

# Forecasting Techniques for Time Series Financial Data

<sup>[1]</sup>T. Soni Madhulatha, <sup>[2]</sup>Dr. S. Sameen Fatima

<sup>[1]</sup> Academic Consultant, CSE Dept, KUCET, Warangal, Telangana, India

<sup>[2]</sup> Professor, CSE Dept, University College of Engineering Osmania University, Hyderabad, Telangana, India

**Abstract** - This paper focuses on Time Series forecasting techniques to financial time series with an autoregressive integrated moving average (ARIMA) and Seasonal and Trend decomposition using Loess (STL) that estimates nonlinear relationships. The data used are historical currency exchange rates of INR/USD, INR/GBP, INR/EURO and INR/YEN from May 2005 to July 2014 provided by Reserve Bank of India. First, we show the Multivariate Analysis and then perform the forecasting on individual currency. Various plots are used for visualization.

**Keywords:** Exchange rate, Time series analysis, forecasting models.

## I. INTRODUCTION

Data mining techniques are designed for efficient handling of huge amounts of data that are typically multidimensional and of various complex types. Statistical techniques for data analysis have been applied extensively to scientific data as well as to data from economics. Statistical data mining technique regression is used for predicting future values. One of the most prominent research areas of data mining in the present emerging technologies is in finance. In the field of finance the use of data mining application are considered to be stock market forecasting, managing financial risk, trading futures, prediction of exchange rates. Daily data is collected from May 2005 to July 2014 for Analysis and prediction.

The foreign exchange markets of a country provide the mechanism of exchanging different currencies with one and another, and thus, facilitating transfer of purchasing power from one country to another. With the multiple growths of international trade and finance all over the world, trading in foreign currencies has grown tremendously over the past several decades. Since the exchange rates are continuously changing, so the firms are exposed to the risk of exchange rate movements. As a result the assets or liability or cash flows of a firm which are denominated in foreign currencies undergo a change in value over a period of time due to variation in exchange rates.

## II. METHODOLOGY

Analysis of multivariate time-series data is carried to obtain parsimonious models for estimation, to extract "useful" information when the dimension is high, to make

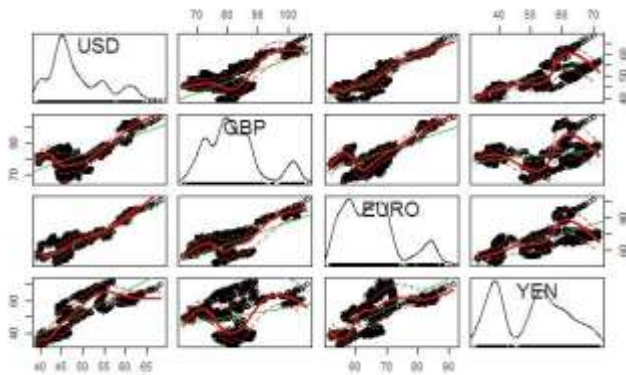
use of prior information or substantive theory and to consider also multivariate volatility modeling and applications the dataset.

*Table 1 Sample dataset of exchange rates*

Date	USD	GBP	EURO	YEN
28/07/2014	60.1013	102.04	80.7385	59
25/07/2014	60.1448	102.2161	81.0173	59.11
24/07/2014	60.002	102.1894	80.6785	59.13
23/07/2014	60.183	102.6842	81.0305	59.34
22/07/2014	60.236	102.865	81.4414	59.32
21/07/2014	60.1853	102.8687	81.4821	59.45
18/07/2014	60.33	103.1824	81.5848	59.54
17/07/2014	60.1401	103.0019	81.334	59.25
16/07/2014	60.1903	103.1361	81.656	59.19
15/07/2014	60.2195	102.7947	81.951	59.27
14/07/2014	60.005	102.7166	81.599	59.18
11/7/2014	60.1855	103.1459	81.872	59.41
10/7/2014	59.875	102.5898	81.6925	59
9/7/2014	59.726	102.3584	81.38	58.79
8/7/2014	59.7968	102.42	81.3225	58.73
7/7/2014	59.9455	102.7826	81.3923	58.71
4/7/2014	59.7939	102.6601	81.3203	58.6
3/7/2014	59.7225	102.4539	81.5415	58.62

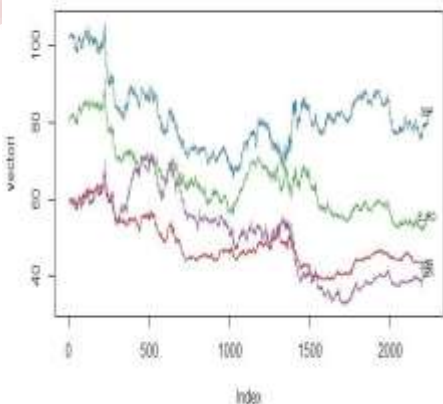
**Multivariate analysis**

Multivariate analysis method constructs various statistics and plots for exchange rates. We use scatterplotmatrix() function to plot the multivariate data. In this matrix scatter plot, the diagonal cells show histograms of each of the variables, in this case USD,GBP, EURO and YEN. Each of the off-diagonal cells is scatterplot of two of the four exchange rates i.e the second cell in the first row is a scatter plot of USD on Y-axis against GBP on X-axis.



