Analysis of Effects of Demonetization on Indian Economy Using Machine Learning Techniques

Rithika Cariappa
Dept. of Computer Science & Engineering
DSATM
Bengaluru, India

Sudhanva R Vasisht
Dept. of Computer Science & Engineering
DSATM
Bengaluru, India

Sujay Dhulekar
Dept. of Computer Science & Engineering
DSATM
Bengaluru, India

Suraj B
Dept. of Computer Science & Engineering
DSATM
Bengaluru, India

Shylaja B
Assistant Professor, Dept. of CSE
DSATM
Bengaluru, India

***Abstract*— The Indian Banknote Demonetization of 2016 was a financial revolution aimed at unearthing black money and corruption rampant in the country. While it was hailed as a visionary action by few, others criticized the move. Nonetheless, this resulted in a huge impact on the Indian economy, some more than the others. In our project, we aim to collect data from various sources before, during and after this time period, like information about stock markets, public responses, RBI reports and so on. The data collected can be classified as Big Data; hence, we will survey for the best method and/or tool to analyse and study this data. We will also make some predictions of future outcome of the parameters and also what could have happened in the absence of the event of Demonetisation. This will help us to come up with results that can help in further studies of Indian economies and help make better decisions in the future.**

***Keywords*—ARIMA Model, Demonetisation, Linear Regression, Time Series Analysis, VAR Model**

I. INTRODUCTION

The Indian Banknote Demonetization of 2016 was a financial revolution aimed at unearthing black money and corruption rampant in the country. This resulted in a huge impact on the Indian economy, some more than the others. In our project, we aim to collect data, related to Foreign Exchange Rates and the factors that affect it, before, during and after this time period from the RBI Data Warehouse.

The data collected can be classified as a time series. Analytics of this time series data can help us in several ways, including uncovering hidden transactions, provide a helping hand for tax administration, checking corruption and terrorism. This will help us to come up with results that can help in further studies of Indian economies and help make better decisions in the future.

To reach the objective, we will make use of Python programming. Python provides a huge number of libraries to work on Big Data. We can also work – in terms of developing code – using Python for Big Data & Time Series much faster than any other programming language. These two aspects are enabling developers worldwide to embrace Python as the language of choice for Data Analytics projects. Some developers say that the performance of Java is better than Python, but it is observed that when working with huge amount of data (in GBs, TBs and more), the performance is almost the same, while the development time is lesser when working with Python on Big Data. The best thing about Python is that there is no limitation to data. We can process data even with a simple machine such as commodity hardware like laptops, desktop and others.

II. LITERATURE REVIEW

Sagar et al mentions that Big Data Analytics will examine the patterns of bank transactions to uncover the hidden fraudulent activities. The cash counting machines equipped with sensors could keep the track of unique number of notes could serve as flood of data that will help to determine the fake currency [2]. A good point to note is that such data can be classified as a Time Series. This is seen in a study by D. Banerjee (2014) where forecast of the future stock market indices is done using time-series ARIMA model. The analysis of the performance of the Indian stock market for six years from January 2007 to December 2012 presented a suitable time-series ARIMA model (1,0,1) which helped in predicting the approximate values of the future indices[1].

Y. Li (2011) et al proposed a model that integrated time-series forecasting and multiple regression analysis to establish a multiple time series regression model. Combination of the two results could obtain good prediction. In addition function of correlation analysis in this system could also assist decision-makers understand the impact about the more direct and quantitative specific factors to the dependent variable. I could help decision-makers filter the most effective, variable factors before making any calculations [4].

C. Ji (2011) cites impulse response function method which based on VAR model and variance decomposition method, to analyze the dynamic relationship among investment, consumption, exports and GDP growth in1979-2009 years in Liaoning province. The simulation results of the impulse response function show that investment, consumption and exports are the important reasons to promote economic growth of Liaoning province; on the other hand, economic growth has an important impact on investment, consumption and exports in Liaoning province [5].

III. OBJECTIVE

With this study, we aim to come at a definitive conclusion that states whether Demonetization of Indian Banknotes in 2016 caused any major changes to factors like Foreign Exchange Rates and predict some future values as well.

IV. BACKGROUND WORK

This research uses two libraries; Scikit-learn and StatsModels for analyzing data in order to create machine learning models. In this research, the Matplotlib & Seaborn libraries are also used to generate data visualizations. We obtain our data from the RBI Data Warehouse regarding Foreign Exchange Rates and factors that affect it.

