*Abstract --* ***Price of currency/stocks/commodities and other financial assets change rapidly because of economic factors. These prices when captured for each moment give rise to huge set of data which is not only difficult to do analysis but also to store and manage. The data generated each second is stored with the help of Big Data and visualization techniques like bar charts, line graph, curves, candlesticks help us in performing analysis. One can take help of tools like RStudio, SASminer, etc. to apply analysis techniques on such dataset and come up with a predictive model to know about the pattern/trend of fluctuations happening. Analysis techniques which are popular now days are regression, clustering, decision trees, association rule mining, text mining etc.***

*Keywords* – ***Analysis, Data, Metadata, Time series, Data mining, R, Currency, ARIMA, STL, Regression, USD, INR, EUR, Forwarder, Forecast, Trend, Prediction, Visualization, Perception, Detect, Market, Model, Graph.***

1. Introduction

Currency is an essential part in today’s world through which you are able to do purchase/sell/trade stuff. Now days currency price is tracked and fluctuated daily depending on the various factors of a particular country. US being one of the strongest economy and popular currency, so we would be keeping US dollars ($) as base currency to compare with Indian (INR) and European currency (EUR). We would also be comparing EUR with INR to visualize the difference in the fluctuations happening between these currencies with time. Analyzing the scenario and historical data through data analytics, data visualization can be very efficient to an investment organization or an individual as it gives you the idea and the knowledge to invest at right time in right place. We have provided efficient decisions based on our analysis and perception after visualizing the data in order to predict values of currency for next months. Also we have rectified upper and lower bounds for these currencies to avoid huge losses in case of some mishappening. Comparing currencies with each other provides factor on which 1 can make the decision, which is most beneficial currency to invest in? We have achieved this goal by identifying attributes, different models and choosing R Studio (based on R language) to perform analysis and visualize the given data.

1. BACKGROUND

Fluctuations in exchange rates are usually caused by actual monetary flows as well as by expectations of changes in monetary flows caused by changes in gross domestic product (GDP) growth, inflation (purchasing power parity theory), interest rates (interest rate parity, Domestic Fisher effect, International Fisher effect), budget and trade deficits or surpluses, large cross-border M&A deals and other macroeconomic conditions. Crisis in United States in 2008 and crisis in European currency in 2014 impacted a lot of currency traders.

1. PROPOSED SYSTEM

R is a highly flexible language that helps in performing not only data analysis but also visualizing these datasets through different plots and packages available. R is known for performing time series analysis on datasets that deal with variation in data according to time and has different models built by renowned programmers and researcher.

Some popular time series models are GARCH, ARFIMA, ARIMA, STL, ARMA, AR (auto regressive), Structure Model, KFKSDS (KALMAN filter model), ZIN (Zero Inflated Model), SETAR, LSTAR and various other models. In our system, we have recognized the trend in behavior of currency conversion rate and used performed time series analysis by using ARIMA and STL models for USD:EUR, EUR:INR, and USD:INR currency datasets.

1. Data Collection

2. Identifying Attributes

3. Model Selection

4. Data Visualization

5. Currency Comparison

*1. Data Collection:* Data collection was divided into two parts.

1.1 Collection of old dataset.

The velocity at which the data for currency fluctuation gets generated it would have been extremely expensive to store these datasets without Big Data (Cloud Technology). Evolvement in the technology has given rise to a large number of websites and organizations that are providing forex services and storing this data. Few of the services are open source while others involve cost to get these type of data. We have collected historical daily trade datasets from one of the open source websites available, ranging as: -

* USD:EUR - 01-03-2005 to 06-20-2016
* USD:INR - 01-03-2005 to 06-19-2016
* EUR:INR - 05-12-2009 to 06-20-2016

1.2 Collecting real time data.

For collecting real time data we developed a forwarder in the form of windows service. The forwarder was running on the server machine and it was collecting real time data after every 2 minutes and saving it into a relational database. For collecting real time data we used an open source, YAHOO FINANCE API.

