Comparing Macroeconomic Indicator Forecasting in Multivariate Framework

By

Fatima Shad

PIDE2016FMPHILETS15

Supervised by

Dr. Amena Urooj

Co-Supervised by

Ms. Uzma Zia

Department of Econometrics and Statistics

PAKISTAN INSTITUTE OF DEVELOPMENT ECONOMICS, ISLAMABAD

2019

Pakistan Institute of Development Economics

CERTIFICATE

This is to certify that this thesis entitled: **"Comparing Macroeconomic Indicator Forecasting in Multivariate Framework"** submitted by Ms. Fatima Shad is accepted in its present form by the Department of Economics & Econometrics, Pakistan Institute of Development Economics (PIDE), Islamabad as satisfying the requirements for partial fulfillment of the degree in **Master of Philosophy in Econometrics**.

Supervisor:

Dr. Amena Urooj
Assistant Professor
PIDE, Islamabad

Co-supervisor:

Ms. Uzma Zia
Senior Research Economist
PIDE, Islamabad

External Examiner:

Dr. Ismail Shah
Assistant Professor
Department of Statistics
Quaid-i-Azam University, Islamabad

Head,
Department of Economics & Econometrics:

Dr. Karim Khan

DECLARATION

I **Fatima Shad** hereby state that my M.Phil thesis titled **"Comparing Macroeconomic Indicator Forecasting in Multivariate Framework"** is my own work and has not been submitted previously by me for taking any degree from Pakistan Institute of Development Economics or anywhere else in the country/world.
At any time if my statement is found to be incorrect even after my Graduation the university has the right to withdraw my M.Phil degree.

Date: 23.12.2019 Signature of Student

Fatima Shad

The dissertation is to dedicated to

my Parents, my believed Brothers

(for their endless love, support, prayers

and encouragement)

ACKNOWLEDGMENT

First of all, I would like to thank Almighty Allah for giving me the strength to complete this work. I would like to pay gratitude to my supervisor ***Dr. Amena Urooj*** and co-supervisor ***Ms Uzma Zia,*** Pakistan Institute of Development Economics, whose consistent guidance and support was helpful throughout the work. I'm greatly indebted to them for the knowledge they imparted and the encouragement they extended to me.

Beside the supervisor, I would like to special thanks to my Ph.D senior **Faridoon khan** and junior **Shafaq Toori** for helping me in different phase of thesis their encouragement and making the hard things out.

I am very much thankful to my Parents, my Brothers for having faith in me and having blind trust that I will do it and Allhamdullilah I have done that.

Last but not the least, I would like to thanks my friends **Rabail Mir, Lubna Bano, Namal Kinza** and others, my classmate for the stimulating discussion, for the sleepless nights we were working together helping each other out before the deadline and all of the fun and happiness that we have had in last 2 years.

Fatima Shad

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

CPI	Inflation
M2	Money supply
REER	Real Effective Exchange Rate
WR	Worker remittances
ELEC	Electricity
NG	Natural Gas
PC	Production of Cement
PG	Production of Gypsum
PL	Production of Lime Stone
PR	Production of rock Salt
PCO	Production of Crude Oil
PGC	Production of Good Coal
CIC	Currency in Circulation
ERO	Export of Rice
EXP	Total Exports
INT	International liquidity total reserves
IMPP	Import of Petroleum
CP	Cash in Pakistan
IMP	Total Imports
INL	International Liquidity gold holding,
BSP	Balance with State Bank of Pakistan
INS	Goods values of imports, insurance
WPIM	WPI Manufacture
WPIF	WPI Food
FIR	Financial Interest Rate, money market rate
NEER	Nominal Effective Exchange Rate
ERC	Exchange rate of China
ERU	Exchange rate of United Arab Amarat
ERS	Exchange rate of Saudi Arabia
ERU	Exchange rate of USA
ERM	Exchange rate of Malaysia
SBP	State Bank of Pakistan
IFS	International Financial Statistics
RT	Regression Tree
XR	Exchange Rate
DMS	Dynamic Model Selection
DMA	Dynamic Model Average
MSE	Mean Square Error

ABSTRACT

This study attempts finding the best forecasting model for multivariate time series in Pakistan by using the variable selection models like, Regression Tree, Bagging, Random Forest, Boosting, and Adaptive LASSO. The monthly time series data on 31 Macroeconomic variables is collected from State Bank of Pakistan and International Financial Statistics for the period 1990-2017. The performance of each technique is gauged on the basis of forecast error. Further, performance is checked by varying the number of predictors in the model i-e, information set ranging from four, fourteen and thirty one. In this study we evaluate the empirical performance of the competition forecasting method with nested subsets of predictor with different value of k, specifically $I_4 \subset I_{14} \subset I_{31}$. This allows us to investigate the impact of utilizing information set of different sizes. In this study we have found that when we increase the numbers of predictors form four to fourteen the predications found are better from four variables and if the number of variables increased from fourteen to thirty one the results showed prediction improved even further. If there are more than 20-31 variables then there is no improvement in the result.

Key words: Bagging, Regression Tree, Random Forest, Boosting, Adaptive LASSO, Auto Regressive Model.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Forecasting is a used for developing the assumptions about future uncertainty with the help of analyzing the available data (present and past data). It is a designing tool. Forecasting is a necessary planning of economy as it provides the direction to the economist for making policy about economy. Forecasting involves use of different methods from statistical estimation and inference to determine the possible future outcome.

Multiple predictors forecasting is a recent idea, offers the chance to have access to the large set of information rather than the traditional approaches of time series forecasting. Multiple predictors forecasting gives, the researcher the opportunity of robustness for the structure variability that leads to less dimensional forecasting. The major feature of multiple predictors forecasting is that it applies structure to control the estimation error and still more information can be accessed. Another way to elaborate the multiple predictor forecasting is to change the dimensionality challenge into a benefit. The positive attribute of multiple predictors forecasting by considering the range of 20 to 30 variables for macroeconomic variables is explained in this study (Stock and Watson, 2004).

One of the most significant objective of time series is forecasting. From policy point of view reliable forecast of macroeconomic variables is critical. Many forecasting models have been proposed in past to achieve this objective, each with its own merits and demerits. Conventional forecasting models such as regression analysis, time series analysis, moving averages and smoothing methods, the Box-Jenkins methodology, and numerous judgmental methods use previous data to construct a model and to develop the concerned variables. These approaches assume that the future will have the same characteristics as the past, except for those variables which are being used by the model to propose a forecast. These models or approaches have many assumptions and many of them are not verifiable.

Forecasting of macroeconomic variables provides us a tool to sense where the economy may be directed, allowing producers, government, consumers and macroeconomic policy planners to plan in advance with some level of balanced certainty. Forecasting usually reduce the level of uncertainty about the future development of the economy. If various macroeconomic variables and persistent nature of the business cycles are carefully analyzed. Policy makers and macroeconomists may form forecasts by considering statistical regularities as procedure for predicting the economy and for the overall efficiency of the economy. In this study we use multivariate framework of macroeconomic variables for forecasting by using methods like Regression Tree, Bagging, Boosting, Random Forest Adaptive LASSO and Autoregressive model. Improved and timely forecasting of macro economy is a recent need of Pakistan economy as it can provide a dimension to policy makers and will provide input for central bank policy planning, government budget planning and business initiatives.

Recent literature focuses on techniques using multiple predictors for forecasting such as Regression Tree, Bagging, Boosting, Random Forest, Least Angle Regression, Adaptive LASSO. These methods which are mainly variable selection methods select and retain important features and helps in better forecast as well.

Among other techniques for prediction, Regression Trees are widely used. It allows a mixture of continuous and categorical variables as predictors. A Regression Tree is constructed through a process known as binary recursive subdividing, in this process data splits into partitions or branches. Bagging takes many bootstrapped samples and fit tree to each bootstrapped sample (James et al 2013). But Boosting RTs has a distinctive property of being sequential. Multiple predictors forecasting can be specified by using dynamic factor predictors rather to use all the variables (Stock & Watson, 2006). Bootstrap aggregation or ‘bagging’ had been used to forecast the inflation and it gave good results of forecasting by (Inoue and Kilian 2008).

The adaptive LASSO was introduced by (Zou 2006) for linear regression and by Zhang and Lu (2007) for proportional hazards regression. Adaptive LASSO, as a regularization method, avoids overfitting penalizing large coefficients. Further, it has the same advantage that LASSO can shrink some of the coefficients to accurately zero, performing a selection of features with the regularization.

1.2 Significance of study

This study attempts to forecast macroeconomic variables by using multiple predictors forecasting techniques like Regression Tree, Bagging, Boosting, Random Forest and Adaptive LASSO. Study also includes forecasting of Macroeconomic variables by using benchmark methods for forecasting i-e. “Mean Model” and “$AR_{(p)}$”. Many previous studies have discussed above mentioned approaches separately along with the benchmark methods. However our study compares the results of multiple predictors forecasting approaches and makes an attempt to conclude about the technique which have the best forecast results of targeted Macroeconomic variables. Our study to explore the impact of information set size in multivariate forecasting.

Forecasting is done for these 31 variables of Pakistan in order to compare the forecasting results of these methods and to find out best forecasting technique. We have also attempted by seeking different size of information set in case Pakistan to explore the role of size of information set in multivariate forecasting for Pakistan.

1.3 Objectives

The goal of the study is to compare forecasted results of multiple predictors forecasting methods and benchmark methods and to find out the most appropriate forecasting technique for Macroeconomic variables in case of Pakistan. Main objectives of study are:

- To forecast multivariate macroeconomic data set comprising 31 monthly Pakistani Macroeconomic variables.
- To compare the performance of various forecasting methods by using results of multiple predictors forecasting methods in case of Pakistan.
- To reveal the advantage of various information set in multiple predictors forecasting of macroeconomic variables in case of Pakistan.

1.4 Frame work of the study

Thirty one macroeconomic variables of Pakistan are selected. Inflation is chosen as the dependent variable in this study. Since it is a significant indicator of economic wellbeing of a country. Methods are used Regression Tree, Bagging, Boosting, Random Forest, Adaptive LASSO and bench mark method (AR). We intend to do multivariate forecasting by adopting different size of information set i-e, i= 4,14, 31 and by concluded.

1.5 Organization of study

This study is organized into five chapters. Chapter 1 offers introduction, objectives of study and significance of the study. Chapter 2 provides the literature review of both national and international studies. Chapter 3 contains data and methodologies, chapter

4 explain data description and methodology while chapter 5 is about result and discussion.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Machine learning is a semi-automated elimination of Knowledge from data; starts with a question that might be responsible using data. Automated removal mean a computer provides the vision. Semi-automated mean it needs many smart conclusions by a human. Machine learning techniques such as Regression Trees, bagging, boosting, random forest, adaptive LASSO. Machine learning is mostly concerned with prediction, which is one of the fields of data mining. Econometrician, statisticians and data mining specialist are looking for insights that can be extracted from data. Econometricians usually use linear regression analysis for noticing and summarizing associations in data where Machine learning offer tools for summarizing nonlinear relationships in data. The purpose of Machine learning is to find some function which give good prediction.

There are number of studies which have discussed multiple predictor forecasting for macroeconomic variables. It also discusses review of studies for both conventional and modern approaches of multiple predictor forecasting for macroeconomic variables. Among other techniques for prediction, Regression Trees are widely used. It allows a mixture of continuous and categorical variables as predictors. Bagging takes many bootstrapped samples and fits tree to each bootstrapped sample (James et al 2013). Multiple predictors forecasting can be specified by using dynamic factor predictors rather to use all the variables (Stock & Watson, 2006). Now the approach which considered the uncertainty and risk factor is bootstrap aggregation or ‘bagging’. Adaptive LASSO, as a regularization method, avoids overfitting penalizing large coefficients.

2.2 Literature on Forecasting Methods From International Sources

In this section,we discuss both national and international literature related to techniques as a evidence that techniques are applied on different theories.

According to Breiman (1996) bagging predictors is a way to obtain a combined predictor from multiple generated predictors. For numerical results, collecting averages completes the version but to predict class it does a plurality vote. Bootstrapping is used to replicate new learning sets. Testing on simulated along with real data by employing regression and classification tree and selecting subset in linear regression showed that bagging yields higher accuracy.

Stock & Watson (2002) forecasted a single time series with large number of predictors using principal components[1]. An approximation model was followed by data, then index was used to encapsulate the predictors and further estimation was done by using principal components. Three major contributions were made in this study. Firstly, this study proposed general conditions on the errors discussion, this study showed that the principal components are consistent estimators of the true latent factors. Secondly, it was shown that the feasible forecasts which were constructed by using the estimated factors along with the estimated coefficients were converging to the infeasible forecast that would be attained if the factors and coefficients were recognized. Again, this study showed that results hold as N, T $\rightarrow\infty$. The feasible

[1] Principle Component Analysis (PCA) is a tool to reduce the dimension of a large set of variables into a small set of variables but still it consists of most of the information of large set of variables.

forecast was first-order asymptotically efficient. Finally, this article showed the robustness of the results to the time changes in the factor model due to temporal instability in macroeconomics forecasting models. The Monte Carlo results suggested that these methods can play a vital role in considerable development in forecasts instead of conventional models using a small number of variables.

Estruch et.al (2004) in their study combined several methods to improve prediction and proposed an algorithm to decrease the computational cost in bagging. Their method is rooted on employing multi-tree structures which allows common parts in trees. Thus, high similarity among trees signifies improvement of computational resource. Applying bagging on several similar decision trees can lead to correlation in misclassification error. However, it can be reduced by organizing trees into a multi tree. In this way accuracy will be improved too.

De mol *et. al* (2008) conducted study to create a linking between Bayesian regression and the conventional literature on forecasting with huge pieces created on principal components. They used Bayesian regression with normal and double exponential priors for forecasting large penal of time series comprising real and nominal variables. Real variables included sectorial industrial production, and hours' worked, while nominal variables included consumer and producer price indices, wages, money aggregates. The results revealed that limited selected variables were bright to capture the space spanned by the common factors. The study suggested that small models with accurately selected variables may do as well as methods that use information on large panels and are based on regressions on linear combinations of all variables. Their analysis provided a guide in the setting of the prior, also construed as a Ridge fining parameter. The results showed that the variable selection provided by the

LASSO regression is not clearly interpretable and they are not the typical ones that a macroeconomist could take**.**

Zhang and LU (2008) analyzed the variable selection method for Cox’s proportional hazards model and offered an integrated selection model and estimation strategy with preferred theoretical characteristics and calculative handiness. The approach combined the various penalties for various coefficients: irrelevant series receive greater penalties as compared to significant one, therefore, significant variables likely to be consider in the selection process, while irrelevant variables tend to be free. Furthermore, theoretical characteristics of Adaptive LASSO such as consistency, degree of merging of estimators, were considered. The study also presented that with suitable selection of regularization parameters, the given estimator had Oracle characteristics. Both computer-generated and real examples indicated that the method accomplished competitively. Selection of weights τj is very important for the Adaptive LASSO method. This study used $\tau_j = l/\beta_j$ 1, though, the β_j's in the presence high dimension data parameters might not be estimated. In the high dimension data covariates is larger than sample size n. β parameters may be unstable where strong collinearity exist. Remove collinearity we suggest that ridge regression estimator is robust estimator.

Potscher and Schneider (2009) studied the dispersal of the Adaptive LASSO estimator (Zou, 2006) in the large-sample limit as well as in limited samples. The big sample distributions were consequent both for the case where the Adaptive LASSO estimator is adjusted to perform conservative model selection as well as for the case where the tuning results in consistent model selection. Further, the study showed that the finite-sample as well as the large-sample distributions were typically highly non-normal, regardless of the choice of the tuning parameter. The finite-sample distribution was found to be a mixture of a singular normal distribution and an absolutely continuous distribution, which is non normal. They further added that the large-sample limit of the distributions was dependent on the choice of the estimator's tuning parameter.