A. Scikit-learn

Scikit-learn is an open-source library for analyzing data mining. Python is used to analyze and create models from various machine learning algorithms, such as classification, regression, and clustering. Scikit-learn can also be used for preparing data in several ways: normalization, standardization, and cleaning outlier data or missing data [6]. Here we use itto perform univariate Linear Regression on the Foreign Exchange Rates data.

B. StatsModels

StatsModels is a Python module that provides classes and functions for the estimation of many different statistical models, as well as for conducting statistical tests, and statistical data exploration. An extensive list of result statistics is available for each estimator. The results are tested against existing statistical packages to ensure that they are correct. Here we use statsmodels.tsa for time series analysis using the ARIMA and VAR models.

C. Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits.

D. RBI Data Warehouse

In addition to traditional channels such as data publications, reports, press releases, RBI has also set up a public website, 'Database on the Indian Economy' (DBIE) for data dissemination. The DBIE can be accessed through web browsers using the URL, dbie.rbi.org.in

E. Seaborn

Seaborn is a Python data visualization library based on matplotlib and closely integrated with pandas data structures. It provides a high-level interface for drawing attractive and informative statistical graphics. It makes exploring and understanding correlations in data easy in our study.

V. DATA & METHODOLOGY

A. Data Collection & Preparation

Machine learning models in this research use Foreign Exchange data from the RBI Data Warehouse website. In this research, 1 month interval exchange data rates in INR versus USD and factors that affect it from April 1966 to March 2019 for univariate analysis and from April 1990 to March 2019 are focused for multivariate VAR model. The datasets are in CSV files. All the datasets in various CSV files were prepared together cumulatively to be extracted from one source, resulting in a single dataset with 348 rows and 6 columns, including the index 'Month'.

B. Checking best factors affecting Foreign Exchange Rates

From both literature research and Seaborn visualization plots, we were able to choose the best factors that could be used in multivariate analysis models. Figure 1 shows the visualization made using Python.

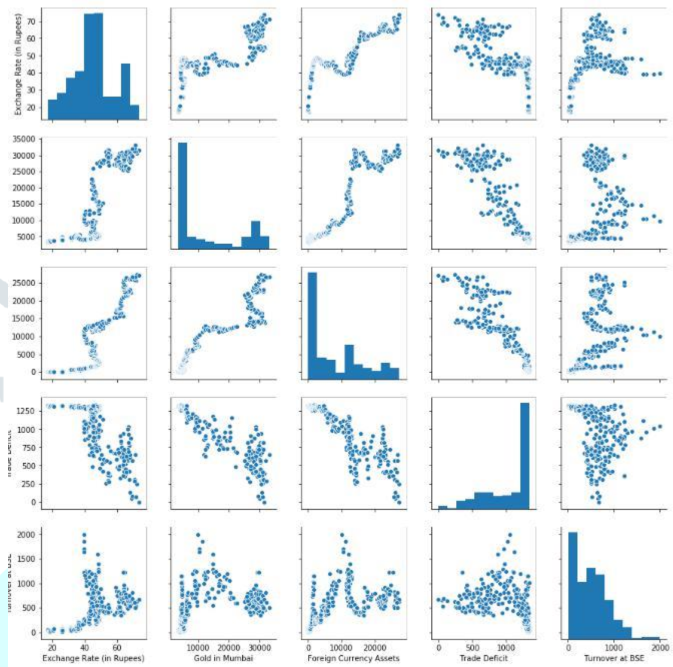

Figure 1. Pairplot to determine best attributes to use in multivariate model

C. Modeling

There are three regression models that we have taken into consideration for analysis of the data available. They are: Linear Regression, ARIMA Model and VAR Model. Our aim here is to see which model gives best analysis and prediction. In order to evaluate the prediction performance, it is necessary to introduce a forecasting evaluation criterion. In this study, the quantitative evaluation used as the accuracy measures is Root Mean Square Error. Root mean Square Error or RMSE can be computed as

$$ \qquad (1)$$

[1] Linear Regression

Given a data set { ,, 1, ... , }1 of n statistical units, a linear regression model assumes that the relationship between the dependent variable y and the p-vector of regressors x is linear. This relationship is modeled through a disturbance term or error variable ε — an unobserved random variable that adds "noise" to the linear relationship between the dependent variable and regressors. Thus the model takes the form

$$ = \quad + \qquad (2)$$

Where = { 1, 2, ..., }1 and = 0, 1, ..., +1

When a Linear Regression model was fitted to the Foreign Exchange Rates with Month as index, it returned a model with R2 score 0.9152221213249985 and RMSE of 34.702270900465294. The resulting equation was:

$$= 0.0034745 \quad - 2498.446289 \qquad (3)$$

This model showed a good R2 score but a better model could be made to fit the data to get a better RMSE value. Data forecasted and actual data can be visualized as seen in Figure 2.

