Predicting next day's movement (up or down) for EUR/USD currency pair.

Below data is used for analysis and predict the price of the next day.



The Open, High, Low, Close, Adjust all have high level of correlation. To build an efficient model, only adjusted values are taken.

Both the below models are tried on the data –

* Generalized Linear model
* ARIMA model.

In Generalized Linear model, the accuracy is (75.61%) but based on Kappa & NIR. This model is not a better fit model.

***So, ARIMA model is built on top of data. Using it we can able to minimize the error and we can able to predict that the price moves in the downward direction.***

**Code – R programming**



**Data Inference :-**

* Open, High, Low, Close, Adjust all have high level of correlation. To build an efficient model, only adjusted values are taken.

**## Open High Low Close Adj.Close  
## Open 1.0000000 0.9993069 0.9942836 0.9997704 0.9997704  
## High 0.9993069 1.0000000 0.9941038 0.9992505 0.9992505  
## Low 0.9942836 0.9941038 1.0000000 0.9942958 0.9942958  
## Close 0.9997704 0.9992505 0.9942958 1.0000000 1.0000000  
## Adj.Close 0.9997704 0.9992505 0.9942958 1.0000000 1.0000000**

* “Adjusted Price” data is negatively Skewed but almost close to 0. The data is slightly not normal.

Kurtosis of the data indicates the data have fat tail.

**-------------------------------------------------------------------------------------------------**

**1 - diff(log(EUR\_USD\_Price$Adj.Close)) (numeric)**

**length n NAs unique 0s mean meanCI'**

**2'342 2'342 0 2'327 15 -0.000096420 -0.000336006**

**100.0% 0.0% 0.6% 0.000143166**

**.05 .10 .25 median .75 .90 .95**

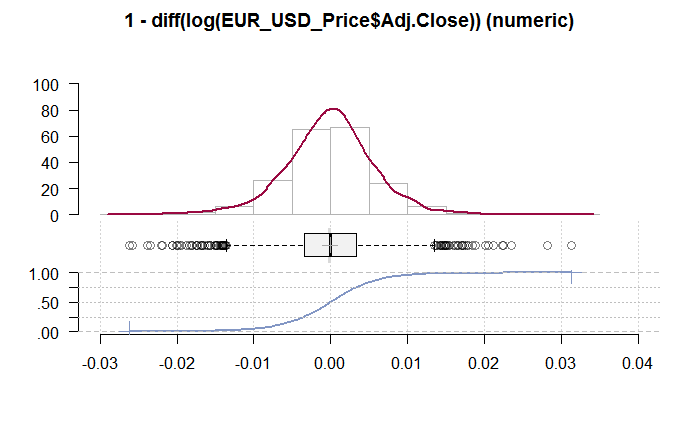
**-0.009662871 -0.007060157 -0.003440378 0.000000000 0.003331970 0.006670400 0.009389041**

**range sd vcoef mad IQR skew kurt**

**0.057457513 0.005912655 -61.321695779 0.005038434 0.006772348 -0.058799800 1.832081438**

**lowest : -0.026195188, -0.025767007, -0.023817388, -0.023396148, -0.022057223**

**highest: 0.022340412, 0.022434959, 0.023494186, 0.028145351, 0.031262325**

* Plot on the Log difference of Adjusted price data. 
* From OHLC chart, it’s clear that there is a drop in the price movement.



**Linear Regression:-**

**Strategy used:** *If Price on day t is more than the price on day t-1, buy else sell.*

* Using the strategy, the direction of the adjusted price is found for the current day.

names(EUR\_USD\_newdata\_LR)<-c("Adjusted","Direction")

* Train/Test data split for the model.

train\_data<-EUR\_USD\_newdata\_LR[1:2300,]  
test\_data <-EUR\_USD\_newdata\_LR[2301:2342,]

* Co-efficient generated for the Linear model(Y=mX+C).

M = 0.03013

C = 1.77953

**Coefficients:**

**Estimate Std. Error z value Pr(>|z|)**

**(Intercept) 0.03013 0.05037 0.598 0.55**

**Adjusted 1.77953 0.08222 21.645 <2e-16 \*\*\***

* Confusion matrix on the test data results with 75.61% accuracy.

Accuracy: 75.61%

No Information Rate : 0.5122

P-Value [Acc > NIR] : 0.00122

**Accuracy is significantly different from NIR. Hence there is imbalance in the data.**

**Reference**

**Prediction 0 1**

**0 14 3**

**1 7 17**

**Accuracy : 0.7561**

**95% CI : (0.597, 0.8764)**

**No Information Rate : 0.5122**

**P-Value [Acc > NIR] : 0.00122**

**Kappa : 0.5142**

**Mcnemar's Test P-Value : 0.34278**

**Sensitivity : 0.8500**

**Specificity : 0.6667**

**Pos Pred Value : 0.7083**

**Neg Pred Value : 0.8235**

**Prevalence : 0.4878**

**Detection Rate : 0.4146**

**Detection Prevalence : 0.5854**

**Balanced Accuracy : 0.7583**

**'Positive' Class : 1**

* Below is the prediction on the test data

signal <- as.data.frame(as.numeric(as.character(pred)))

signal<-na.locf(signal)

summary(signal)