Carriero and Kapetanios (2009) attempted at forecasting the XRs by using Bayesian VAR. They took data on monthly average of XR of US dollar via three data streams WMR/Reuters, Global Trade Information Services and FED, New York FED. They used Bayesian VAR with a normal inverted Wishart prior, imposing a prior univariate drift less random walk appear to be related more to intermittent use of information in the large panel than to change in the persistence of the XR. The study performed forecasting for 7 years (84 months) using rolling estimation. The gains of the forecast were in the range of 2-3% but in some cases of Euro-Dollar and GBP-Dollar these went up to 6-9%. The simple trading strategy suggested by Bayesian VAR was positive returns, higher than Random Walk (RW) forecasts. The best performance of BVAR was just because of irregular use of data in the large panel as compared to changes in persistence of XRs.

Herman *et.al* (2009) forecasted German GDP growth with a large factor model. In addition to German GDP, the study included data from Euro-area and G7 countries. The factors used in study were principle component as well as variables pre-selection prior. The results supported the use of targeted predictors from a large set of national and international data. In line with earlier findings from Boivin and Ng (2006), it was noted that more data does not always improve factor forecasting rather careful pre-selection of variables helps exploiting the additional information.

Using data of New Zealand Eickmeier and Ng (2009) used the approaches that are "data rich" factor and shrinkage method. The study aimed at investigating if data-rich methods were useful for forecasting using large quantity (hundreds of series) and variety (many countries and types) of available international data. They found that international evidence drawn from large international datasets significantly improved the forecasts of New Zealand GDP growth and outperformed forecasts based data and traded weighted aggregate of international variables. Shrinkage methods (RR and EN) and PLS performed mostly fit on international dataset, whereas Principal Component (PC)-based methods performed poorly. The reason could be the slightly weak factor structure or the low commonality of the international dataset. The paper also suggested that the result will be relevant for other forecasting techniques using international datasets.

Stock and Watson (2012) used Generalized Shrinkage Methods for forecasting using many predictors. Their study investigated the functioning of several methods of forecasting considered for a great number of orthogonal predictors (such as principal components). Their approach included pretest method, Bayesian model averaging, empirical Bayes, and bagging. Further, they compared empirical forecasts from these approaches to dynamic factor model (DFM) forecasts using a quarterly data of U.S. macroeconomic with 143 variables ranging from 1960-2008. For most of variables, together with measures of real economic activity, the shrinkage forecasts were less meaningful than DFM forecasts.

Koop and Korobilis (2012) aimed to forecast quarterly US inflation using generalize Philips curve by employing the Dynamic Moving Average (DMA) and Dynamic Models Selection (DMS). The number of the predictors are large if the model changes over time. DMA and DMS allow forecasting model to change, at the same time, they also allow coefficient in each model to evolve over time. Among other predictors, their study used GDP deflator and personal consumption expenditure (PEC) deflator for the purpose of forecasting.

Kim and Swanson (2014) empirically assessed the predictive accuracy of the large group of models of several macroeconomic variables. The study was based on the use of principle components and the different methods including Bayesian model averaging, LASSO and boosting. Moreover they found that the grouping of factor and other methods often yielded superior predictions. Their study focused on pure principal components models constructed using subset of variables selected via elastic net and shrinkage techniques. They concluded that some of the selected methods, when jointly used with factor examination, performed better than the other models. The monthly data is used from the period of 1960 month one to 2009 month five. Variables are used treasury bond yield, consumer price index, producer index, non- , non-farm payroll employment, housing starts, industrial production, M2, the S&P 500 index, and gross domestic product.

Raviv and Dijk (2014) analyzed forecasting with many predictors by allowing for non-linearity. They used data set of 126 U.S. macroeconomic and financial variables of the monthly frequency for the period April 1959 - September 2009. They applied the empirical procedure in two steps. In the first step, they screened for exciting effects by simple t-tests, accounted for controlling False Discovery Rate (FDR). In the second step, they used Ridge-Regression to lessen possible over fitting. Results showed that accuracy was achieved by allowing for both squares and first level interactions of the original explanatory variables. In sum, the study proposed a refinement to the line of research in the form of controlling the FDR under general correlation structure. Results of the empirical application of study suggested that allowing for a non-linear relation could lead to substantial accuracy gains.

Li and Chen (2014) forecasted macroeconomic time series by using LASSO-based approaches and their forecast combinations with dynamic factor models. They examined a group of LASSO-based approaches and evaluated the abilities of prediction for forecasting twenty important macroeconomic series. . The data set contained 107 macroeconomic indicators and used monthly data set. The study asserted both analytically and empirically that combing forecasts from LASSO-based models with those from dynamic factor models could decrease the Mean Square Forecast Error (MSFE) further. For most of the series under analysis, all the LASSO-based models outperformed dynamic factor models in the out-of-sample forecast evaluations. Lastly, they concluded that the combined forecasts were significantly better than dynamic factor model forecasts and as well as the naïve random walk benchmark.

Kascha and Trenkler (2015) compared various selection and penalized regression methods by using the approach for forecasting with autoregressive model for USA data. The variables used in this study are CPI and GDP growth, quarterly data taken from federal reserve bank from time period 1959Q1 to 2012 Q2. The study investigated the effect of system size and prior specification on over all forecasting performance of the method it also found that all lags order is more important than the selection method which means that for this data set the subset selection method not achieved well performance.. The results indicated that increasing the system size can be helpful, depending on the employed shrinkage method.

Heij *e*t.al (2016) used kernel ridge regression method to forecast macroeconomic variables of United States. The variables production, consumption, income, sale, employment, monetary aggregate, price, interest rate, and XR were taken from U.S and from January 1959 to December 2003. kernel ridge regression is confirmed by Monte Carlo simulations as well as empirical application to various key measures of real economic activity, produced more accurate forecasts than outdated linear methods for distributing with many predictors based on principal component regression. The results of four key U.S. macroeconomic variables showed that Kernel-based methods were better, and competitive with, well established autoregressive and principal-components-based methods. Kernel ridge regression showed a relatively consistent good analytical performance, also during the crises period in 2008-9. However, it was outperformed by linear principal components only in those periods when the latter method performed outstandingly well. The study concluded that valuable addition to the macroeconomic forecasting is kernel methodology.

Jiang *et.al* (2017) used Australian macroeconomic data to find out additional information that can improve the forecasting accuracy of small models. For this purpose they used several estimation techniques including Bayesian vector Autoregression (VAR), Univariate ARIMA (BOX and Jenkins), Dynamic Factor Model, Ridge Regression and Least Angle Regression (LARs), Naïve forecast and vector Autoregression (VAR). Quarterly data of selected 151 variables was taken from 1984-2015. LARS and bagging LARS performed best with three predictors in the model. As the information set increased from 3 to 13 Ridge and Bayesian vector Auto regression (BVAR) provided similar forecasting as LARs result. The Dynamic Factor Model (DFM) however was least accurate for all cases.

Fattahi *et.al (*2017) tried to model the relationship between the recession and major indicators. Their study made use of statistical data and decision making trees and attempted to forecast the next recession in Iran. Data on selected variables was taken from central bank, statistical center of Iran. The method used was boosted RTs (BRT) with the variables oil export, unexpected momentum inflation, real total import and inflation. Results showed that momentum imports, revenue from oil exports, unexpected momentum INF, real total import, and INF were more effective in Iran's recession forecast. The results also indicated that BRT method can be a useful technique for analysis of economic policies and could be further used in forecasting other macro-economic variables. The study concluded that even though machine learning techniques provided important intuitions on working of an economy, but from the perspective of business cycle forecast applications, they couldn't replace empirical forecasts of business cycle.

Similarly, Dopke *et.al* (2017) also attempted at predicting recession on Germen data using business cycle as an indicator by using Boosted RTs (BRT). The monthly data is taken from 1973 M1 to 2014 M12 and the variables are financial indicators, surveys, real economy variables, prices, and composite leading indicators. Study concluded that in German, the short term interest rate and the term spread were important leading indicators of recession. Further, the predictive power of the short-term interest rates declined over time. On the other hand, term spread and the stock market faced an increase in their predictive power. The paper concluded that machine learning techniques can yield significant visions into how an economy works. Moreover, the BRT approach made it possible to study complex marginal effects of leading indicators on the recession probability. Lastly, this approach also supported to recover the composite ways in which leading indicators interrelated in predicting recessions.

2.3 Literature on forecasting for Pakistan using large number of Predictors

This section includes the literature about the macroeconomic forecasting for Pakistan using large number of predictors.

Tahir (2014) forecasted the three prior i-e, CPI Index based measure of Inflation and Credit macro variables in order to forecast Output Gap and Monetary Aggregates. In their study output gap was estimated using the Kalman filter in a state space frame work. Bayesian shrinkage has proper techniques for forecasting to use the large number of macroeconomic variables. The estimation of the robust forecasting showed that Bayesian VAR had good result for the interpretation. The study recommended Bayesian shrinkage as a proper tool for forecasting using large number of variables.

Forecasting electricity consumption and generation for various sectors of Pakistan economy, Hussain et al. (2016) also used a large number of predictors i-e, reserves money (M0), broad money (M2), public sector borrowing (PSB), private sector credit (PSC) output gap and CPI inflation. For this purpose two techniques ARIMA and Holt Winter were applied on secondary data from 1980-2011, taken from different issues of economic survey of Pakistan. Comparing different sectors, forecasting models showed that increase in energy generation was less than the increase in total electricity consumption. Thus, there is room for policy options to cope up with electricity shortage challenge. Estimation results showed that Holt-Winter forecasting model had minimum RMSE (root mean square error) and MAPE (mean absolute percentage error) when compared to ARIMA model.

According to (Potscher and Schneider 2009) the huge sample distributions are derivative both for the condition where the Adaptive LASSO estimator is used to to make conventional model range as well as for the situation where the tuning results in reliable model selection. The author show that the finite-sample as well as the large-sample distributions are typically highly non-normal, regardless of the choice of the tuning parameter. The theoretical results, which are obtained for a regression model with orthogonal design, are complemented by a Monte Carlo study using non-orthogonal regressors. The theoretical study assumes an orthogonal regression model. The finite-sample distribution was found to be a mixture of a singular normal distribution and an absolutely continuous distribution, which is non normal. The large-sample limit of the distribution depends on the choice of the estimator's tuning parameter.

2.4 Conclusion

The literature concludes that in case of Pakistan, no previous study has compared the benchmark methods of forecasting with multiple predictors forecasting by considering 31 Macroeconomic variables. The most recent research Tahir (2014) only used Bayesian shrinkage to forecast macroeconomic variables. Therefore the current study will consider the comparison of benchmark methods and multiple predictors forecasting methods including RT, Bagging, Boosting, Random Forest and Adaptive LASSO.

CHAPTER 3

DATA AND METHODOLOGY

3.1 Introduction

To forecast the macroeconomic variable different techniques are used such as: Regression Tree, Bagging, Boosting, Random Forest, Adaptive LASSO and benchmark method which is AR.

In this chapter the multiple predictors forecasting methods are explained in detail.

3.2 Regression Tree (RT)

Among other techniques for prediction, RTs are widely used. It allows a mixture of continuous and categorical variables as predictors. A RT is constructed through a process known as binary recursive subdividing, which is an iterative process that splits the data into partitions or branches, and then continues splitting each partition into smaller groups as the method moves up each branch. The algorithm selects the cut off points for separating and adds the split to the tree. By splitting, data is divided into several subsets with each instance belonging to one subset. A decision tree is created when each decision node in the tree contains a test on some input variable's value. The terminal nodes of the tree contain the predicted output variable values. The decision rule for this prediction is based on the most frequent outcome.

The prediction ability of the rule is assessed using mean square error. Further, it is cross validated by re-splitting of the sample and then calculating the error each time to obtain the average error for all the trees. The tree model's parameters are "if-then" linkages which divide the data as per the observed inputs' values. This is in contrast with linear models, where the parameters are linear coefficients of each regressor. The classification and RTs (CART) algorithm is probably the most popular algorithm for tree induction.

As a first step, data in the training set is segregated into equal portions. Then using binary splitting method on each field, the algorithm divides the data into two groups/branches. The splitting criteria is such that the sum of squared deviations from mean in both groups is minimum. Same rule is applied for splitting at every step. Splitting terminates only when the node size approaches the minimum node size specified by the user. The final nodes are then termed as the terminal nodes. However, even without reaching the minimum size, a node can become a terminal node if its sum of the squared deviations from the mean is zero.

3.2.1 Tree Pruning

Tree pruning is used to check over fitting the model of the tree. If the produced tree for training data doesn't fit well for test data then there is a problem of over fitting. There are few reason from which over fitting arises such as lack of representative sample, presence of noise etc. A good model must have small generalization error (expected error based on unseen records or observations) as well as low training error (based on training data or re-substitution error).

Reason of over fitting is that the method works so hard on training data or memorizes the training data and finds patterns of unknown function caused by random chance instead of true pattern. Overgrowing the tree faces the problem of over fitting, the smaller the tree the lower its variance and gives better interpretation but at the cost of little bias. The solution to this problem is to grow a tree as long as residual sum of square (RSS) decrease due to each split exceeding some threshold. This method produces smaller tree but it is short sighted because a seemingly useless split might follow a best split, which gives low RSS. Instead of that we should grow a large tree and then prune it back .the pruned tree gives lowest test error rate.

3.2.2 Cross validation

Cross validation is the estimation of testing error method that estimates the model on the basis of testing error. Like RT, cross validation also divides the data set into two parts i.e. testing and training data. Training data is used to check or fit a model.

3.3 Bootstrap Aggregation /Bagging

To increase the prediction accuracy and decrease the variance, bootstrap aggregation is used. CART faces the problem of high variance due to random split of data set in to training and testing data. If repeatedly we fit the model on different training set we get different result. If we have n independent observationX_1,X_2,X_3,…….X_n have variance σ^2 the variance there mean is σ^2/n of the observations. The mean of the observation reduce the variance. (James et al; 2013)

3.4 Boosting

Like bagging boosting also combine a lot of trees but bagging takes many bootstrapped samples and fit tree to each bootstrapped sample while boosting make trees sequentially. (James et al; 2013)

3.5 Random Forest

Random forest is a best approach than Bagging. It decorrelate the trees while in bagging trees are highly correlated and when the average of highly correlated trees are taken there is not that much reduction in variance than taking the average of uncorrelated trees. In bagging most of the trees look alike because most of the trees use the strongest predictor as a first split because we have random sample of m predictors at each split out of full p predictors. While random forest force each split to consider only a subset of predictors. On average (p-m)/p of the splits are not even allowed to use the strongest predictor like this other predictors will have chance for being split. In random forest if m=p then it is same like bagging and there will be no de correlation of trees. Random forest uses m= p which reduce the variance of the resultant tree.

The parameter coefficient (β) correlates with the tuning parameter value. Whenever λ=0 then the penalty term has no effect and we will get the same coefficients as simple linear regression. When $\lambda =\infty$ then all the coefficients are zero. When λ is in between the two extremes ($0< \lambda<\infty$) then we are balancing the two ideas. Such as; fitting a linear model of Y on X and shrinking the coefficients. The range of tuning parameter is between zero to infinity and it is a crucial value for the identification of the true model.

3.6 Adaptive LASSO

Zou (2006) showed the LASSO estimator does not enjoy the oracle property, and proposed a simple and effective solution the properties are:

• Identifies the right subset model, {j :βˆ

j = 0} = A

• Has the optimal estimation rate, √n(βˆ(δ)A − β∗

A) →d

N(0,∗), where ∗ is the covariance matrix knowing the true subset model.