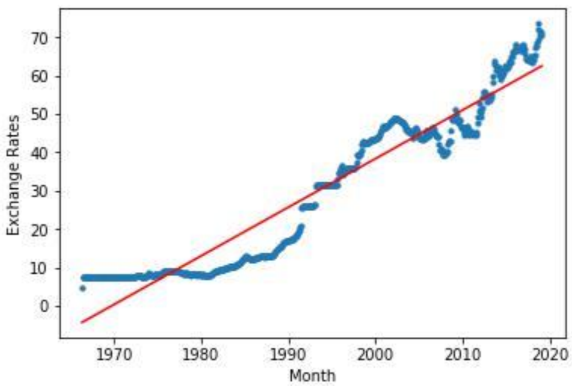

Figure 2: Univariate Linear Regression Model for the data

[2] ARIMA Modeling

ARIMA stands for Auto-Regressive Integrated Moving Averages. The ARIMA forecasting for a stationary time series is nothing but a linear (like a linear regression) equation. How it varies is in its calculation of the values to be predicted. The ARIMA(p,d,q) model follows the equation:

$$(1-\sum \quad {}' \quad _{=1} \quad)(1- \quad) \quad =(1+\sum \quad _{=1} \quad)$$

Given a time series of data where t is an integer index, the set comprises of real numbers, L is the lag operator, are the parameters of the moving average part and are error terms.

Most of the time series models work on the assumption that the time series is stationary. Intuitively, we can say that if a time series has a particular behavior over time, there is a very high probability that it will follow the same in the future. Also, the theories related to stationary series are more mature and easier to implement as compared to non-stationary series. We check for stationarity by plotting rolling statistics and performing a Dickey-Fuller test. It was observed that the time series was nonstationary. We make it stationary by eliminating trend and seasonality by Smoothing and First Order Differencing techniques respectively.

The predictors depend on the parameters (p,d,q) of the ARIMA model. The Auto Regressive terms (p) are just lags of dependent variable. The Moving Average terms (q) are lagged forecast errors in prediction equation. The term d represents the number of nonseasonal differences. In this case we took the first order difference so we can set d=1. To compute p and q we use the Autocorrelation and Partial Autocorrelation functions respectively. On plotting the ACF and PACF we take p=2 and q=2 since the lag value in both scenarios crosses the upper confidence interval for the first time at lag axis value= 2 as seen in Figure 3.

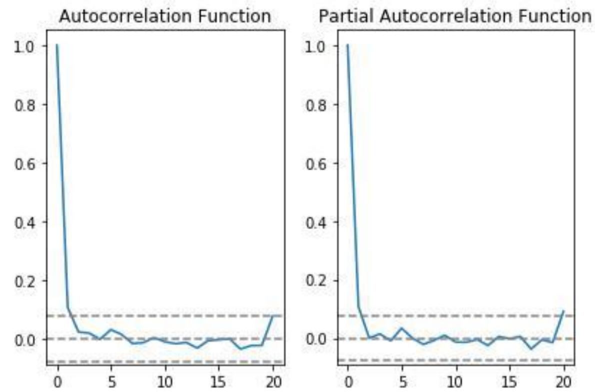

Figure 3: The ACF and PACF plots for the univariate ARIMA analysis

Once the lag values are found, we fit the data to the ARIMA model and take it back to the original scale. Since we did a log transform for removing the trend, we take the exponent values of the predicted values and compare the values to the original values. The predictive model proves a better alternative to the Linear Regression model with a reduced RMSE value of 5.3441 as seen in the Figure 4.