*2. Identifying Attributes:* On every trade day, market opens and closes with an opening and closing amount. Other than opening and closing amount market also has higher and lower limit and different fluctuating rates.

We have performed our analysis on closing price of every trade day. Since there is a new instance of closing price on each day, so we are considering date as an independent variable and closing price becomes the dependent variable.

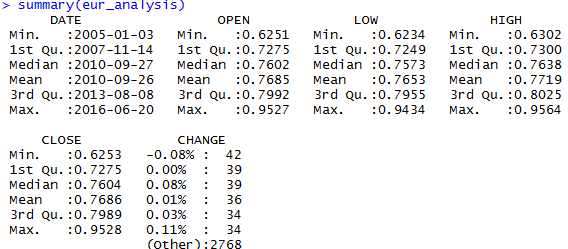


Figure 3.2(a) USD:EUR

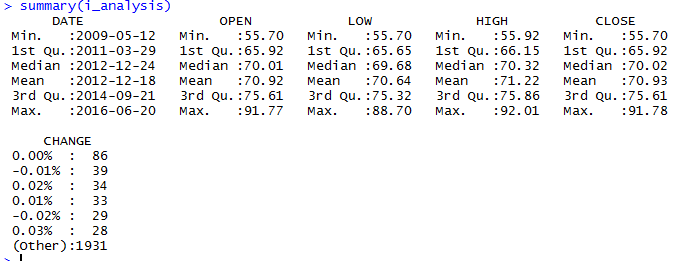


Figure 3.2(b) EUR:INR

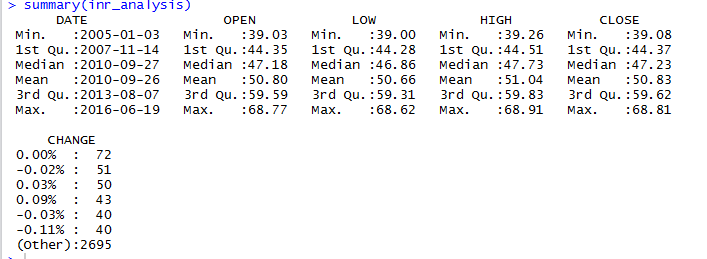


Figure 3.2(c) USD:INR

Summary of dataset for different currency comparisons.

*3. Model Selection:* Currency has tendency to change every second, every minute, day, month, year so we have modeled the historical data on the basis of closing price of each working day. We researched/studied the pattern of currency on various websites and forex platforms thereby recognized that there is a trend being generated according to each date by the currency conversion rate.

We started with basic linear regression analysis in R by generating plot with the least square estimate (**lsfit Model**), of exchange rate relation with date. We selected conversion rate as dependent variable and date as independent variable,

**Conversion Rate = a.Date + b**

(Follows the relation **Y = a.X + b**, where a and b are the coefficients)

R having **ARIMA** as one of the best Auto-Regressive Integrated Moving Average Model being a better option for forecasting than linear regression, became next selection criteria for our analysis. ARIMA was first written in FORTRAN programming language and it is used to find the best fit for past values of a time series. We worked on historical data sets so ARIMA proved to be a valuable choice for performing analysis and formulating moving average along with Upper and Lower Bound range.

Although, **STL** model (seasonal, trend and irregular components using Loess Model) designed by St. Louis was far better than ARIMA in terms of currency comparison but we modeled our dataset on both, since we worked on historical data analysis and forecasting future currency conversion rates.

*4. Data Visualization:* Scatter plot for different currencies, along with linear regression, has been plotted to see the linear relationship and differences between rates for various instances.

From the below graph it can be observed that there is a drastic decrease in USD value from year 2006 to 2008 and decrease in EUR value from year 2014 to mid-2015 thereby proving the fact that crisis occurred in United States and Europe.

