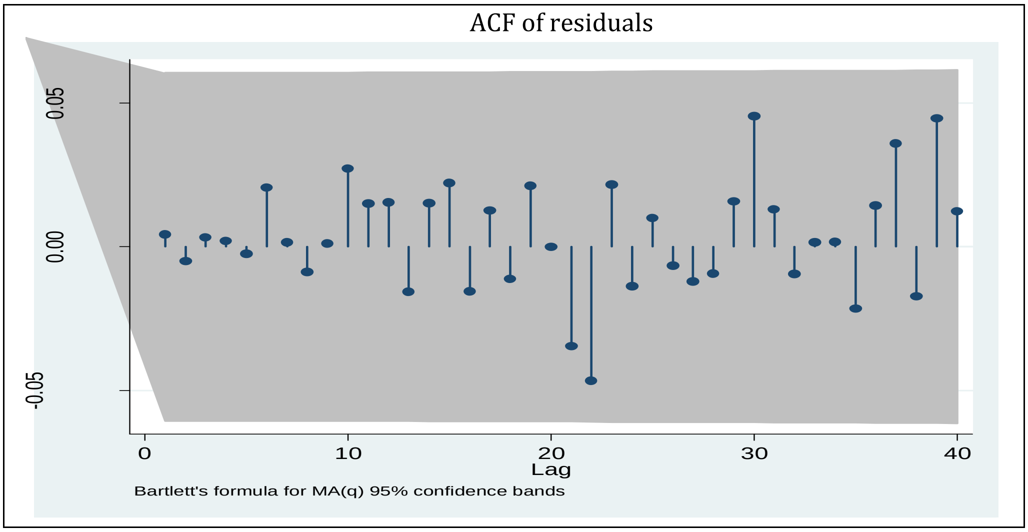

Fig 8

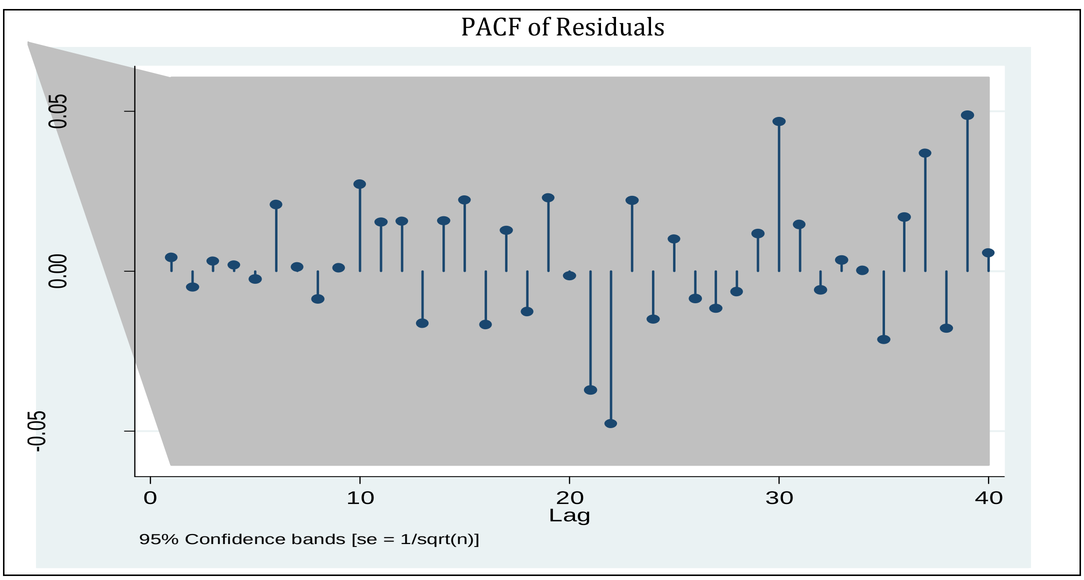

Figure 9

From the figure we see that ACF and PACF of residuals lie within the 95% confidence interval band confirming that model is best fit and residuals follow white noise process.