In this technique, Adaptive weights are being utilized for penalizing various coefficients in the panel. This represents that Adaptive LASSO utilizes the oracle characteristics, mostly, it executes well if the true original models are given in advance. The Adaptive LASSO is similar to the LASSO in the regard that it is revealed to nearby minimum optimal point. Additionally, Adaptive LASSO can be resolved by using the effective algorithm. The Adaptive LASSO, which employs different weights to different coefficients. Adaptive LASSO is modified form of the least absolute shrinkage and selection operator (LASSO). Suppose that we have data (x_i , y_i), i = (1, 2, . . ., N), where x_i = $(x_{i1}, \dots x_{ip})^T$ are the predictor variables and y_i are the responses. Letting $\hat{\beta}$ = $(\hat{\beta}_1, \dots \hat{\beta}_p)^T$ the LASSO estimate ($\hat{\beta}_{LASSO}$) is defined as

$$\hat{\beta}_{LASSO}(\lambda) = argmin \sum_{i=1}^{n} (y_i - \beta_o - \sum_{j=1}^{p} \beta_j \, x_{ij})^2 + \lambda \sum_{i=1}^{p} |\beta_j| \qquad (1)$$

Where we have n observations, p covariates, λ is shrinkage parameter (a nonnegative regularization parameter) and $|\cdot|$ is the absolute value so $\sum_{i=1}^{p} |\beta_j|$ is the $l_1$norm (l_1 penalty). The $l_1$norm does not only reduces the absolute value of the coefficients compared to the value of the estimation by OLS, but also performs the selection of variables, i.e. which are relevant to explain or predict values of y and those that are not.(6). Zou (2006) proposes the Adaptive LASSO, which employs different weights to different coefficients (Konzen & Ziegelmann, 2016).

$$\hat{\beta}_{adpLASSO}(\lambda) = argmin \sum_{i=1}^{n} (y_i - \beta_o - \sum_{j=1}^{p} \beta_j \, x_{ij})^2 + \lambda \sum_{i=1}^{p} \widehat{\omega}_j |\beta_j| \qquad (2)$$

where $\widehat{\omega}_j$ (j = 1,..., p) are the adaptive data-driven weights, which can be estimated by $\widehat{\omega}_j$= $\left|\hat{\beta}_j^{ridge}\right|^{-t}$, t is a positive constant and $\hat{\beta}_j^{ridge}$ is an initial consistent estimator of β obtained through least squares or ridge regression if multicollinearity is important [10]. The weights ωj can also be obtained from OLS estimates; however, this would be limited to the case where $n > p + 1$(Konzen & Ziegelmann, 2016).

3.7 Data

To identify the challenges faced in macroeconomic forecasting, this study aims to conduct an extensive review of procedure for forecasting in time series. . Data of thirty one series (Inflation, Money supply, Real Effective Exchange Rate, worker remittances, electricity, natural gas, production of cement, production of gypsum, production of lime stone, production of rock salt, production of crude oil, production of good coal, currency in circulation, cash in Pakistan, import of petroleum, international Liquidity gold holding, export of rice, total export, total import, intentional liquidity, balance with state bank of Pakistan, WPI food, WPI manufacture, finical interest rate, nominal effective XR, XR of China, XR of United Arab Amarat, XR of Saudi Arabia, XR of USA, XR of Malaysia.) is taken from SBP and IFS sources. The monthly data of Pakistan is taken from the period of (1990-2017).

The data of all 31 economic variables is taken from SBP and IFS for the economy of Pakistan. From this we have to examine the multiple predictor forecasting.

3.7.1 Pakistan's Macroeconomic variables

Macroeconomics is the study of the economy as a whole, and the macro variables are usually thought to be responsible to control the macro-economy. Macroeconomics does not provide the explanation of economic measures only but it also expand the economic policy as well and provides tools for the assessment of economic policy. It regulates the government policy meant to control and stabilize the economy over time, that is, to reduce fluctuations in the economy. As government policies tries to stabilize the economy, the policy makers tries to formulate suitable policies in consultation with economists. Moreover, suitable Macro policies make it favorable to control factors like inflation and deflation, unemployment, production, consumption and moderate violent booms and recessions in any economy.

Forecasting of macroeconomic variables provides us a tool to sense where the economy may be directed, allowing consumers, producers, government, and macroeconomic policy planners to plan in advance with some level of rational certainty. Forecasting usually reduce the level of uncertainty about the future development of the economy. Forecasting can be easier by carefully analyzing various macroeconomic variables, and persistent nature of the business cycles. Policy makers and macroeconomists may form forecasts by considering statistical regularities as procedure for predicting the economy and for the overall efficiency of the economy. In this study we use macroeconomic variables for forecasting by using RT, Bagging, Boosting, Random Forest, Adaptive LASSO and Auto Regressive Model. Improved and timely forecasting of macro economy is a recent need of Pakistan economy as it can provide a dimension to Pakistan's government for further and will provide input for government budget planning, central bank policy planning and business initiatives.

3.7.1.1 Consumer price index (CPI)

CPI is the main measure of price changes at retail level. It measure the changes in the cost of buying representative predefined basket of goods and services and to gauge the increase in the cost of living in reporting period. Laspeyer's formula used to compute CPI is:

$$CPI=[\varepsilon(p_{n/}p_{o)}w_i/w_i]*100$$

Where P_n = Price of an item in the current period.

P_o= Price of an item in base period

W_i= Weight of the ith item in the base period.

3.7.1.2 Broad money (M2)

Broad Money (M2) is the indicator for gauging the liquidity in an economy. It incorporates the currency in circulation, deposits held by the SBP and other demand and time deposits held by the scheduled banks.

3.7.1.3 Wholesale Price Index (WPI)

Wholesale price index is calculated to identify the price level movements in wholesale sector. The items in the selected basket are those sold in bulk by producers and manufacturers. The prices used are those being paid by the primary buyers at market or ex-factory level.

3.7.1.4 Currency in Circulation (CIC)

The Currency in circulation is the money held by the general public. It is the amount of money existing in the economy outside its banking sector.

3.7.1.5 Workers' Remittances (WR)

Workers' remittances is one major form of inflow of money into the economy from abroad. It is the amount of money being sent by the migrants who are employed abroad, to their families residing within the country.

3.7.1.6 Nominal effective exchange rate (NEER)

It is an index of a country's exchange rate relative to the major trade partners. It calculated by incorporating the relative weight of trading partner in the host economy's total foreign trade.

3.7.1.7 Electricity Generation (ELE)

For infrastructure of an economy is the amount of electricity generation via all possible sources including coal, gas and oil. Coal includes several kinds (brown, black etc.), natural gas other than liquid gas, petrol derived products and crude oil. Since

Pakistan is a net importer of most of these sources of electricity generation, the import bill significantly affects the level of inflation prevailing in the economy.

3.7.1.8 Exchange Rate (XR)

One of the most significant and critical macroeconomic indicator is the XR of any economy. It gauges the strength of economy relative to other countries. It's calculated by dividing the local currency with the foreign currency (usually US dollar). XR significantly affects the inflation in economy.

3.7.1.9 Total reserves includes gold, (current US$)

The total foreign reserves of an economy being held by the central bank are an indicator of its strength and health. A growing economy would have huge reserves and vice versa. Reserves consist of the gold, SDR, forex and reserves with the IMF. Amount of reserves is usually measured in dollars and is greatly influenced by the inflation in an economy.

CHAPTER 4

RESULTS AND DISCUSSION

In this chapter, We estimate five techniques of forecasting i-e, regression tree (RT), Bagging ,Random Forest, Boosting, Adaptive LASSO and auto regressive model (AR). AR method is used as benchmark method. We have evaluated all these techniques with the help of mean square error calculated using Pakistani data. For this purpose we divide the data set into two groups training and testing. 1990 M1-2014 M12 data is considered as training and 2015 M1-2016 M12 data is considered as test data. In the first step we take twelve lags with time of the data to check the stationarity.

4.1 Regression Tree for information set i = 4

In estimation of regression tree the data set consist on four variables which are inflation, real effective exchange rate, money supply and currency in circulation.

4.1.1 Using inflation as dependent variable

We use inflation as dependent variable to check the determinants or corelate of inflation.

a) *On Train Data Set*
   The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 4.

Table 4.1 Regression Tree Results of Inflation for Train Data

<table>
<tr><td colspan="2">Number of terminal nodes</td><td colspan="4">26</td></tr>
<tr><td colspan="2">Residual mean deviance</td><td colspan="4">0.00002 = 0.005369 / 268</td></tr>
<tr><td colspan="6">Distribution of residuals</td></tr>
<tr><td>Min
-0.01213</td><td>1st Qu
-0.00281</td><td>Median
-0.0003</td><td>Mean
0.000</td><td>3rd Qu
0.00281</td><td>Max
0.0139</td></tr>
<tr><td colspan="6">Variables: lag9M2, lag8M2, lag7M2, lag5M2, lag1M2, lag4M2, lag3REER, lag9CIC, lag4CIC, lag11CIC, lag1REER, lag3CPI, lag2REER, CIC, lag5CPI, lag1CPI, lag9REER, lag4CPI.</td></tr>
</table>

The Above Table 4.1 shows that inflation is used as dependent variables whereas real Effective exchange rate (REER), money supply (M2) , currency in circulation (CIC), their lags and lags of dependent variables are used as regressors. From among all variables lag 1, 4, 5, 7, 8, 9, of M2, lag 1, 2, 3, 9 of REER, lag 1, 4, 9, 11 of CIC and lag 1, 3, 4, 5 of CPI are important which means that the tree select all these variables for constructing the tree. Residual mean deviance is used as error which is 0.00002.

4.1.1.1 Cross validation on train

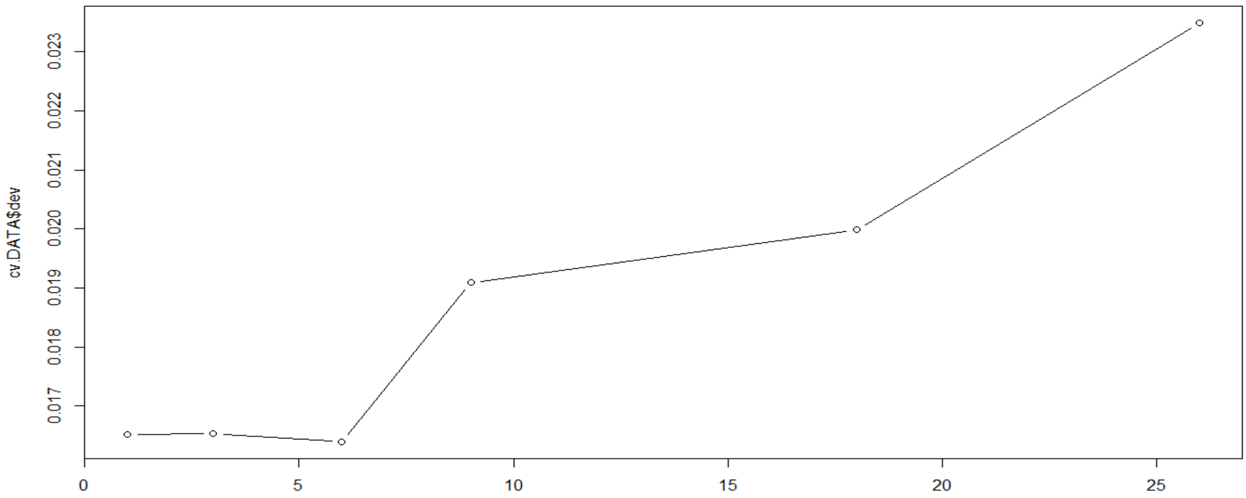

Figure 4. 1: Cross Validation on Train Data

Cross validation Graph 4.1 is used to check the size of tree against the deviance. When the size of tree is large and not cleared then cross validation is used. Here the minimum deviance is six so we use the prune tree to minimize the size of tree.

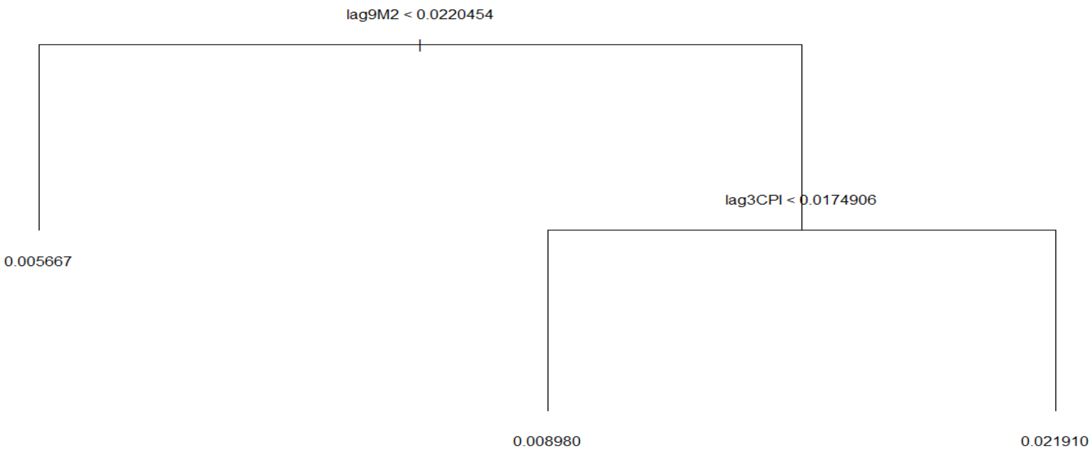

Figure 4.2 Regression Tree of Inflation in Train Data

The Above Plot 4.2 shows that when lag9 of money supply is less than 0.0220 the predicted inflation value is 0.0051 and the lag9 of money supply value is greater than 0.0220 and the value of lag3 of inflation is less than 0.0175 the predicted inflation value is 0.0089 and the value of lag3 of inflation is greater than 0.0175 the predicted inflation is 0.022.

b) *Inflation on Test Data Set*

The data used for test data set is from 2015 M1- 2016M2 and information set is i=4.

Table 4.2 Regression Tree Results of Inflation for Test Data

Number of terminal nodes	3				
Residual mean deviance	0.000033 = 0.0004 / 15				
Distribution of residuals					
Min	1st Qu	Median	Mean	3rd Qu	Max
-0.0056	-0.0024	0.0001	0.000	0.0014	0.0079
Variables: M2, lag5REER					

The above Table 4.2 is use for inflation to use the test data in which inflation is used as dependent variable whereas money supply (M2), real effective exchange rate (REER) currency in circulation (CIC), their lags and the lags of dependent variables is used as regressors. From among all variables M2 and lag5 of REER is important variables and is used for construction of the tree. It means that change in REER affects inflation after 5lags. Residual mean deviance is 0.000015 which is used as mean square error.