**Figure 1: Scatterplotmatrix for exchange rates**

Another type of plot is a Profileplot which shows the variation in each of the variables. It is clear from the profile plot that the mean and standard deviation for GBP variable is higher than the other three variables.



**Figure 2: Profile plot for the dataset**

Next we will apply the statistical function to each column of the dataset. First we will find the mean and next standard deviation.

```
USD GBP EURO YEN
48.384 81.736 64.907 50.417
```

Standard deviation result

```
USD GBP EURO YEN
6.336 8.642 8.667 10.881
```

Correlation coefficient

Pearson's product-moment correlation data: rates\$USD and rates\$GBP

t = 45.473, df = 2240, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0.95  
percent confidence interval:

0.6706676 0.7137552

sample estimates: cor 0.6928293

The correlation coefficient is 0.6928293 which is a weak correlation.

If we apply the most highly correlated function on dataset and print the top 5 pair-wise correlation coefficient.

	First.Variable	Second.Variable	Correlation
9	USD	EURO	0.8912105
13	USD	YEN	0.7755069
15	EURO	YEN	0.7233079
5	USD	GBP	0.6928293
10	GBP	EURO	0.6780547

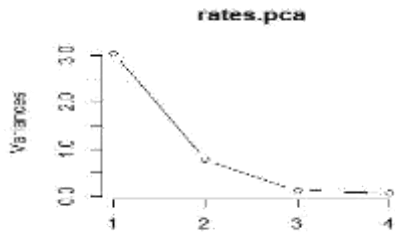
We observe that USD and EURO pair of variables with the highest linear correlation coefficient i.e correlation=0.8912

Finally we carry out a principal component analysis to investigate whether we can capture most of the variation between samples. Hence standardize the variables and apply the principal component analysis function.

Summary of PCA: Importance of components:

```
PC1 PC2 PC3 PC4 Standard deviation 1.7422 0.8771
```

0.35623 0.26186 Proportion of Variance 0.7588 0.1923  
 0.03172 0.01714 Cumulative Proportion 0.7588 0.9511  
 0.98286 1.00000. In order to decide how many principal components should be retained, make screen plot



**Figure 3: Screen plot for exchange rate dataset**

From the plot we can decide that component 1 can be retained.

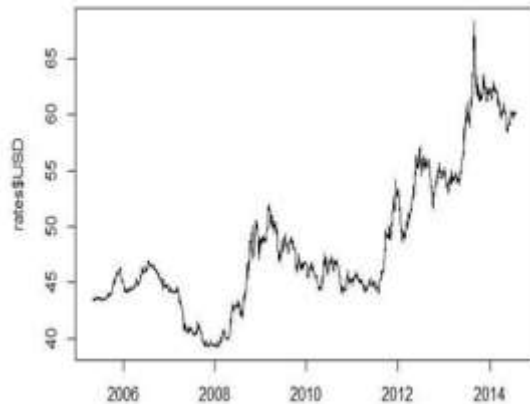
Another way of deciding how many components to retain is to use Kaiser's criterion that we should only retain principal components for which the variance is above 1

PC1 PC2 PC3 PC4 are 3.03521631 0.76931455 0.12689721 0.06857193 respectively

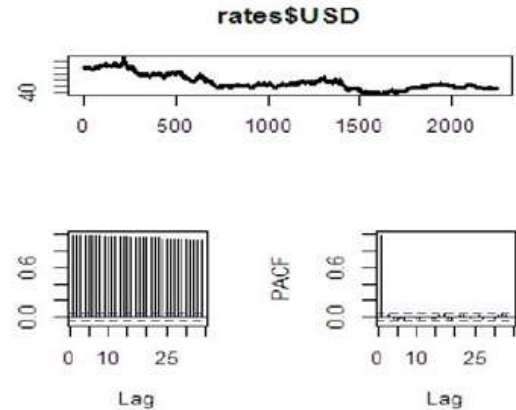
We see that the variance is above one for principal component 1