We also find the Portmanteau test

$H_o: \rho_k = 0$ for all k=1,2,3,...............

$H_a: \rho_k \neq 0$

We find the Portmanteau test and see if the series is white noise or not.

Portmanteau (Q) statistic	16.2174
Prob > chi2(40)	0.9997

Since p-value is more than 0.05 this implies we do not reject the null hypothesis that residuals are white noise.

Now since we know that residuals follow white noise process we go to next step of forecasting.

4.4 Forecasting

We try to predict the value of next seven periods based on the estimated model of AR(3) process.

On the basis of that, we find that actual and predicted values are close to each other.

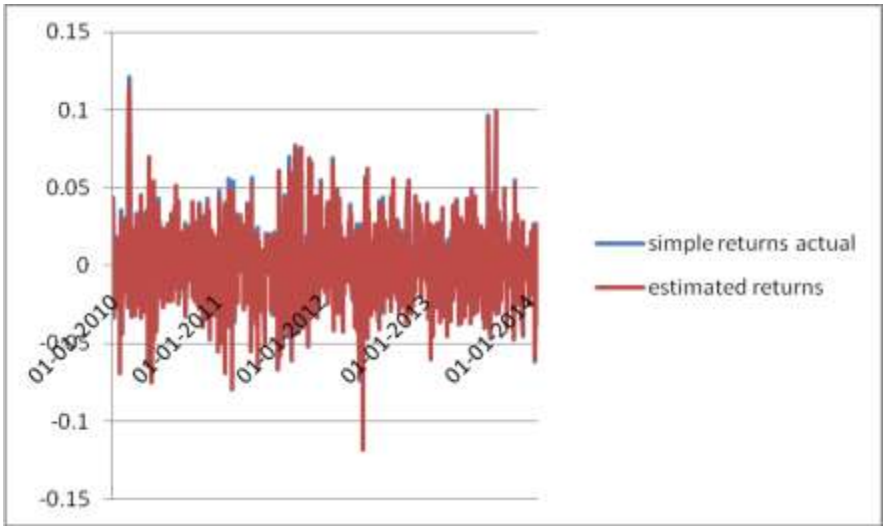

Figure 10

There is a very small error or □ difference between actual and predicted values.

The above figure shows that the actual and estimated series of returns almost coincide. The only discrepancy between the two is due to some random error term

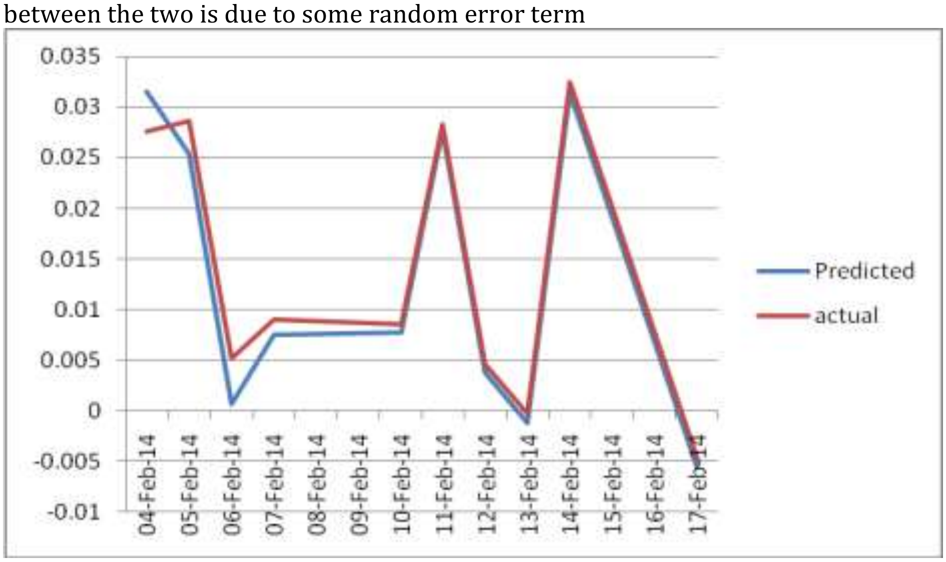

Figure 11

As seen from the figure the actual value is very close to predicted value, difference implies the error term which is almost negligible.