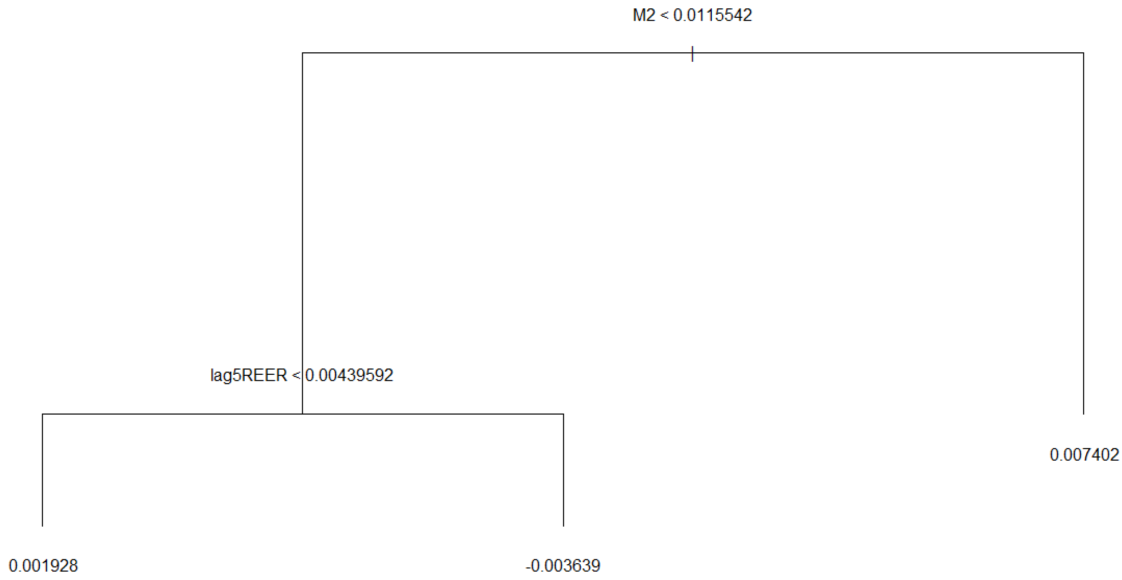

Figure 4 3 : Regression Tree of Inflation in Test Data

The Above Figure 4.3 show that when M2 is less than 0.0012 and the lag 5 of REER is less than 0.0044 the predicted inflation is 0.0019 and the lag 5 of REER is greater than 0.0044 the predicted inflation value is 0.0037. On the other hand when M2 is greater than 0.0012 the predicted inflation value is 0.0074.

4.1.2 Using M2 as Dependent Variable

We use M2 as dependent variable to check the determinants or corelate of inflation.

a) *On Training Data*

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 4.

Table 4.3 Regression Tree Results of Money Supply for Train Data

Number of terminal nodes			5		
Residual mean deviance			0.1773 = 51.23 / 289		
Distribution of residuals					
Min	1st Qu	Median	Mean	3rd Qu	Max
-5.5290	-0.0077	0.00183	0.000	0.0171	1.411
Variables: lag5CPI, lag1M2, CIC, lag11CIC					

The Above Table 4.3 shows that money supply is used as dependent variable whereas inflation, REER, currency in circulation, their lags and the lags of dependent variable is used as regressors. Lag 5 of inflation, lag1 of money supply lag 1 and 11 currency in circulation are used to construct the tree which means that they are important variables. Residual mean deviance is 0.1773 which is used as mean square error.

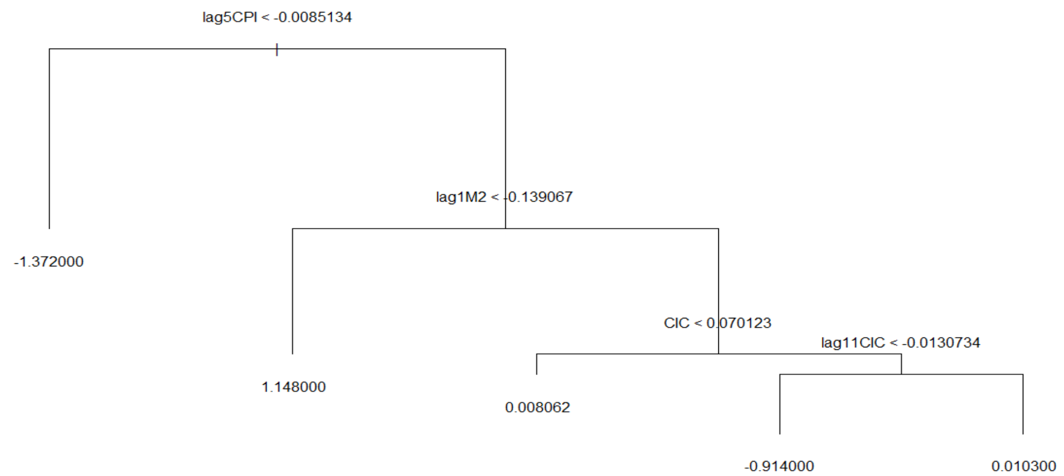

Figure 4.4 Regression Tree of Money Supply in Train Data

The Above Figure 4.4 shows that when lag5 of inflation is less than 0.0085 the predict money supply is -1.372. On the other hand when lag5 of inflation is greater than 0.0085 and lag1 of M2 is less than 0.139 the predicted money supply is 1.148 and lag1 of money supply is greater than 0.1391 and currency in circulation is less than 0.0701 the predicted supply money is 0.0081. When the value of currency in circulation value is greater than 0.0701 and the lag1 of circulation is less than 0.0131 the predicted supply money is -0.914 and the lag1 of currency in circulation is greater than -0.0131 the predicted value of money supply is 0.0103.

b) On Test Data
c) The data used for test data set is from 2015 M1- 2016M2 and information set is i=4.

Table 4.4 Regression Tree Results of Money Supply for Test Data

<table>
<tr><td colspan="3">Number of terminal nodes</td><td colspan="3">3</td></tr>
<tr><td colspan="3">Residual mean deviance</td><td colspan="3">0.00010 = 0.00154 / 15</td></tr>
<tr><td colspan="6">Distribution of residuals</td></tr>
<tr><td>Min
-0.0262</td><td>1st Qu
-0.0039</td><td>Median
-0.00015</td><td>Mean
0.000</td><td>3rd Qu
0.0028</td><td>Max
0.0233</td></tr>
<tr><td colspan="6">Variables: lag6M2, lag7REER</td></tr>
</table>

The Above Table 4.4 shows that money supply (M2) is used as dependent variable whereas inflation (CPI), real effective exchange rate (REER), currency in circulation (CIC), their lags and the lags of dependent variables. Here lag6 M2 and lag7 of REER is important variable and is used to construct the tree. There terminal nodes are used in this tree and the residuals mean deviance is 0.0001 which is used as mean square error.

4.1.3 Using Real Effective Exchange Rate as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 4.

Table 4.5 Regression Tree Results of Real Effective Exchange Rate for Train Data

Number of terminal nodes			25		
Residual mean deviance			0.00011 = 0.0299 / 269		
Distribution of residuals					
Min -0.0355	1^{st} Qu -0.0062	Median 0.0002	Mean 0.000	3^{rd} Qu 0.0064	Max 0.0335
Variables: lag4CIC, lag1REER, lag1CPI, lag2M2, lag11M2, lag9CIC, lag7M2, lag7REER, lag9M2, lag10CPI, CPI, lag2CPI, CIC, lag8REER, lag2CIC, lag3M2					

The Above Table 4.5 shows that real effective exchange rate (REER) is used as dependent variable whereas money supply (M2), inflation (CPI), currency in circulation (CIC), their lags and the lags of dependent variables is used as regressors. From among all variables lag 1,2,4,9 of CIC, lag 1,7,8 of REER, lag 1,2,3,10 of inflation, 2,3,7,9,11 of M2 are important and are used to construct the tree. Twenty five terminals nodes are used in this tree and the residual mean deviance is 0.00011 which is used as regressors.

4.1.3.1 Cross validation

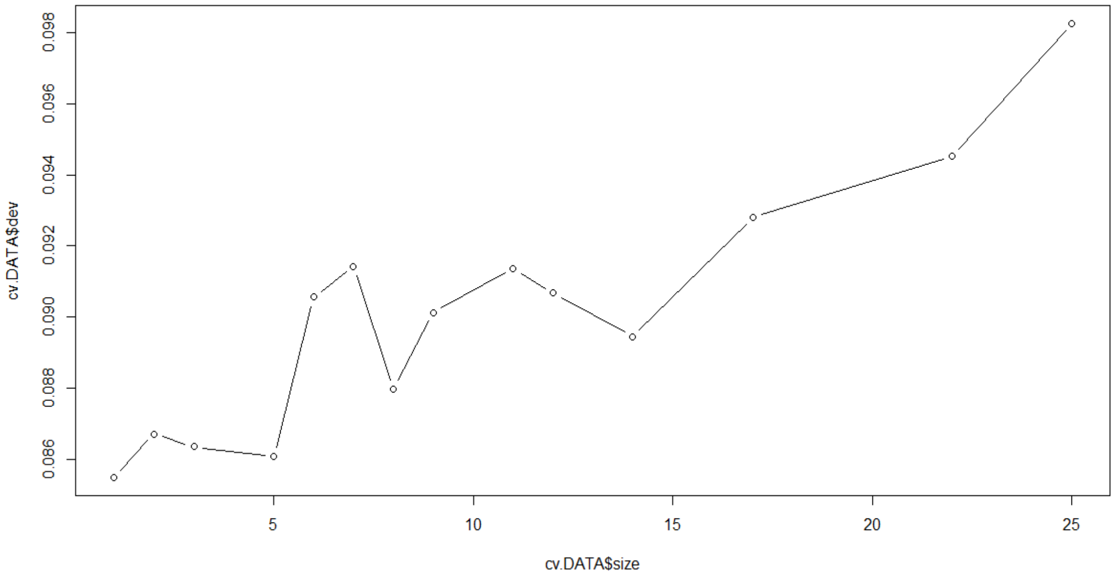

Figure 4.5: Cross Validation of Money Supply

Figure 4.5 of cross validation is used to check the size of tree against the deviance. Here the minimum deviance is 2. Here used the tree pruning.

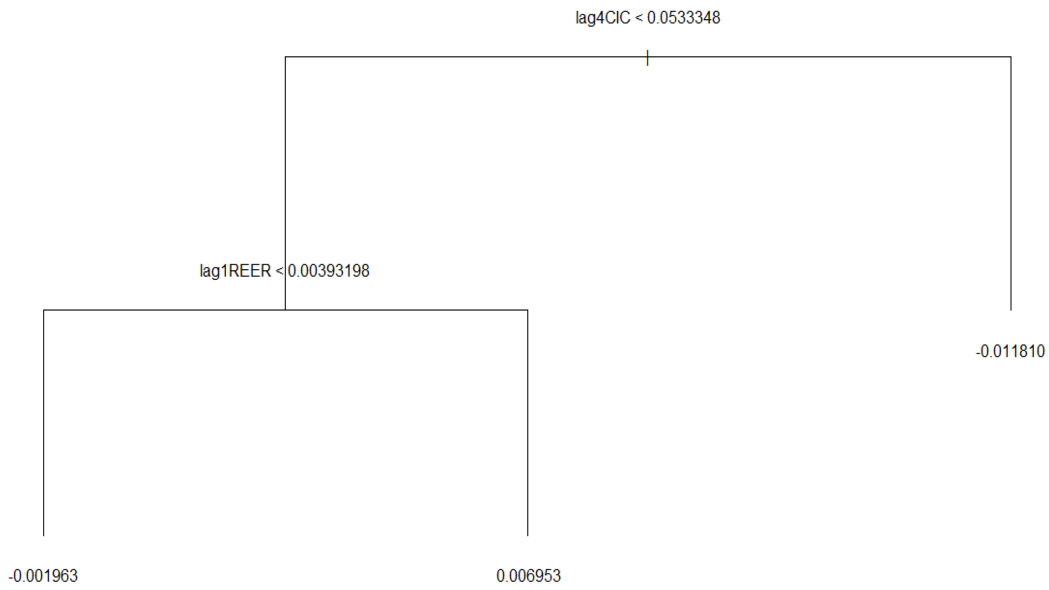

Figure 4.6: Regression Tree of Real Effective Exchange Rate in Train Data

The Above Figure 4.6 shows that when lag4 of currency in circulation is less than 0.0533 and the lag1 of REER is less than 0.0039 the predicted value of REER is -0.0019 and the value of lag1 of REER is greater than 0.0039 the predicted value of REER is 0.0069. On the other hand when lag4 of currency in circulation value is greater than 0.0533 the predicted value of REER is -0.0118.

b) *On Test Data*
The data used for test data set is from 2015 M1- 2016M2 and information set is i=4.

Table 4.6 Regression Tree Results of Real Effective Exchange Rate for Test Data

Number of terminal nodes			3		
Residual mean deviance			0.000037 = 0.0005487 / 15		
Distribution of residuals					
Min -0.0133	1st Qu -0.0037	Median -0.0002	Mean 0.000	3rd Qu 0.0035730	Max 0.0104300
Variables: lag11CPI, CPI					

The Above Table 4.6 shows that real effective exchange rate (REER) is used as depen dent whereas money supply (M2), inflation, currency in circulation (CIC), their lags a nd the lags of dependent variables are used as regressors. From among all variables la g11 of inflation and inflation is important variable and is used to construct the tree. Th ere are tree terminals nodes are used and the mean deviance is 0.00004 which is used as mean square error.

4.1.4 Using Currency in Circulation as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 4.

Table 4.7 Regression Tree Results of Currency in Circulations for Train Data

<table>
<tr><td colspan="3">Number of terminal nodes</td><td colspan="3">6</td></tr>
<tr><td colspan="3">Residual mean deviance</td><td colspan="3">0.08002 = 23.05 / 288</td></tr>
<tr><td colspan="6">Distribution of residuals</td></tr>
<tr><td>Min
-1.887</td><td>1st Qu
-0.024</td><td>Median
-0.0014</td><td>Mean
0.000</td><td>3rd Qu
0.0245</td><td>Max
1.986</td></tr>
<tr><td colspan="6">Variables: lag1CIC, lag3REER, lag8CPI, lag3CPI</td></tr>
</table>

The Above Table 4.7 shows that currency in circulation (CIC) is used as dependent variable whereas inflation, money supply (M2), REER (REER), their lags and the lags of dependent variable is used as regressors. From among all variables lag1 of CIC, lsg3 of REER lag 8 and 3 of inflation are important variables and is used for constructing the tree. Number of terminal nodes is 6 and residual mean deviance is 0.0800 which is used as mean square error.

4.1.4.1 Cross validation

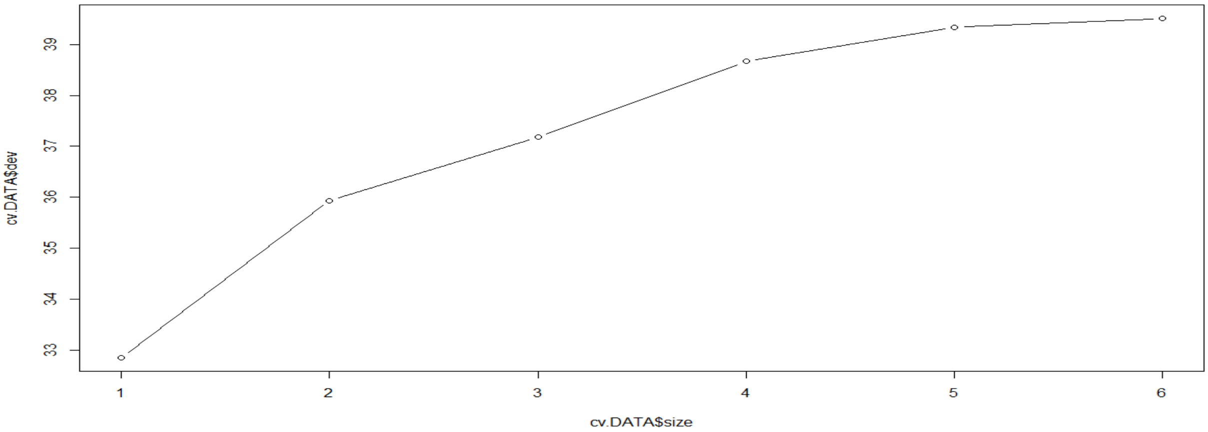

Figure 4.7: Cross Validation of Currency in Circulation

Cross validation is used to check the size of tree against the deviance. Here the minimum deviance is 1.