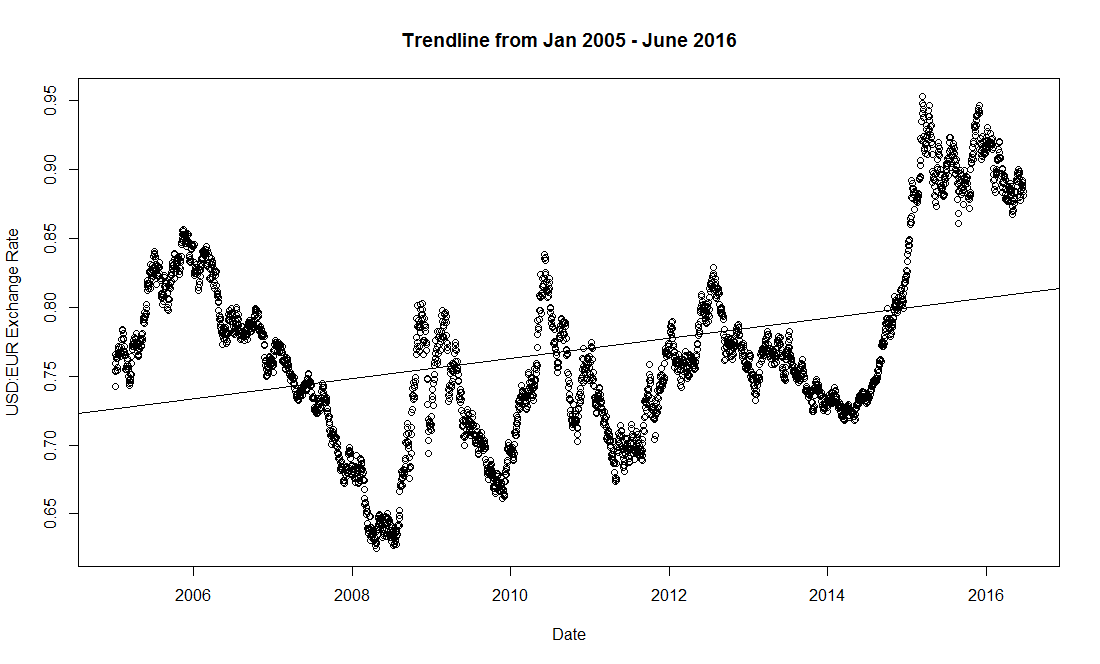


Figure 3.4(a) USD:EUR

In Figure 3.4(b), also weakness in European currency can be observed decrease in EUR value from year 2014 to mid-2015 thereby proving the fact that crisis occurred in United States and Europe.