**Forecasting**

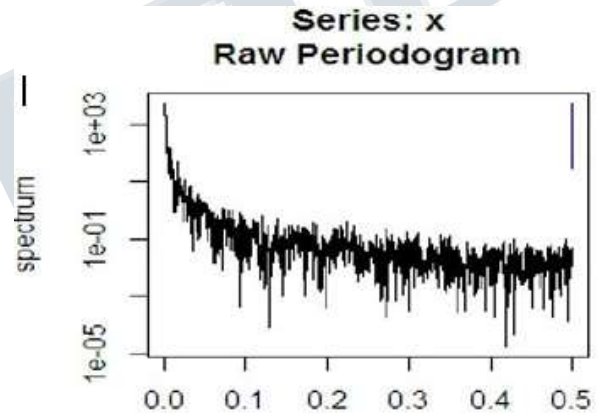
Form the dataset we are going to forecast the INR/USD exchange rates using ARIMA and STL models. The simple plot between the dates and INR/USD are show as



**Figure 4: Simple plot of INR/USD rates**



**Figure 5: Time series display plot for INR/USD rates**



**Figure 6: Periodogram for the dataset**

In order to apply the ARIMA or STL we need to find out the number of days per year which will the window size or frequency used. As per the dataset the frequency/window size calculated is 233.

Code for Forecast with STL model

```
forecastStl <- function(x, y){
myTs <- ts(x$USD, start=1, frequency=233) fit.stl <-
stl(myTs, s.window=233)
sts <- fit.stl$time.series trend <- sts[,"trend"]
fore <- forecast(fit.stl, h=y, level=95) plot(fore)

pred <- fore$mean upper <- fore$upper lower <-
fore$lower
```

```

output <- data.frame(actual = c(x$USD, rep(NA, y)),
trend = c(trend, rep(NA, y)),

#pred = c(trend, pred),

pred = c(rep(NA, nrow(x)), pred), lower = c(rep(NA,
nrow(x)), lower), upper = c(rep(NA, nrow(x)), upper),
dates<-as.Date(x$Date,"%d/%m/%Y") date = c(dates,
max(dates) + (1:y))
)
return(output)
}

```

```

trend <- as.numeric(fore$fitted) pred <-
as.numeric(fore$mean)

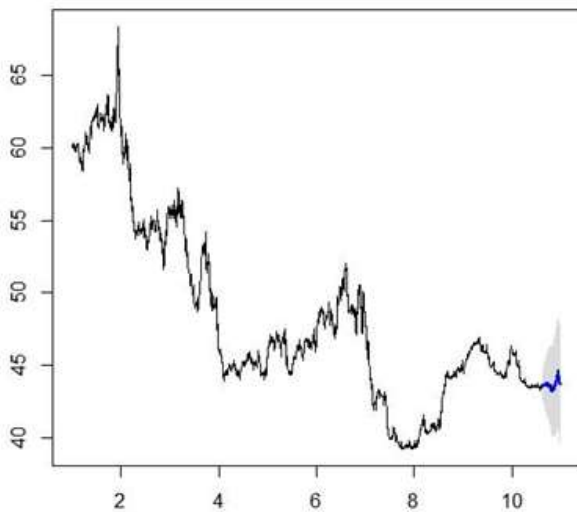
output <- data.frame(actual = c(x$USD, rep(NA, y)),
trend = c(trend, rep(NA, y)),

#pred = c(trend, pred),

pred = c(rep(NA, nrow(x)), pred), lower = c(rep(NA,
nrow(x)), lower), upper = c(rep(NA, nrow(x)), upper),
date = c(x$Date, max(x$Date) + (1:y))
)
return(output)
}

```

**Forecasts from STL + ETS(M,N,N)**



**Figure 7: Forecast plot with STL between years and exchange rate**

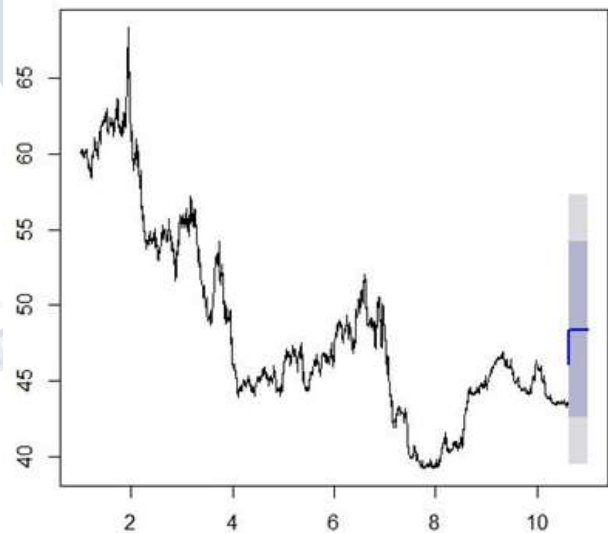
Code for Forecast with ARIMA model

```

forecastArima <- function(x, y){
myTs <- ts(x$USD, start=1, frequency=233) fit.arima <-
arima(myTs, order=c(0,0,1)) fore <- forecast(fit.arima,
h=y)
plot(fore)
upper <- fore$upper[,'95%'] lower <- fore$lower[,'95%']