The following table shows the predicted value of next ten periods based on the model and also shows the mean square error:

Date	Predicted Value	Actual Value	Mean Square Error
04-Feb-14	3.148	2.7629	0.14830201
05-Feb-14	2.5358	2.8621	0.10647169
06-Feb-14	0.0627	0.52	0.20912329
07-Feb-14	0.7415	0.9087	0.02795584
10-Feb-14	0.771	0.00859	0.007744
11-Feb-14	2.7608	2.8297	0.00474721
12-Feb-14	0.3746	0.4675	0.00863041
13-Feb-14	-0.128	-0.027	0.010201
14-Feb-14	3.1414	3.2451	0.01075369
17-Feb-14	-0.566	-0.464	0.010404

Table 6

As seen from the table, the predicted value is close to the actual value and also the mean square error is also close to zero indicating that the estimated model is a perfect fit for the series of Tata Motors.

Date	Predicted Value	Actual Value
04-Feb-14	347.196	345.9
05-Feb-14	356.00019	355.8
06-Feb-14	356.2232	357.65
07-Feb-14	358.86	360.9
10-Feb-14	361.62	364
11-Feb-14	371.603	374.3
12-Feb-14	372.995	376.05
13-Feb-14	372.5175	375.95
14-Feb-14	384.2197	388.15
17-Feb-14	382.045	386.35

Table 7

The above table shows the forecasted and actual value of price series. As observed from the table the predicted and actual values are close to each other.

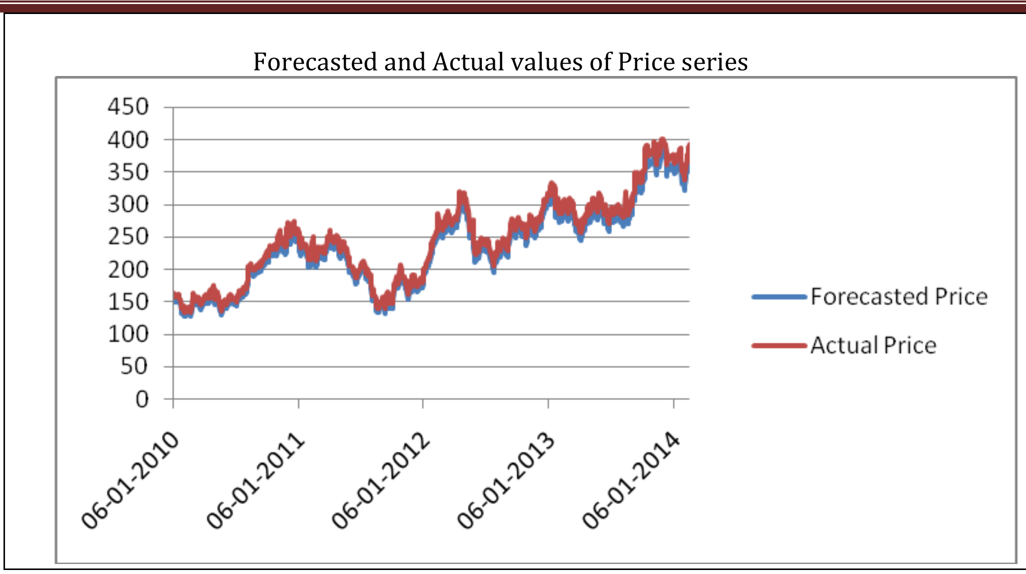

Figure 12

The above figure shows the value of forecasted and actual price series of Tata Motors. As seen from the figure the predicted and actual values lie close to each other, although it cant be perfectly forecasted.