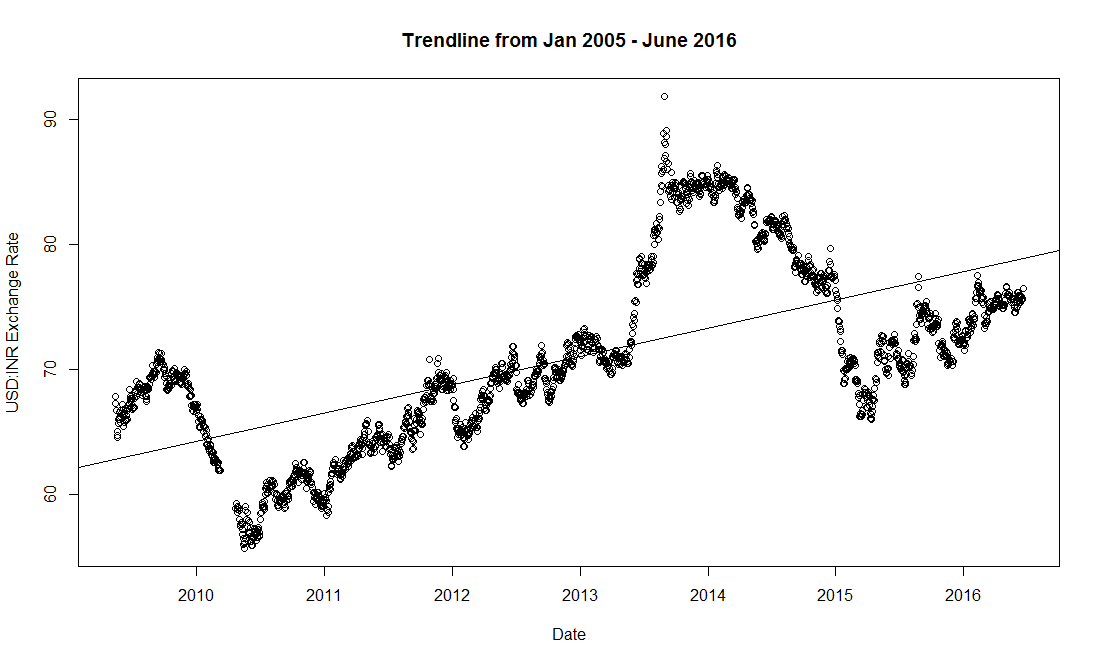


Figure 3.4(b) EUR:INR

US crisis can be seen here also during the 2006-2008 but the proportion of decline/weakness in USD is less compared to that in Figure 3.4(a).

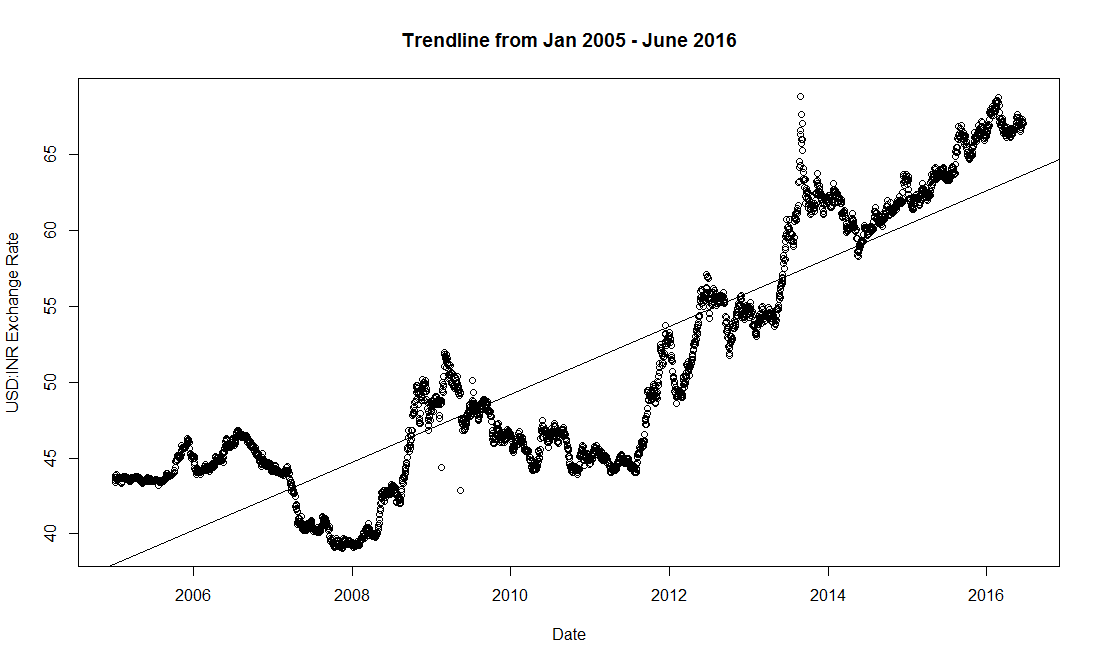
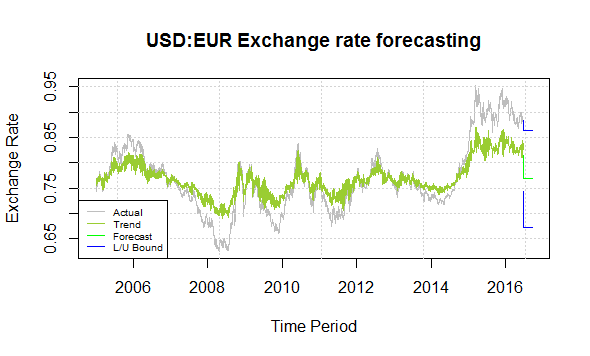