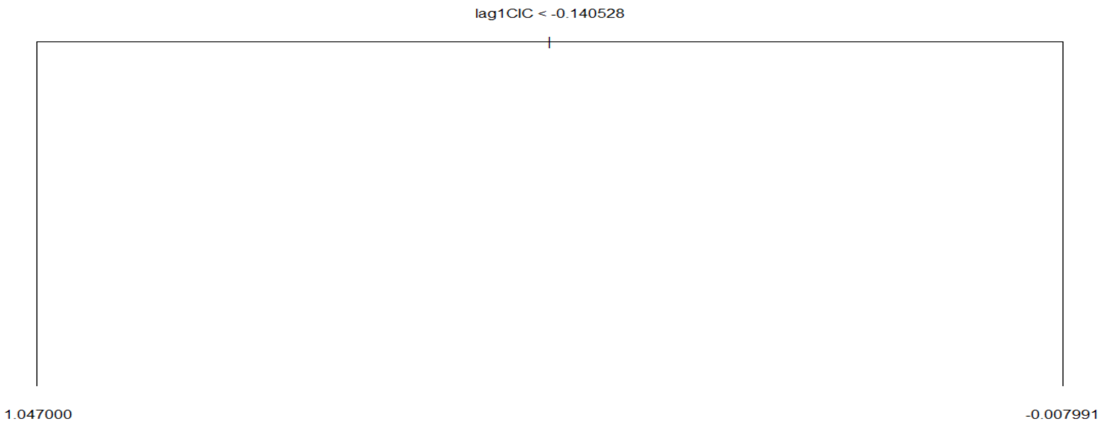

Figure 4 8: Regression Tree

The Above Figure 4.8 show that when lag1 of CIC is less than -0.1405 the predicted value of currency in circulation is 1.047 and lag1 of CIC is greater than -0.047 the predicted currency in circulation value is -0.0079.

b) *On Test Data*

The data used for test data set is from 2015 M1- 2016M2 and information set is i=4.

Table 4.8 Regression Tree Results of Currency in Circulations for Test Data

Number of terminal nodes		3			
Residual mean deviance		0.0072 = 0.008607 / 15			
Distribution of residuals					
Min	1st Qu	Median	Mean	3rd Qu	Max
-0.0622	-0.0073	0.0023	0.000	0.0093	0.0459
Variables: lag6REER, lag10CIC					

The Above Table 4.8 shows that currency in circulation is used as dependent variable whereas inflation, money supply (M2), real effective exchange rate (REER), their lags and the lags of dependent variable is used as regressors. From among all variables lag6 of (REER), and lag10 of CIC are important variables and is used as regressors. There are three terminal nodes in this tree the mean deviance is 0.0006 which is used as mean square error.

4.2 Boosting for Inflation Using Information set i = 4

In boosted Regression Tree four variables are used. Inflation is used as dependent variable whereas real effective exchange rate, money supply and currency in circulation are used as regressors.

4.2.1 Summary of Boosted Regression Tree (BRT)

Variables are used in the BRT inflation is used as dependent variable whereas money supply (M2), Real effective exchange rate (REER), currency in circulation (CIC), their lags and the lags of dependent variable is used as regressors. A gradient boosted model, with gaussian loss function 5000 iteration were performed and there are 47 regressors are used of which 47 had non zero influence.

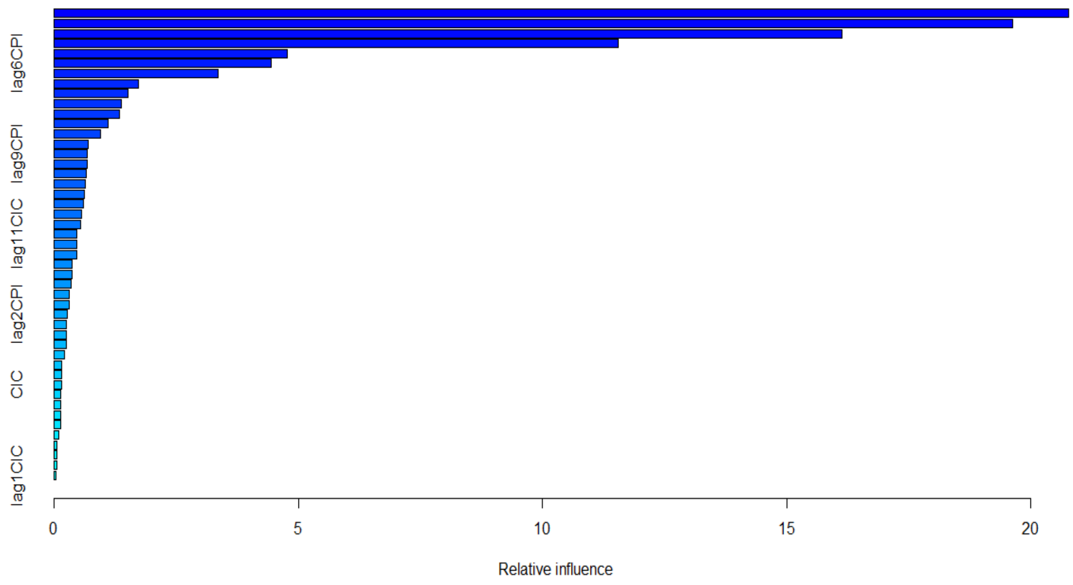

Figure 4.9: Boosting of inflation

The Above Figure 4.9 shows that the highest bar shows the relative importance of the variables. The first highest bar shows, lag7 of money supply, which is the most significant influential variable, then lag6 of currency in circulation, then lag6 of money supply and so on. The mean square error of boosting is 0.000014.

4.2.2 Summary of Boosted Regression Used M2 as Dependent Variable

Variables are used in the boosted RT money supply (M2) is used as dependent variable inflation, real effective exchange rate (REER), currency in circulation (CIC), their lags and the lags of dependent variable is used as regressors. A gradient boosted model, with gaussian loss function 5000 iteration were performed and 47 regressors are used out of which 47 had non zero influence. The mean square error is 0.000076.

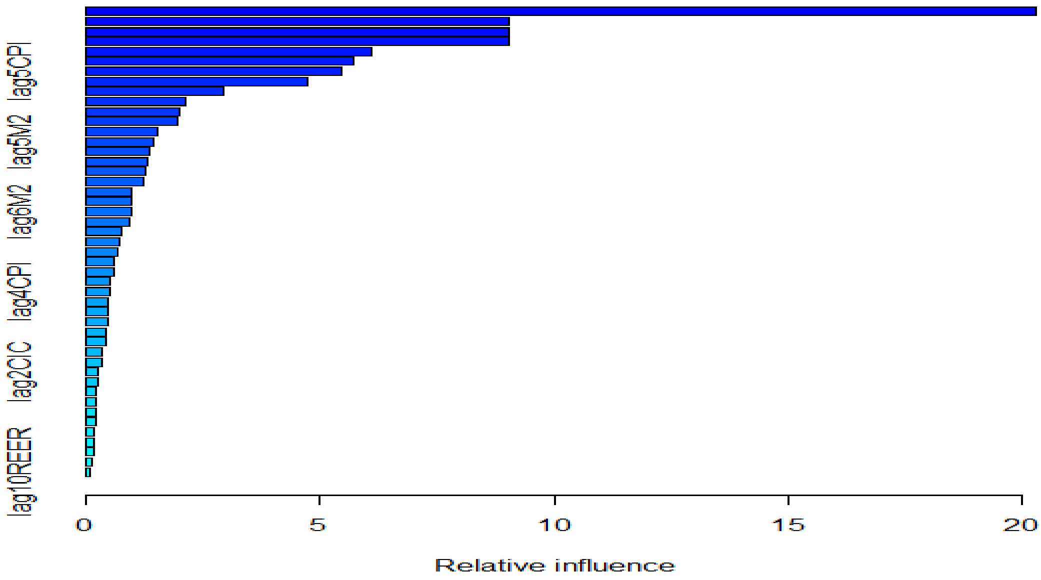

Figure 4.10: Boosting of Money Supply

The Figure 4.10 of relative influence shows that the highest bar is most important variable. In this graph lag3 of M2, lag8 of CIC, lag2 of M2, and CIC are the most important and so on.

4.2.3 Summary of Boosted Regression Tree Real Effective Exchange Rate as Dependent Variable

Variables are used in the BRT real effective exchange rate (REER) is used as dependent variable inflation, money supply (M2), currency in circulation (CIC), their lags and the lags of dependent variable is used as regressors. A gradient boosted model, with gaussian loss function 5000 iteration were performed and there are 47 regressors out of which 47 had non zero influence.

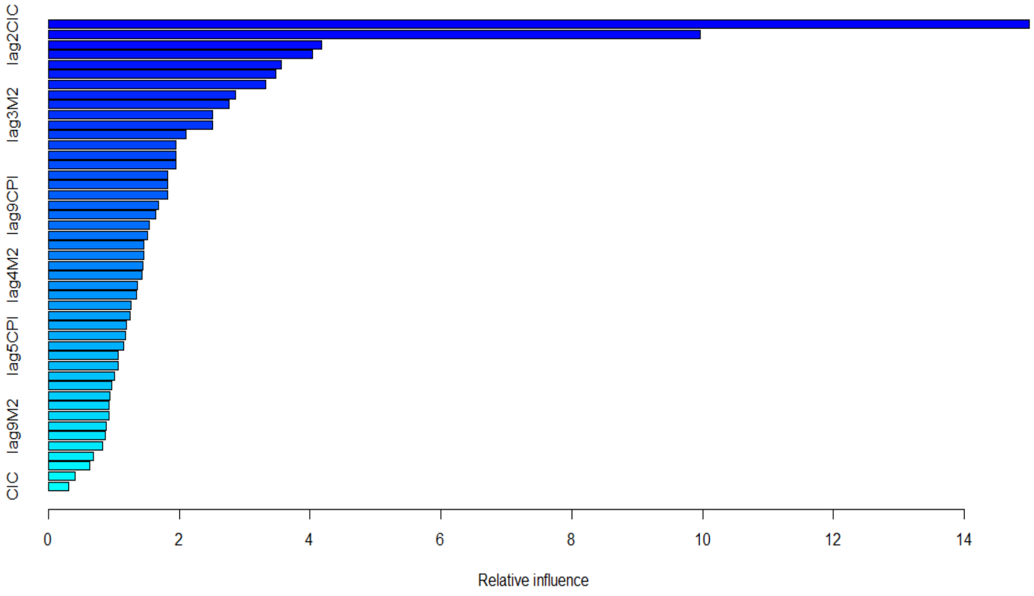

Figure 4.11: Boosting of Real Effective Exchange Rate

In the Above Figure 4.11 the highest bar shows the importance of variables. First or the highest bar show lag 2 of M2, then lag 2 and 3 of CIC and so on. The mean square error of the boosted regression tree is 0.00043.

4.2.4 Summary of Boosting Regression Tree Currency in Circulation (CIC) as Dependent Variable

Variables are used in the boosted regression tree currency in circulation (CIC) is used as dependent variable inflation, real effective exchange rate (REER), money supply (M2), their lags and the lags of dependent variable is used as regressors. A gradient boosted model, with gaussian loss function 5000 iteration were performed and there are 47 regressors are used of which 47 had non zero influence.

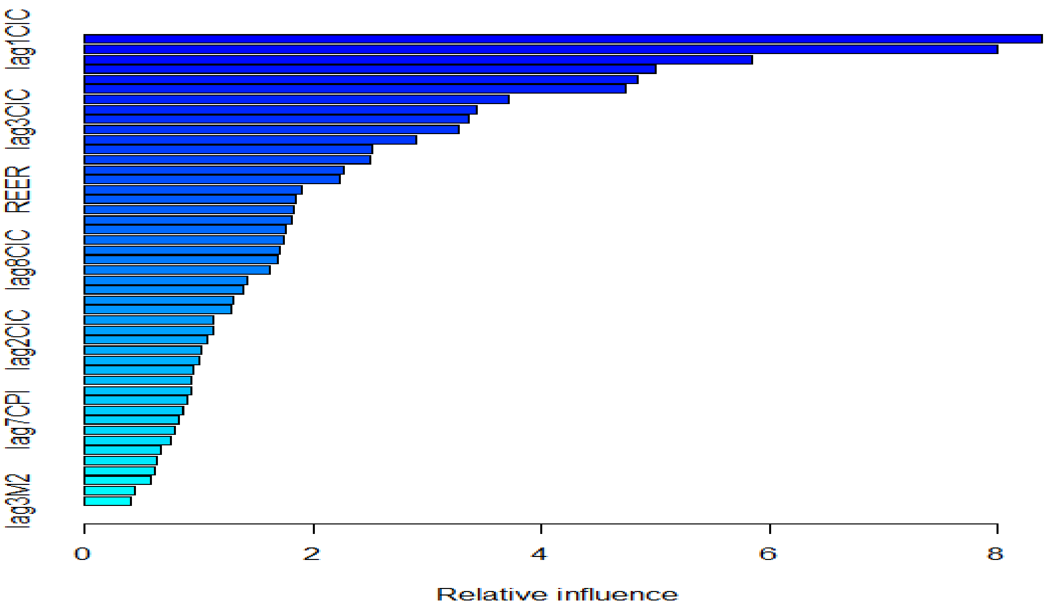

Figure 4.12: Boosting of Currency in Circulation

The Above Figure 4.12 of relative influence shows that the highest bar shows the most important variables. First or the highest bar i-e, lag 1 of CIC, then lag 1 and 2 of REER and so on.

4.3 Auto Regressive Model (AR)

4.3.1 Inflation

To check the stationarity and nonstationary of the data set ADF test is applied following table shows the statistics for ADF test.

Table 4.9 ADF Test Result

Series	Deterministic part	Lags	Tcal	Integration order
Inflation	None	6	-5.14	I(0)

4.3.1.1 Identification of input lags

Identification for AR lags Partial autocorrelation (PACF) function is used, Figure 1 represents the ACF.