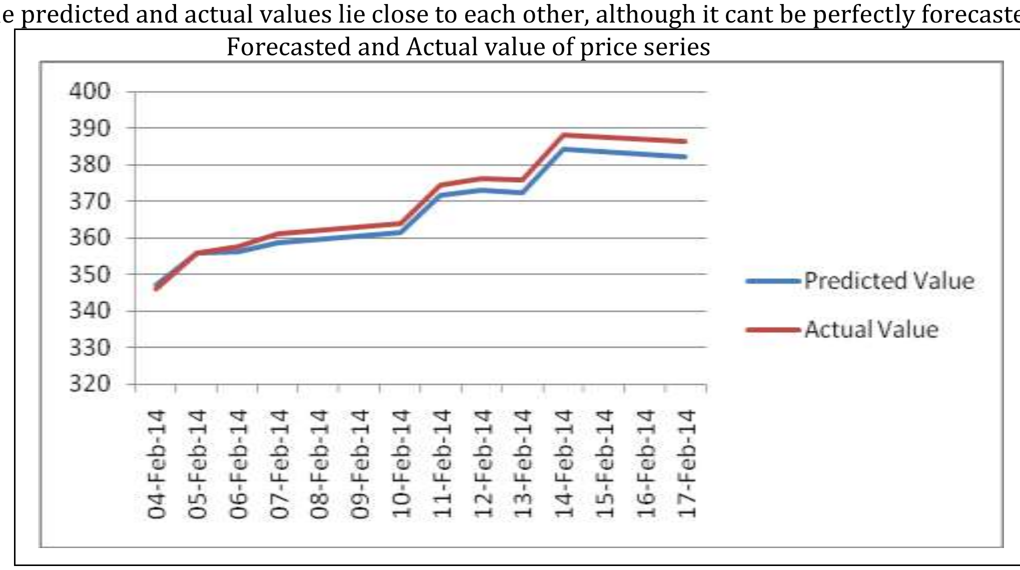

Figure 13

The figure above shows the forecasted and actual values of next ten days. As observed from the table the actual and predicted values are not exactly the same.

5. CONCLUSION

Although we have forecasted values for the next period but there is accurate precision. There is some level of error .If one could forecast tomorrow's price correctly on the basis of today's and past period's price, then one could become billionaire.

The forecasted values, however are not very close to the observed value, but tend to follow the trend of the observed prices

There is one interesting finding after analysing the data that the forecasts obtained from 6 months (small period data) have more precision in forecasting than those obtained by taking 4 years(long period).

APPENDIX

This table shows the ARIMA model for (5,0,0)

Simple Returns	Coefficient	Standard Error	z	P>|z|	[95% Conf. Interval]	
constant	.0010124	.0006795	1.49	0.136	-.0003193	.0023441
Lag1	.0555965	.0266405	2.09	0.037	.003382	.107811
Lag2	.0034277	.0301051	0.11	0.909	-.0555771	.0624326
Lag3	-.0878218	.0305687	-2.87	0.004	-.1477355	-.0279082
Lag4	-.0138866	.028647	-0.48	0.628	-.0700336	.0422604
Lag5	-.0705865	.0288566	-2.45	0.014	-.1271444	-.0140286

References

1. David G. Loomis and James E. Cox, Jr.*"A Course in Economic Forecasting: Rationale and Content",* The Journal of Economic Education,.(31), No. 4 (Autumn, 2000), pp. 349-357
2. Time Series Analysis by James D Hamilton
3. The Econometric Modelling of Financial Time Series by Terence C Mills
4. www.bseindia.com
5. http://in.finance.yahoo.com/q/hp?s=TATAMOTORS.BO