```

**Forecasts from ARIMA(0,0,1) with non-zero mean**



**Figure 8: Forecast plot with ARIMA between years and exchange rate**

Now we use the code to plot the forecast results

```

plotForecastResult <- function() { resul.stl <-
resul.stl[order(resul.stl$date),]

max.val <- max(c(resul.stl$actual,resul.stl$upper),
na.rm=T)

min.val <- min(c(resul.stl$actual, resul.stl$lower),
na.rm=T)

```

```

rates<-read.table("edata.txt",sep="",header=TRUE)
dates<-as.Date(rates$Date,"%d/%m/%Y")

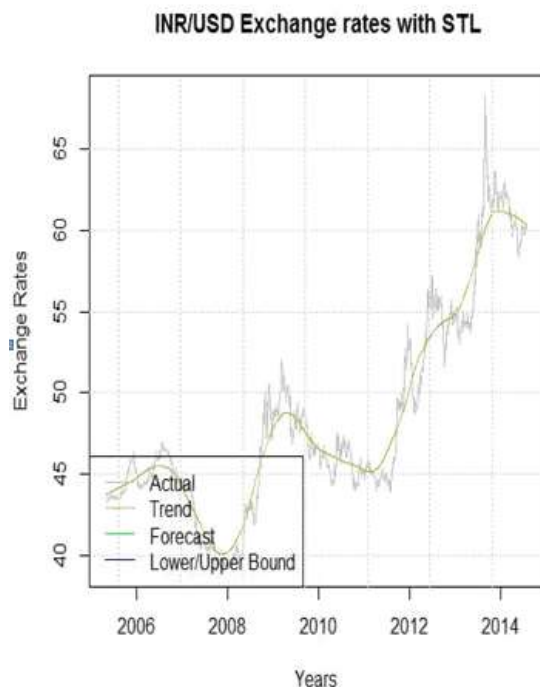
plot(dates, resul.stl$actual[1:2242], type="l", col="grey",
main="Exchange rate forecasting with STL",

xlab="Time", ylab="Exchange Rate", xlim=c("2005-05-
02","2014-07-28"),

ylim=c(39.27,68.3611))
grid()

lines(dates, resul.stl$trend[1:2242], col="yellowgreen")
lines(dates, resul.stl$pred[1:2242], col="green")
lines(dates, resul.stl$lower[1:2242], col="blue")
lines(dates, resul.stl$upper[1:2242], col="blue")
legend("bottomleft", col=c("grey", "yellowgreen",
"green",
"blue"), lty=1, c("Actual", "Trend", "Forecast",
"Lower/Upper Bound"))
}

```



**Figure 9: Forecasting plot for INR/USD**

### III. CONCLUSION

The Multivariate analysis methods present a lot of useful information about the exchange rates. Component retain analysis was performed on summarization. Predictive models show significant results in exploring hidden information. To improve the model more advanced predictive model like Neural Networks can be applied.

### REFERENCES

- [1]. Factors affecting Currency Exchange Rate, Economical Formulas and Prediction Models IJAIEM Volume 3, Issue 3, March 2014 ISSN 2319 – 4847
- [2]. Lillie Lam, Laurence Fung, Ip-Wing Yu “Comparing Forecast Performance of Exchange Rate Models” Research Department, Hong Kong Monetary Authority, Working Paper June-2008.
- [3]. V. Ravi, Ramanuj Lal, N. Raj Kiran, “Foreign Exchange Rate Prediction using Computational Intelligence Methods”, International Journal of Computer Information Systems and Industrial Management Applications. Vol.4 (2012) pp. 659-670.
- [4]. Statistical data mining for forecasting Time series data on foreign currency exchange rates in Global journal of management and technology(GJMT) Special Issue September, 2013
- [5]. Data Mining Techniques For Time Series Data in International Journal of Engineering Sciences Research (ISSN: 2230-8504, e-ISSN: 2230-8512 )
- [6]. Mining inflation rate using predictive and descriptive data mining techniques in IEEE Explore digital library INDIACom-2014.
- [7]. Han, J. and Kamber, M. Data Mining: Concepts and Techniques, 2001
- [8]. <http://www.tradingeconomics.com>