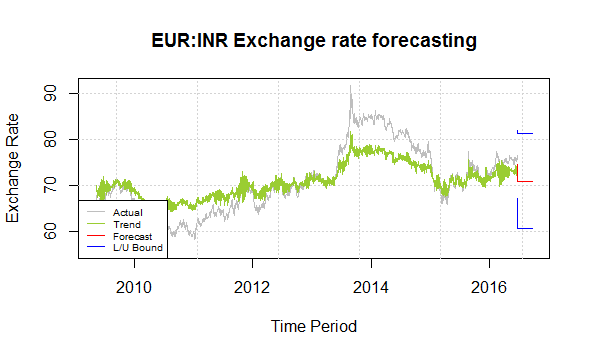
Figure 3.4(c) USD:INR

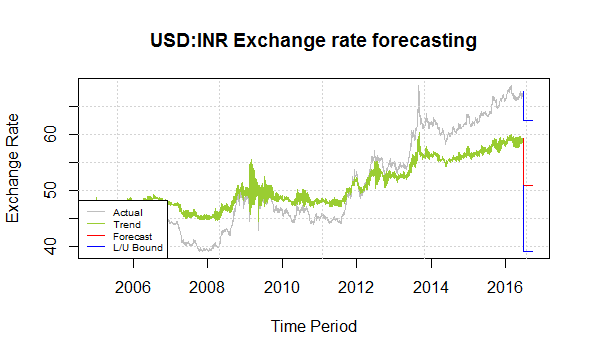
Also we can observe that the conversion rate in case of comparison with INR in figure 3.4(b) and 3.4(c) are close to regression line or the relation drawn with the help of linear regression. According to visual analysis one can recognize that in case of USD vs EUR there were 2 mishappening i.e. crisis in Europe and United States but in other 2 cases there was only 1 i.e. either in Europe and US. One can term this mishappening as an error which is resulting in high deviation from the plotted linear model.

We have also plotted graphs with the help of STL and ARIMA models to get the trend lines for currency conversion rate for different time periods.

**Historical Currency Analysis Using ARIMA Model:**

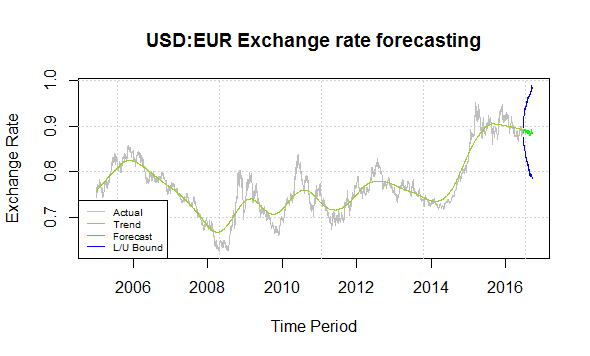


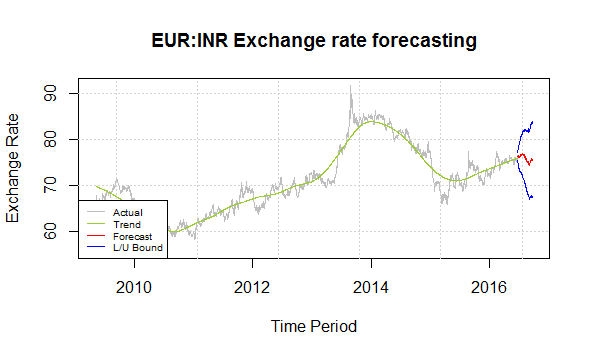


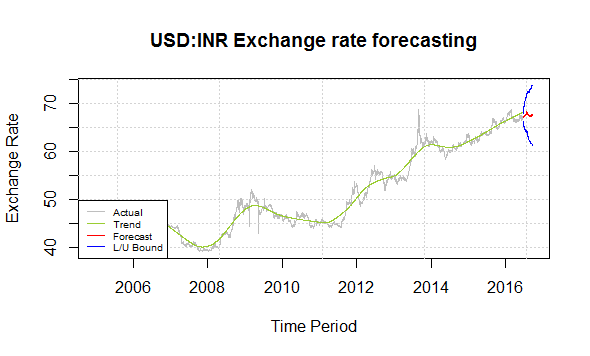


After plotting, both the models we can see that there is a significant difference in upper/lower bound range, along with the plotted trend line for the given time period.