* **The Signal for the 2343 day is ‘1’.**

as.numeric(as.character(pred))

1 1

2 0

3 1

4 1

5 0

6 0

7 0

8 0

9 1

10 1

11 1

12 1

13 1

14 0

15 1

16 1

17 0

18 0

19 0

20 1

21 1

22 0

23 1

24 0

25 1

26 1

27 1

28 0

29 0

30 1

31 0

32 0

33 1

34 1

35 1

36 1

37 0

38 1

39 1

40 0

41 1

**Based on Kappa and NIR, this model is not a better fit.**

**ARIMA model:-**

* Correlation Test on “Adjusted Closing Price”
  + Data is highly correlated data based on the historical data
  + PACF indicates that Significant correlations at the first or second lag, followed by correlations that are not significant. Hence auto-regression is present in the data.

|  |  |
| --- | --- |
|  |  |

* Test for Normality

P-value < 2.2e-16 . P < 0.5

Hence the data is statistically significant. Ie., Auto-correlation present in the data.

**Title:**

**Jarque - Bera Normalality Test**

**Test Results:**

**STATISTIC:**

**X-squared: 330.37**

**P VALUE:**

**Asymptotic p Value: < 2.2e-16**

* Test for stationary

The augmented Dickey–Fuller (ADF) statistic, is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit root at some level of confidence.

Hence there is stationary in the data.

* From Ljung Box test, its again evident that auto-correlation present in the data.

P-value < 2.2e-16 . P < 0.5

Hence Auto-correlation present in the data.

**Box-Ljung test**

**data: EUR\_USD\_Price\_ARIMA**

**X-squared = 40.692, df = 20, p-value = 0.004079**

* Best model turns out to be - ARIMA(1,0,0) among of the below analysis.
  + The best model seems to be with Auto-regression and no Moving average.
  + RMSE value is very minimal 0.005899829. So the error in model prediction would be less

Fitting models using approximations to speed things up...

ARIMA(2,0,2) with non-zero mean : Inf

ARIMA(0,0,0) with non-zero mean : -17382.7

ARIMA(1,0,0) with non-zero mean : -17493.76

ARIMA(0,0,1) with non-zero mean : Inf

ARIMA(0,0,0) with zero mean : -17384.08

ARIMA(2,0,0) with non-zero mean : -17331.11

ARIMA(1,0,1) with non-zero mean : Inf

ARIMA(2,0,1) with non-zero mean : Inf

ARIMA(1,0,0) with zero mean : -17495.71

ARIMA(2,0,0) with zero mean : -17333.02

ARIMA(1,0,1) with zero mean : Inf

ARIMA(0,0,1) with zero mean : Inf

ARIMA(2,0,1) with zero mean : Inf

Now re-fitting the best model(s) without approximations...

ARIMA(1,0,0) with zero mean : -17389.77

Best model: ARIMA(1,0,0) with zero mean

Series: EUR\_USD\_Price\_ARIMA

ARIMA(1,0,0) with zero mean

Coefficients:

ar1

-0.0666

s.e. 0.0240

sigma^2 estimated as 3.482e-05: log likelihood=8696.89

AIC=-17389.77 AICc=-17389.77 BIC=-17377.58

Training set error measures:

ME RMSE MAE MPE MAPE MASE ACF1

Training set -0.0001046289 0.005899829 0.004400612 NaN Inf 0.6913262 0.003266312

* Prediction of next 10 day data (predict(model1,10))

**The negative value indicate that the data is going in downward direct. This means the price goes downward for the next 10 days.**

**Error in prediction**: 0.005901472 / 1.000512 = 0.005898452 ie., 0.58% which is even less than 1

$pred

Time Series:

Start = 17894

End = 17903

Frequency = 1

[1] -1.000512 -1.000069 -1.000098 -1.000096 -1.000096 -1.000096 -1.000096 -1.000096 -1.000096

[10] -1.000096

$se

Time Series:

Start = 17894

End = 17903

Frequency = 1

[1] 0.005901472 0.005914630 0.005914688 0.005914688 0.005914688 0.005914688 0.005914688

[8] 0.005914688 0.005914688 0.005914688

* Logarithmic Forecast of the model for the next 10 days.

forecast1 <- forecast(model1, h = 2)

Point Forecast Lo 80 Hi 80 Lo 95 Hi 95

17894 -5.116064e-04 -0.008074648 0.007051435 -0.01207828 0.01105507

17895 -6.855139e-05 -0.007648454 0.007511351 -0.01166101 0.01152391

* Last price from the excel : 1.143105

Based on the Forecast, its 95% confident that the value might be 1.105.

**Hence the price moves in downward direction.**