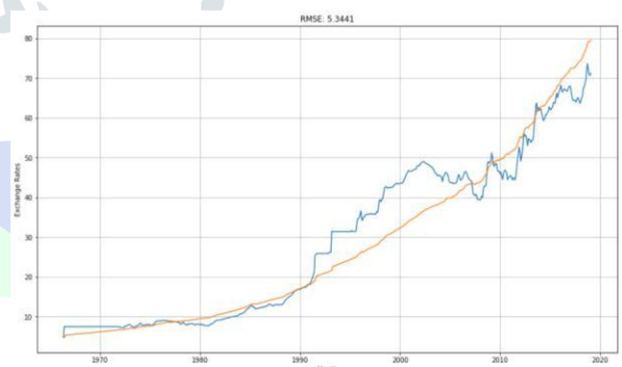

Figure 4: ARIMA(2,1,2) model for univariate analysis of Foreign Exchange Rates

[3] VAR Modeling

A Multivariate time series has more than one time-dependent variable. Each variable depends not only on its past values but also has some dependency on other variables. This dependency is used for forecasting future values. One of the most commonly used methods for multivariate time series forecasting is Vector Auto Regression (VAR). In a VAR model, each variable is a linear function of the past values of itself and the past values of all the other variables. The Var model can be represented mathematically as

$$(\ ) = \quad + \ _1 \ _{-1} + \cdots + \quad _{-} + $$

The term in the equation represents multivariate vector white noise. For a multivariate time series, should be a continuous random vector that satisfies the following conditions:

- $(\ ) = 0$; Expected value for the error vector is 0
- $(\ _1,\ _2') = \ _{12}$;Expected value of εt and εt‘ is the standard deviation of the series

Once again, we need to check the stationarity of this dataset with multiple variables. For a series to be stationary, the Eigen values of |Φ(L)-1| should be less than 1 in modulus. On performing the augmented Dickey Fuller tests, it is observed that the dataset is stationary.

TABLE I. EIGEN VALUES OF |Φ(L)-1| FOR EACH ATTRIBUTE OF THE DATA

Attribute	Eigen Values of |Φ
Exchange Rate (in Rupe	0.26705944
Gold in Mumbai	0.10
Foreign Currency Assets	0.045683
Trade Deficit	0.01
Turnover at BSE	0.00050043

Next, we split the data into training and validation sets. Once this is done, we fit the data to the VAR model and check its prediction performance with the RMSE values. We observe the lowest RMSE value so far, 3.8896. This is illustrated in the Figure 5 below.

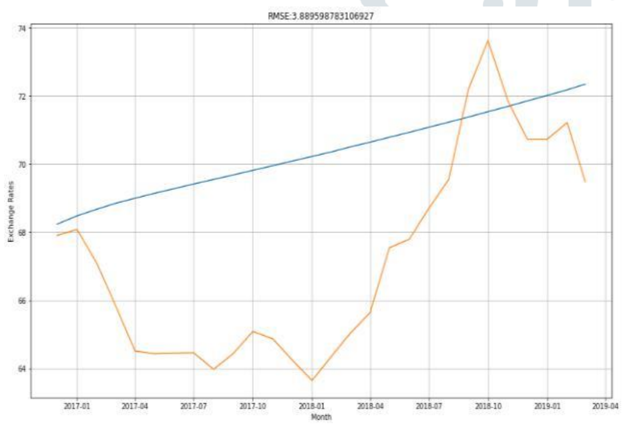

Figure 5. VAR model for multivariate analysis of dataset

This model shows better prediction values as compared to ARIMA and Linear Regression Models. We can also predict future values as shown below in Figure 6.

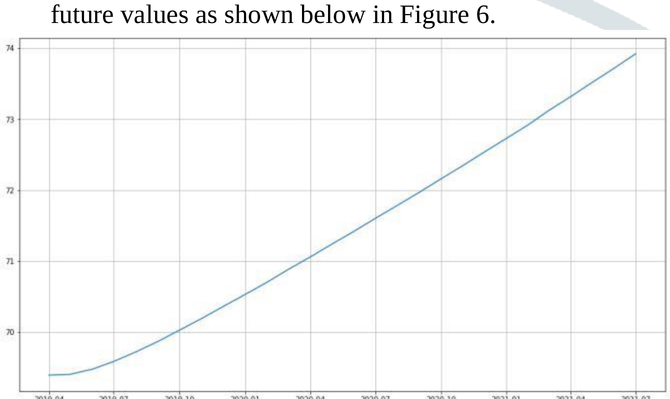

Figure 6. Forecasted Values from April 2019 to July 2021

VI. RESULTS AND DISCUSSIONS

From the above models and datasets, it was clear that VAR multivariate model was the best among all the other models for analysis. This is seen in a quantative manner in terms of RMSE scores as mentioned in Table 2.