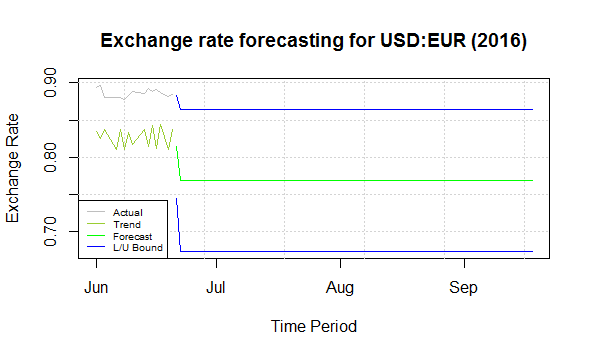
**Historical Currency Analysis Using STL Model:**

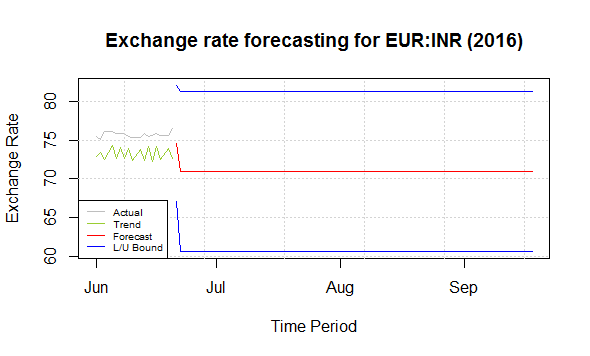


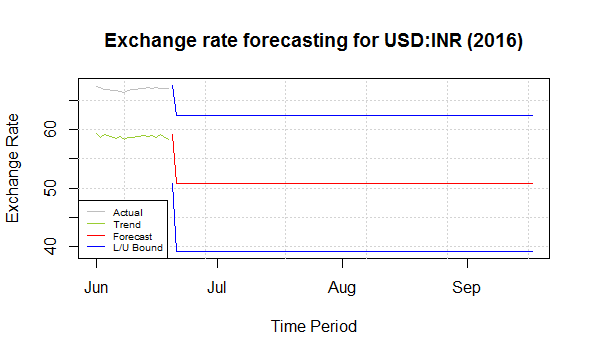




**Current Currency Analysis Using ARIMA Model:**

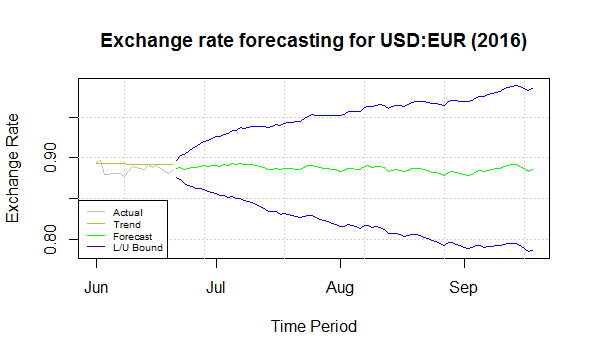


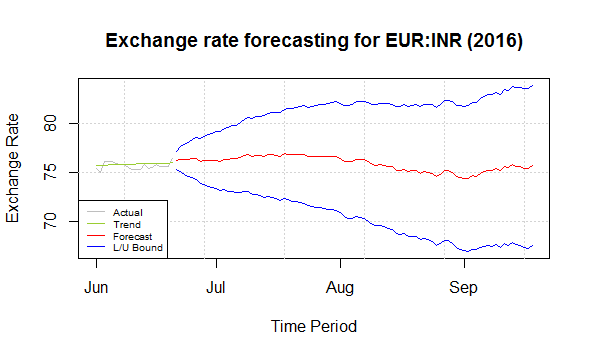


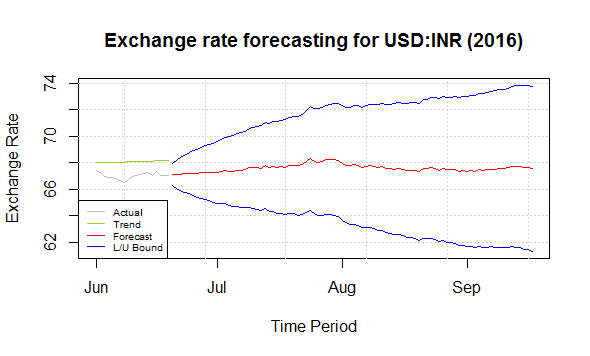


One can clearly see from the plotted ARIMA models for 3 different currency comparisons that the actual is far away from the trend line and the forecast. Also as said earlier that ARIMA is the best fit for historical data but is not good for performing forecasting, we can see a straight line of forecast with 0 slope which shows no variation in exchange rate for different instances.

**Current Currency Analysis Using STL Model:**







On the other hand, one can clearly observe the forecasting through STL. The actual line in case of STL is somewhat similar to the trend which is becoming the base for forecasting and providing profound results. We took buffer by reducing our data set by 3 days to compare actual fluctuations in market. It was found that the tracked real time data and forecasted trend was

IV. CONCLUSION

Data mining along with visual analytics helps in analyzing huge datasets in an efficient manner. It not only saves time but provides accurate results which are useful for making decisions in real life. One should have the ability to rectify the known and the unknown parameters/attributes of any given dataset before choosing any model. The probability of the analyzed and forecasted result depends upon the ability of the selected model to fit for a given particular scenario/dataset in real life. We clearly saw the difference in ARIMA and STL while forecasting the currency conversion rate for given timestamp.

After comparing both the models we came to conclusion that one can perform analysis on the currency dataset using STL in order to forecast trend for next three months. Also one can achieve a model with higher accuracy and forecasting probability by doing future works while dealing with currency.