Figure 4.11: ACF for CPI series

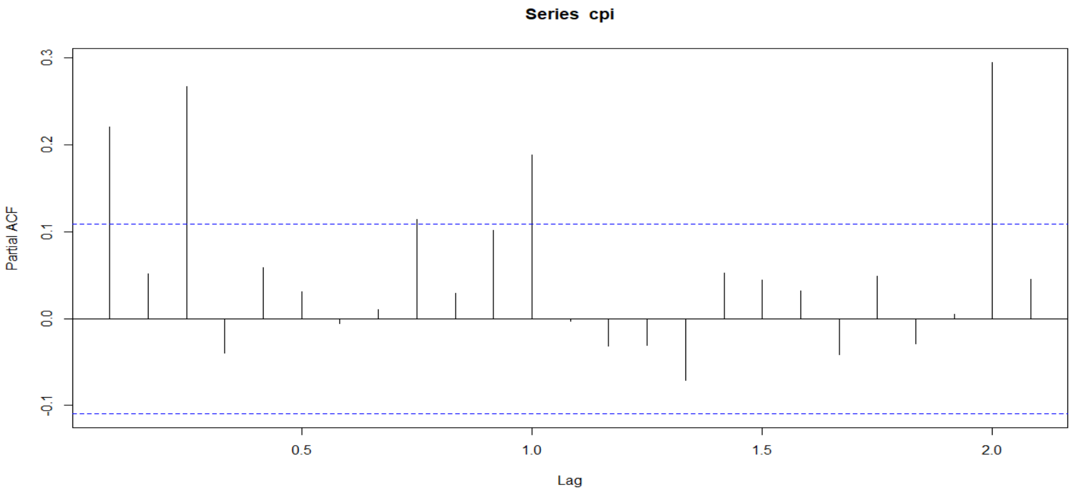

Figure 4.13: ACF of Inflation

PACF for CPI series is given up to 24 lags in which dotted line shows the $\pm\frac{2}{\sqrt{T}}$ confidence interval. If PACF spikes lie outside this dotted lien then it can be concluded that lags are significant otherwise not. It can be seen from the plot that spikes at nonseasonal lags 1 and 3 and seasonal lags 12 and 24 exceeds the significance bound. The estimated coefficients for obtained SAR models is given in Table 4.9.1 as:

Table 4.10 SAR Results for Inflation

Lags	**Coefficients**	**Standard error**	**t-statistics**
C	0.0069	0.0009	7.666667
AR(1)	0.1781	0.0544	3.273897
AR(3)	0.2345	0.0545	4.302752
SAR(1)	0.2321	0.0573	4.050611
Sigma^2 estimated	4.731e-05	AIC	-2203.74
Log likelihood	1106.87	JB test (P-value)	0.2062

Where sigma^2 is the variance assumed by the model, loglikelihood is the probability of desired data to be observed by the model and AIC is the value for to choose the bes t model among various chosen models. In the next step 1 to 12 ahead forecasts are done. JB test (p-value >.05) shows that residual are normally distributed. The RMSE and graphical representation for actual versus forecast values are given as:

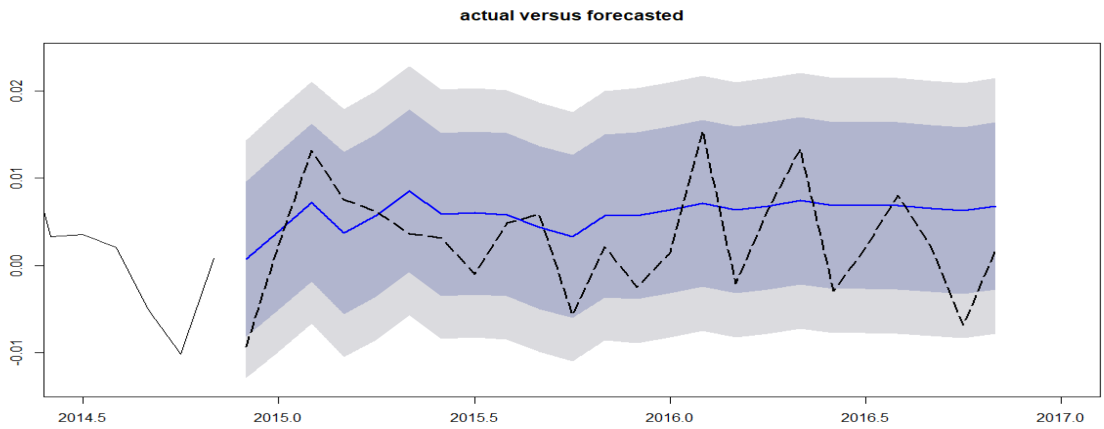

Figure 4.14: Forecasting

4.3.2 Real Effective Exchange Rate

Table 4.11 ADF Test Results

Series	Deterministic part	Lags	Tcal	Integration order
REER	None	6	-9.305	I(0)

4.3.2.1 Identification of input of lags

Identification for AR lags Partial autocorrelation (PACF) function is used, Figure 1 represents the ACF.

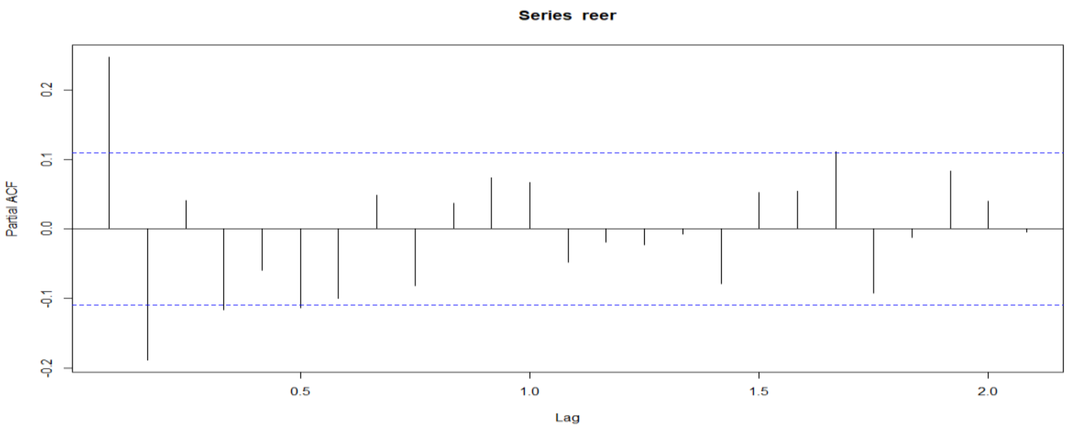

Figure 4.15: ACF of Real Effective Exchange Rate

PACF for REER series is given up to 24 lags in which dotted line shows the $\pm\frac{2}{\sqrt{T}}$ confidence interval. If PACF spikes lie out side this dotted lien then it can be concluded that lags are significant otherwise not. It can be seen from the plot that spikes at nonseasonal lags 1 and 2 the significance bound. The significant values of lag 1 and 2 are shows below the table.

Table 4.12 SAR Results for Real Effective Exchange Rate

Lags	**Coefficients**	**Standard error**	**t-statistics**
AR(1)	0.2872	0.0570	5.0386
AR(2)	-0.1936	0.0571	-3.391
Sigma^2 estimated	0.0002579	AIC	-1610.49
Log likelihood	808.24	JB test (P-value)	0.0887

Where sigma^2 is the variance assumed by the model, loglikelihood is the probability of desired data to be observed by the model and AIC is the value for to choose the best model among various chosen models. In the next step 1 to 12 ahead forecasts are done.

JB test p-value 0.0887 shows that residual are normally distributed.

4.3.3 Currency in Circulation

Table 4.13 ADF Test Results

Series	Deterministic part	Lags	Tcal	Integration order
Currency in circulation	None	6	-11.189	I(0)

4.3.3.1 Identification of lags

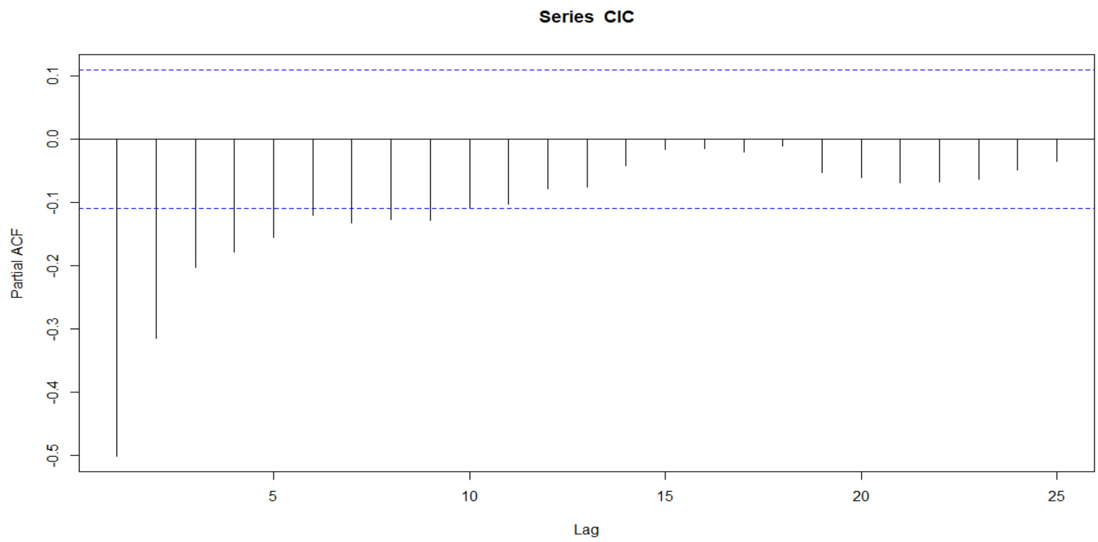

Figure 4.16: ACF of Currency in Circulation

Above Figure 4.17 shows that the five significant lags, as spikes of given series are lies in the significant area of partial-ACF. Five lags results are as fallows.

Lags	Coefficients	Standard error	t-statistics
AR(1)	-0.7862	0.0572	-13.745
AR(2)	-0.5789	0.0710	-8.153521
AR(3)	-0.4126	0.0747	-5.523427
AR(4)	-0.2944	0.0707	-4.164074
AR(5)	-0.1537	0.0567	-2.710758
sigma^2 estimated	0.06692	aic	54.71
log likelihood	-20.36	Jb test	0.143

Where sigma^2 is the variance assumed by the model, loglikelihood is the probability of desired data to be observed by the model and AIC is the value for to choose the best model among various chosen models. In the next step 1 to 12 ahead forecasts are done. JB test p-value 0.143 which is greater than 0.05 which shows that residual are normally distributed.

4.3.4 Money Supply

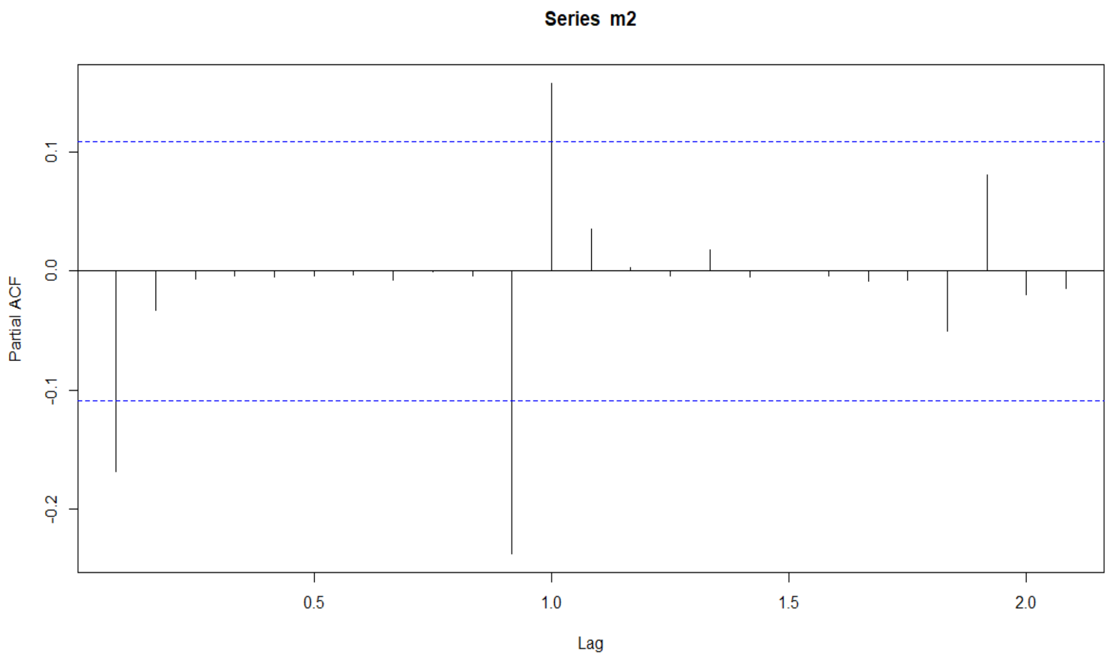

Figure 4.17: ACF of Money Supply

Above Figure 4.18 shows that the five significant lags, as spikes of given series are lies in the significant area of partial-ACF. Five lags results are as fallows.

Lags	**Coefficients**	**Standard error**	**t-statistics**
AR(1)	-0.1684	-0.0542	-3.107011
AR(11)	-0.2218	0.0533	-0.5714286
sigma^2 estimated	0.211	aic	407.27
log likelihood	-199.63	Jb test	0.091

Where sigma^2 is the variance assumed by the model, loglikelihood is the probability of desired data to be observed by the model and AIC is the value for to choose the best model among various chosen models. In the next step 1 to 12 ahead forecasts are done. JB test p-value 0.091 which is greater than 0.05 which shows that residual are normally distributed.

4.4 Adaptive LASSO for information set i = 4

4.4.1 Using Inflation as Dependent Variable

Table 4.13 Adaptive LASSO Results of Inflation

	Variables	**Adaptive LASSO**
M2	Money supply	0.0013
CIC	Currency in Circulation	0.0022
LAG1CPI	Lag1 of inflation	0.2055
LAG1CIC	Lag1 of currency in circulation	-0.0012
LAG1REER	Lag1 of real effective exchange rate	-0.0522
LAG2CPI	Lag2 of Inflation	-0.0004
LAG3CPI	Lag3 of inflation	0.2007
LAG7M2	Lag7 of Money Supply	0.0011
LAG8REER	Lag8 of Real Effective Exchange rate	-0.0316
LAG9CPI	Lag9 of inflation	0.1224
LAG9REER	Lag9 Real Effective Exchange Rate	-0.0233
LAG11CPI	Lag11 inflation	0.0824
LAG11M2	Lag11 Money Supply	-0.00004
LAG11REER	Lag11 of Real Effective Exchange Rate	-0.04401

In the Above Table 4.15 it shows the inflation is used as dependent variable whereas money supply (M2), REER (REER), currency in circulation (CIC), their lags and the lags of dependent variable are used as regressors. The Adaptive LASSO is ignoring the high correlation between variables. There are 48 variables used for the Adaptive LASSO. Among all variables M2, lag 11 of M2,CIC, lag 1,2 3 9,11 of inflation, lag 1 0f CIC, lag 1, 8, 9 11 of REER are significant which means that they are important variables and the forecasted mean square error is 0.000033.

4.4.2 Using Money Supply as Dependent Variable

Here money supply (M2) is used as dependent variable whereas inflation, real effective exchange rate (REER), currency in circulation (CIC), their lags and the lags of dependent variable are used as regressors. The Adaptive LASSO is ignoring the high correlation between variables. There are 48 variables used for the Adaptive LASSO. All variables are insignificant which means that there is no effect on money supply.

4.4.3 Using Currency in Circulation as Dependent Variables

Table 4.14 Adaptive LASSO Results of Currency in Circulation

	Variables	Adaptive LASSO
Lag1 M2	Lag1of Money Supply	-0.4985

In the Above Table 4.16 it shows the currency in circulation (CIC) is used as depende nt variable whereas money supply (M2), real effective exchange rate (REER), inflatio n, their lags and the lags of dependent variable are used as regressors. The Adaptive L ASSO is ignoring the high correlation between variables. There are 48 variables used for the Adaptive LASSO. From all variables lag1 of M2 is important and significant v ariable and the mean square error is 0.0014.

4.4.4 Using Real Effective Exchange Rate as Dependent Variable

Table 4.15 Adaptive LASSO Results of Real Effective Exchange Rate

	Variables	Adaptive LASSO
CIC	Lag1 of currency in circulation	1.9423

In the Above Table 4.17 it shows the REER (REER) is used as dependent variable wh ereas, money supply (M2), inflation, currency in circulation (CIC), their lags and the l ags of dependent variable are used as regressors. The Adaptive LASSO is ignoring the high correlation between variables. There are 48 variables used for the Adaptive LAS SO. From among all variables lag1 of CIC is important and significant variable. The mean square error is 0.00012.

4.5 Bagging Using Information set i = 4

4.5.1 Using Inflation as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 4.

Table 4.16 Bagging Results of Inflation for Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	47
Mean of square residual	0.000048
% variation explained	12.75

Bagging method is the one of assemble method its give better prediction than the RT but in our case mean square of residual of bagging method is greater than RT. So we use another assembling method like Random forest and boosting.

b) On Test Data
The data used for test data set is from 2015 M1- 2016M2 and information set is i=4.