TABLE II. ANALYSIS MODELS AND THEIR RMSE SCORES

Analysis Model	RMSE Score
Linear Regression Model	34.7022
ARIMA (2,1,2) Model	5.3441
VAR Model	3.8896

In the VAR model, we took the values till November 2016 to train the model. Post that, which denotes the time period post Demonetisation, we fitted the validation dataset to the model to check if the absence of the event of Demonetisation caused any major changes. It is clear from Figure 5 that while there was a deviation from the trend for few months post Demonetisation, there was actually no major deviations as the data progressed towards the current date. As any other such financial policy change, there was definitely some discrepancies in Foreign Exchange rates but it gradually started fitting the predictive model. So we were able to conclude that while there were some changes due to Demonetisation, it does not affect Foreign Exchange Rates by a lot in the long run. This is also seen with the Root Mean Square Error. Also, in terms of what could happen in the future with Foreign Exchange Rates given the following trend, we can also predict some values with the VAR model with multivariate analysis. This is illustrated in Figure 6 and a sample of these forecasted values is as seen in Table 3.

TABLE III. FORECASTED FOREIGN EXCHANGE RATES

Month	Forecasted Exchange
April 2019	69.3
May 2019	69.4
June 2019	69.4
July 2019	69.5
August 2019	69.7

VII. CONCLUSION

From our study, the goal was to come at a logical statement if Demonetisation caused any major changes to the Economy of India through Foreign Exchange Rates. It was found that like after most major policies, there was some disruption initially. But later on, it started to follow the predicted model and shows that it did not cause much change to the economy in the long run. There may have been other factors contributing to changes in these attributes and socio economic policies which cannot be easily quantifiable. There is no evidence that Foreign Exchange Rates and factors affecting it are perfectly linear. Hence in future we might implement non-linear analysis using soft computing techniques like ANN and fuzzy time series analysis. Forecasting accuracy will be enhanced if we study the probability distribution nature of the same dataset. To train the model better, data should be constantly updated and other features can also be added by checking their correlation with other data attributes.

VIII. REFERENCES

[1] D. Banerjee, "Forecasting of Indian Stock Market using

Time-series ARIMA Model," 2014 2nd International Conference on Business and Information Management (ICBIM),Durgapur, 2014, pp. 131-135

[2] R. K. Sagar and Y. Gupta, "Demonetization: How big data analytics can unearth the black money of India," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, 2017, pp. 67-70.

[3] I. Chandrakar, "Analysis of the effects of currency demonetization in India using Bayesian prediction," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, 2017, pp. 117-122.

[4] Y. Li and F. Ying, "Multivariate Time Series Analysis in Corporate Decision-Making Application," 2011 International Conference of Information Technology, Computer Engineering and Management Sciences,

[5] Chengjun Ji, Jinping Li and Hongxia Zhao, "The long term interactive relationship analysis of influence factor of economic growth in Liaoning province Based on VAR model," 2011 International Conference on Business Management and Electronic Information, Guangzhou, 2011, pp. 92-95.

[6] T. Phaladisailoed and T. Numnonda, "Machine Learning Models Comparison for Bitcoin Price Prediction," 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE), Kuta, 2018, pp. 506-511.

[7] S. A. Fahad and A. E. Yahya, "Big Data Visualization: Allotting by R and Python with GUI Tools," 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), Shah Alam, 2018, pp. 1-8.

[8] F. Gupta and S. Singal, "Sentiment analysis of the demonetization of economy 2016 India, Regionwise," 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, Noida, 2017, pp. 693-696.

[9] Banerjee, Porinita, and Vasimraja Sayyed. "A Study on Impact of Demonetization on E-Commerce Industry." IJEAST, 2 (3), 66 72 (2017).

[10] Sivarajah, P. (2016). Demonetisation: Madurai Corporation makes record tax collection on a single day | Chennai News - Times of India. [online] The Times of India. Available at: https://timesofindia.indiatimes.com/city/chennai/Demonetisation-Madurai-Corporation-makes-record-tax-collection-on-a-single-day/articleshow/55374378.cms