V. FUTURE WORK

Currently we have performed analysis on the daily closing price of the existing data available to us. In future, we would be performing analysis on the live data set that we are capturing through currency analyzer developed by us. We would be creating a system to track that data into R and apply time series analysis to forecast intraday values.

Right now we have performed analysis on daily “closing” conversion rate on a trade day but while performing research we also analyzed and recognized that the other attributes like volume of currency trade, daily change %, daily high/low, governmental issues, weather etc. can also impact currency of a country. Seeing the scope, we have planned to create a system that will have high probability of forecasting the currency rates considering different parameters.

# VI. REFERENCES

**Conference Papers:**

[1] Chao Luo, Yanchang Zhao, et al. Exception Mining on Multiple Time Series in Stock Market. In Proc. of the 2008 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT 08) Workshops, Sydney, Australia, December 9-12, 2008, pp. 690-693.

[2] Chao Luo, Yanchang Zhao, et al. Outlier Mining on Multiple Time Series Data in Stock Market. In Proc. of the 10th Pacific Rim International Conference on Artificial Intelligence (PRICAI 08), Hanoi, Vietnam, December 15-19, 2008, pp. 1010-1015.

**Books:**

[3] An Introduction to Statistical Learning with Applications in R Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani.

**Online Resources:**

[4] http://finance.yahoo.com/

[5] http://www.investing.com/currencies/

[6] http://www.r-bloggers.com/

[7] https://www.kaggle.com/

[8] http://www.r-chart.com/

[9] http://www.rdatamining.com/

[10] http://en.tradimo.com/

[11] https://stat.ethz.ch/R-manual

[12] https://cran.r-project.org/web/views/TimeSeries.html

[13] https://en.wikipedia.org/wiki/Foreign\_exchange\_market