Table 4.17 Bagging Results of Inflation for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	5
Mean of square residual	0.000031
% variation explained	5.7

Mean square error of the test data is also greater than regression tree.

4.5.2 Bagging Using Money Supply as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 4.

Table 4.18 Bagging Results of Money Supply for Train Data

Type of random forest	Regression
No. of trees	500
No. of variables tried at each split	47
Mean of square residual	0.3224

% variation explained	32.78

Bagging method is the one of assemble method it's give better prediction than the RT but in our case mean square of residual of bagging method is greater than RT. So we use another assembling method like Random forest and boosting.

b) On Test Data

The data used for test data set is from 2015 M1- 2016M2 and information set is i=4.

Table 4.19 Bagging Results of Money Supply for Test Data

Type of random forest	Regression
No. of trees	500
No. of variables tried at each split	47
Mean of square residual	0.00026
% variation explained	37.28
Test data of Mean square error is also greater than regression tree.	

4.5.3 Bagging Using REER as dependent variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 4.

Table 4.20 Bagging Results of Real Effective Exchange Rate for Train Data

Type of random forest	Regression
No. of trees	500
No. of variables tried at each split	47
Mean of square residual	0.000263
% variation explained	6.7

Bagging method is the one of assemble method it's give better prediction than the RT but in our case mean square of residual of bagging method is greater than RT. So we u se another assembling method like Random forest and boosting

b) On Test Data

The data used for test data set is from 2015 M1- 2016M2 and information set is i=4.

Table 4.21 Bagging Results of Real Effective Exchange Rate for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	47
Mean of square residual	0.000014
% variation explained	33.75

Test data of Mean square error is also greater than regression tree.

4.5.4 Bagging using currency in circulation as dependent variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 4.

Table 4.22 Bagging Results of Currency in Circulation for Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	32
Mean of square residual	0.1111
% variation explained	0.14

Variable use in the bagging method are inflation ,money supply ,REER and currency in circulation used as a dependent variable. Mean of square residual is consider as a error rate of the model. Error rate of bagging technique on the train data set is greater than RT so bagging technique give not better prediction than the RT.

b) On Test Data

The data used for test data set is from 2015 M1- 2016M2 and information set is i=4.

Table 4.23 Bagging Results of Currency in Circulation for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	32
Mean of square residual	0.00085
% variation explained	11.61

Again we apply bagging method on test data set results show that the mean square residual of test data set is less than train data set that indicates that there is no overfitting problem.

4.6 Random Forest Using information set i= 4

4.6.1 Inflation Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 4.

Table 4.24 Random Forest Results of Inflation for Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	16

Mean of square residual	0.000047
% variation explained	14.59

Random forest is an assemble method. It can provide better forecasting. Type of random forest is the RT because our dependent variables is numerical. We have choose n 500 number of trees randomly mean of square residual indicated the error rate of model. The error is 0.000047 is slightly less than bagging.

b)On Test Data

The data used for test data set is from 2015 M1- 2016M2 and information set is i=4.

Table 4.25 Random Forest Results of Inflation for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	5
Mean of square residual	0.000033
% variation explained	0.2

While applying technique on test data set the mean square residual is less than train data set. That's indicate that model is not over fitting because we will easily see that the mean square of error is less than train data set.

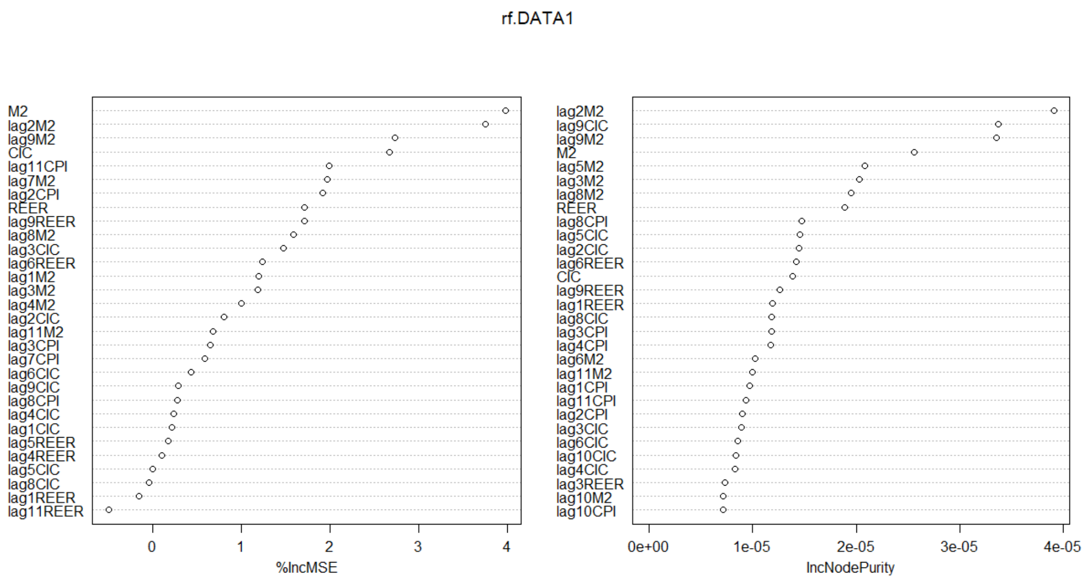

Figure 4.18: Important Variables

The Above Figure 4.18 shows the importance variable in each random forest, that how it's affect the node purity and MSE. This plot represents the variable importance table graphically. Money supply is the most important variable, then CIC, REER, lag1 of inflation and so on.

4.6.2 Random forest Using Money Supply as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 4.

Table 4.26 Random Forest Results of Money Supply for Train Data

Type of random forest	Regression
No. of trees	500
No. of variables tried at each split	16
Mean of square residual	0.2600
% variation explained	7.08

Random forest is an assemble method. It can provide better forecasting. Type of random forest is the RT because our dependent variables is numerical. We have choose n 500 number of trees randomly mean of square residual indicated the error rate of model. The error is 0.2600 is slightly less than bagging.

b) On Test Data

The data used for test data set is from 2015 M1- 2016M2 and information set is i=4.

Table 4.27 Random Forest Results of Money Supply for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	16
Mean of square residual	0.00024
% variation explained	42.11

While applying technique on test data set the mean square residual is less than train data set. That's indicate that model is not over fitting because we will easily see that the mean square of error is less than train data set.

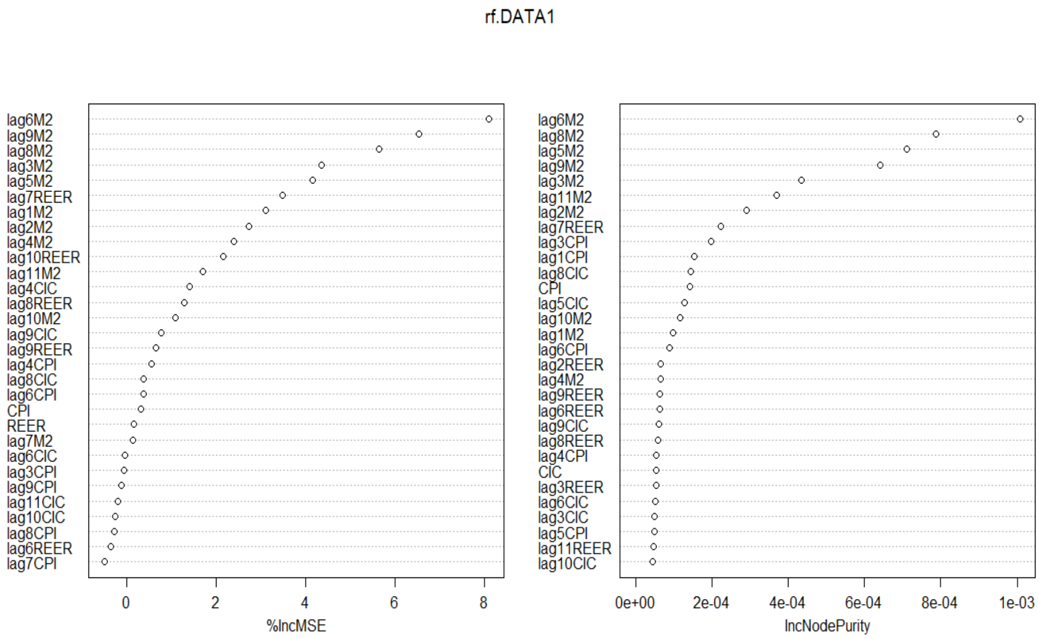

Figure 4 19: Important Variables

The Above Figure 4.20 shows the variable importance in random forest that how it affect the node purity and MSE. The most important variable is lag 6, 9 ,8 3 and 5 of M2 and so on. The least important variable is lag6 of REER and lag7 of inflation.

4.6.3 Random Forest Using REER as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 4.

Table 4.28 Random Forest Results of Real Effective Exchange Rate for Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	16
Mean of square residual	0.00026
% variation explained	7.25

Random forest is an assemble method. It can provide better forecasting. Type of random forest is the regression tree because our dependent variables is numerical. We have choose n 500 number of trees randomly. If the number of trees increase there is no effect on mean of square residual which indicated the error rate of model. The error is 0.000026 is slightly less than bagging.

a) On Test Data

Table 4.29 Random Forest Results of Real Effective Exchange Rate for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	47
Mean of square residual	0.00012

% variation explained	17.13

While applying technique on test data set the mean square residual is less than train data set. That's indicate that model is not over fitting because we will easily see that the mean square of error is less than train data set.

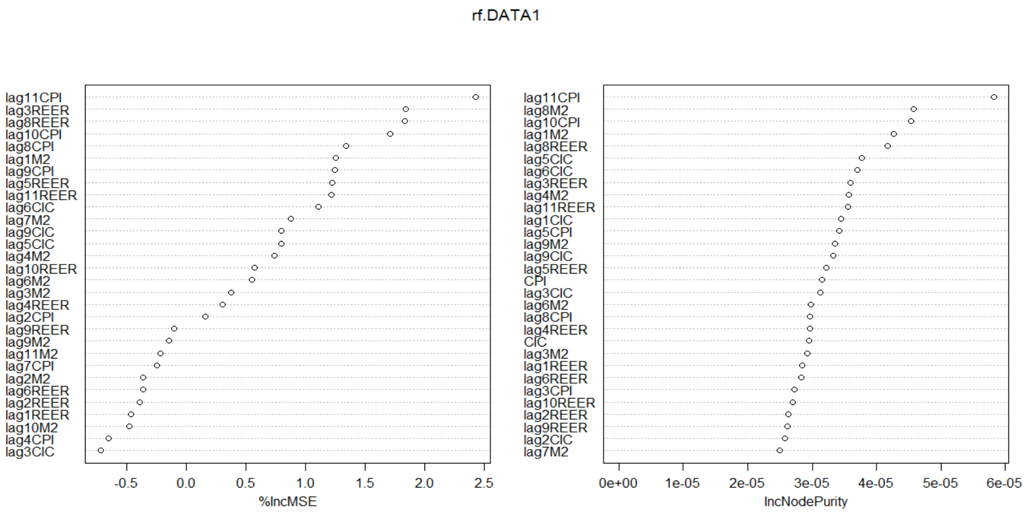

Figure 4 20: Important Variables

The Above Figure 4.21 shows the important of variables in random forest that how effect the node purity and MSE. Here the most important variable is lag 11 of inflation, then lag3 of REER, then lag8 of REER and so on. The least important variable is lag4 of CPI and lag3 of CIC.

4.6.4 Random Forest Using Currency in Circulation as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 4.

Table 4.30 Random Forest Results of Currency in Circulation for Train Data

Type of random forest	**Regression**
No. of trees	500

No. of variables tried at each split	16
Mean of square residual	0.107
% variation explained	4.05

Random forest is the common ensemble method technique its give better result than the other ensemble method like bagging and Boosting. Results show that the mean of square residual is less than RT ,bagging and Boosting .random forest give better prediction .

On Test Data

Table 4.31 Random Forest Results of Currency in Circulation for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	16
Mean of square residual	0.00102
% variation explained	5.75

When we apply on test data then result show that the error rate of random forest model is less than test data set .give more better prediction and there is no evidence of overfitting problem.

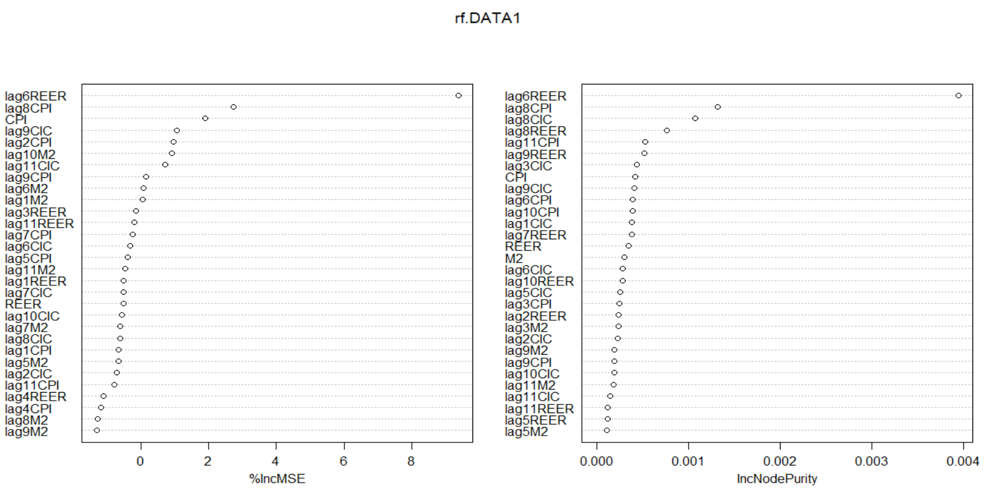

Figure 4 21: Important Variables

The Above Figure 4.22 shows the important of variables in random forest that how effect the node purity and MSE. Here the most important variables are lag6 of REER, lag9 of inflation, inflation, and so on. The least important variables are lag8 and lag9 of money supply.

4.7 Regression Tree for Information Set i = 14

4.7.1 Regression Tree Using Inflation as Dependent Variable

a) For Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 14.

Table 4.32 Regression Tree Results of Inflation for Train Data

Number of terminal nodes			25		
Residual mean deviance			0.000016 = 0.00426 / 269		
Distribution of residuals					
Min -0.0108	1st Qu -0.0026	Median -0.0003	Mean 0.000	3rd Qu 0.0026	Max 0.0131
Variables: lag10NG, lag11CPI, lag2NG, lag2CPI, lag4PGC, lag4CPI, lag8PRS, lag12PGC, lag6PGC, lag8CPI, lag6CIC, lag6ELEC, lag3PLS, lag2PC, lag10CPI, lag6PLS, lag7PG, lag4PCO, lag5REER, lag5CIC, lag6M2, lag5PC, lag2PRS, lag8PG					

The Above Table 4.32 shows that inflation is used as dependent variable whereas that currency in circulation (CIC), money supply (M2), REER (REER), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) and export of rice (EXR), its lags and the lags of dependent variable is used as regressors. From among all variables lag10NG, lag11CPI, lag2NG, lag2CPI, lag4PGC, lag4CPI, lag8PRS, lag12PGC, lag6PGC, lag8CPI, lag6CIC, lag6ELEC, lag3PLS, lag2PC, lag10CPI, lag6PLS, lag7PG, lag4PCO, lag5REER, lag5CIC, lag6M2, lag5PC, lag2PRS, lag8PG variables are important and is used for constructing the tree. The residuals mean deviance is 0.000016 which is used as mean square error.

4.7.1.1 Cross validation

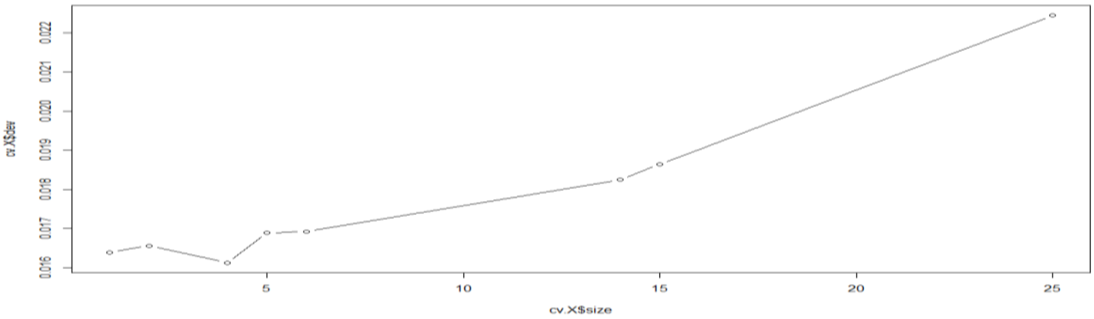

Figure 4 22: Cross Validation

The cross validation is used to check the size of tree among the deviance. Here the minimum deviance is 4.

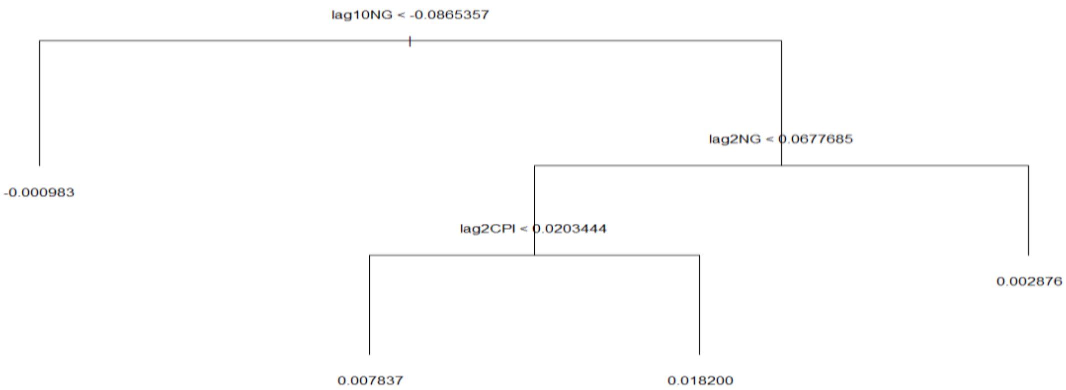

Figure 4 23: Regression Tree in Train Data

The Above Figure 4.24 shows that when lag10 of NG is less than -0.0865 the predicted inflation is -0.0009. On the other hand when lag10 of NG is greater than -0.0865 and lag2 of NG is less than 0.0678 and lag2 of inflation is less than 0.0203 the predicted inflation is 0.0078 and lag2 of inflation is greater than 0.0203 the predictor of inflation is 0.0182 and lag2 of NG is greater than 0.0678 the predicted inflation is 0.0029.

b) On Test Data

4.7.1.2 Summary of Regression Tree

Table 4.33 Regression Tree Results of Inflation for Test Data

Number of terminal nodes		3			
Residual mean deviance		0.000014 = 0.0002091 / 15			
Distribution of residuals					
Min -0.0056	1st Qu -0.0023	Median -0.0002	Mean 0.000	3rd Qu 0.0014	Max 0.0079
Variables: lag3PG, lag4PRS					

The Above Table 4.35 shows that Inflation is dependent variable whereas money supply (M2), real effective exchange rate (REER), currency in circulation (CIC),worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) and export of rice (EXR), their lags and the lags of dependent variable are used as regressors. Form among all the variables lag3 of PG and lag4 of PRS are important variables and is used for constructing the tree. Three terminal nodes are used and the mean deviance is 0.000014 which is used as mean square error.

4.7.2 Regression Tree Money Supply as Dependent Variables

a)On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 14.

Table 4.34 Regression Tree Results of Money Supply for Train Data

Number of terminal nodes			4		
Residual mean deviance			0.179 = 51.9 / 290		
Distribution of residuals					
Min -5.596	1st Qu -0.0079	Median 0.0018	Mean 0.000	3rd Qu 0.0173	Max 1.339
Variables: lag5NG, lag2M2, lag10CIC					

The Above Table 4.34 shows that broad money (M2) is dependent variable whereas , inflation, REER (REER) currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) and export of rice (EXR), their lags and the lags of dependent variable. From among all the variables lag5 of NG lag2 M2 and lag10 of CIC are important variables and is used for constructing the tree. There are four terminal node used for the tree and the mean deviance is 0.179 which is used for mean square error.

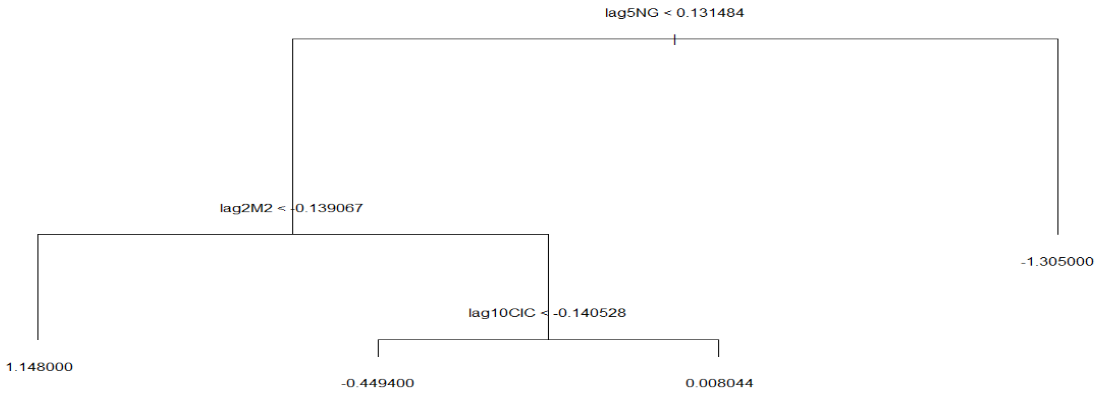

Figure 4.24: Regression Tree

The Above Figure 4.24 shows that when lag5 of NG is less than 0.1315 and lag2 of M2 is less than 0.1391 the predicted money supply is 1.1480. when lag2 of M2 is greater than 0.1391 and lag10 of CIC is less than -0.1405 the predicted money supply is 0.0080. on the other hand when lag5 of NG is greater than 0.1315 the predicted money supply is -1.3050.

b) On Test Data

Table 4.35 Regression Tree Results of Money Supply for Test Data

Number of terminal nodes	3
Residual mean deviance	6.488e-05 = 0.0009732 / 15
Distribution of residuals	

Min	1st Qu	Median	Mean	3rd Qu	Max
-0.0173	-0.0029	0.0002	0.000	0.0028	0.0218
Variables: lag3PG, lag5PC					

The Above Table 4.35 shows that money supply (M2) is dependent variable whereas currency in circulation (CIC), inflation, real effective exchange rate (REER), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) and export of rice (EXR), their lags and the lags of dependent variable are used as regressors. From among all variables lag3 of PG and lag5 of PC are important and is used for constructing the tree. Three terminal nodes are used and the mean deviance is 0.000065 which is used as mean square error.

4.7.3 Regression Tree Using REER as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 4.

Table 4.36 Regression Tree Results of real Effective Exchange Rate for Train Data

Number of terminal nodes			25		
Residual mean deviance			0.00009 = 0.02416 / 269		
Distribution of residuals					
Min	1st Qu	Median	Mean	3rd Qu	Max
-0.0238	-0.0059	0.000011	0.000	0.0057	0.0034
Variables: lag5CIC, lag2REER, lag10WR, lag2PLS, lag3REER, PG, lag12M2, lag6REER, lag9CIC, lag12PC, lag7WR, lag10PC, lag5PGC, lag9WR, lag3CPI, lag12CIC, lag2PG, lag8ELEC, lag7PCO, lag3CIC, PRS, lag3PCO					

The Above Table 4.38 shows that real effective exchange rate (REER) is dependent variable whereas money supply (M2), inflation, currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) and export of rice

(EXR), their lags and the lag of dependent variable are used as regressors are important variables and is used for constructing the tree. There are 25 terminal nodes are used and the mean deviance is 0.00009 which is used as mean square error.

4.7.3.1 Cross Validation

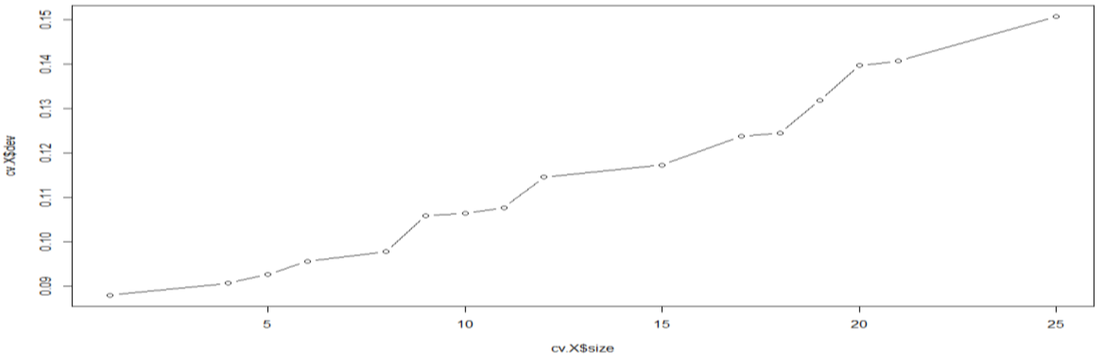

Figure 4 25: Cross Validation

The cross validation Figure 4.25 is used for the size of RT against the deviance. The minimum deviance is 2.

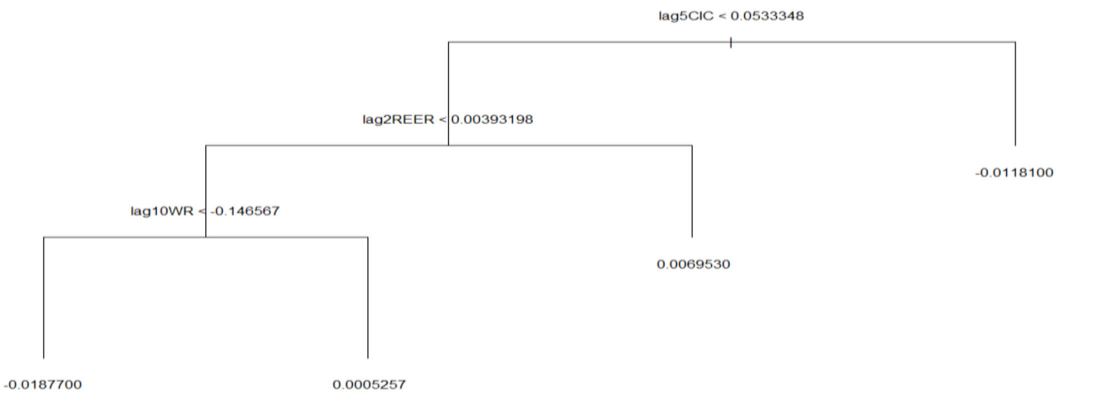

Figure 4 26: Regression Tree

The Above Figure 4.26 shows that when lag5 of CIC is less than 0.0533 and lag2 REER is 0.0039 and lag10 WR is less than-0.1466 the predicted REER is -0.0188 and lag10 of WR is greater than -0.0188 the predicted REER is 0.00053 and lag2 of REER is greater than 0.0039 the predicted REER is 0.0069. On the other hand when lag5 of CIC is greater than 0.0533 the predicted REER is -0.0018.

b)On Test Data

Table 4.37 Regression Tree Results of Real Effective Exchange Rate for Test Data

Number of terminal nodes			3		
Residual mean deviance			0.00004 = 0.00059 / 15		
Distribution of residuals					
Min -0.0148	1st Qu -0.0019	Median 0.0006	Mean 0.000	3rd Qu 0.0025	Max 0.0093
Variables: lag7PC, lag12REER					

The Above Table 4.37 shows that real effective exchange rate (REER) is dependent variable whereas money supply (M2), inflation, currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) and export of rice (EXR), their lags and the lags of dependent variable are used as regressors. From among all variables lag7 of PC and lag12 of REER are important and are used for constructing the tree. Three number of nodes are used and the mean deviance is 0.00004 which is used as mean square error.

4.7.4 Regression Tree Using Circulation in Circulation is used as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 14.

Table 4.38 Regression Tree Results of Currency in Circulation for Train Data

Number of terminal nodes			6		
Residual mean deviance			0.0706 = 20.33 / 288		
Distribution of residuals					
Min -1.750	1st Qu -0.0236	Median -0.0020	Mean 0.000	3rd Qu 0.0226	Max 1.850

Variables: lag2CIC, lag3EXPR, PRS, lag10PLS

The Above Table 4.38 shows that currency in circulation is dependent variable whereas money supply (M2), inflation, real effective exchange rate (REER), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) and export of rice (EXR), their lags and the lags of dependent variable are used as regressors. From among all variables lag2 of CIC, lag3 of EXPR, PRS and lag10 of PLS are the important variables and are used for constructing the tree and six terminal nodes are used in this tree. The mean deviance is 0.0706 which is used as mean square of error.

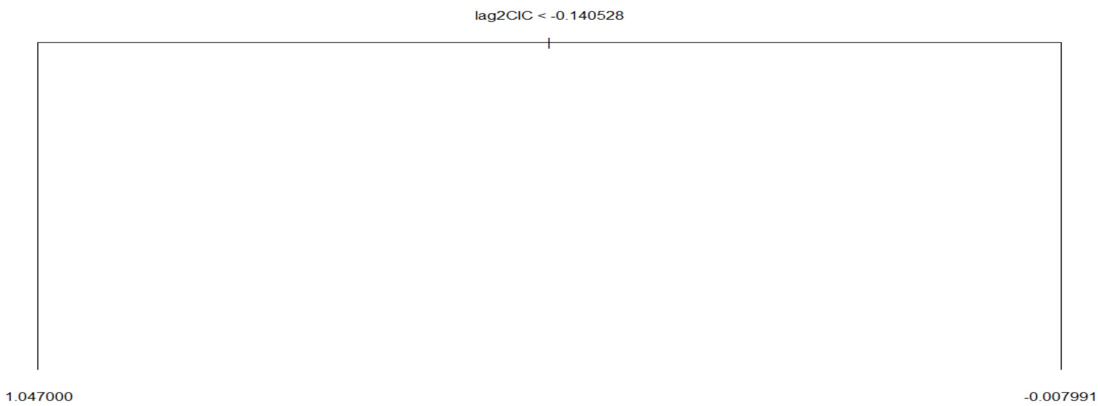

Figure 4 27: Regression Tree

The Above Figure 4.27 shows that when lag2 of CIC is than -0.1405 the predicted currency in circulation is 1.047 and lag2 of CIC is greater than -0.1405 the predicted currency in circulation is -0.0079.

b) On Test Data

Table 4.39 Regression Tree Results of Currency in Circulation for Test Data

Number of terminal nodes			3		
Residual mean deviance			0.00046 = 0.0068 / 15		
Distribution of residuals					
Min -0.0348	1st Qu -0.0111	Median 0.00015	Mean 0.000	3rd Qu 0.0099	Max 0.0514
Variables: lag12PGC, lag12CPI					

The Above Table 4.39 shows that currency in circulation is dependent variable whereas money supply (M2), inflation, real effective exchange rate (REER), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) and export of rice (EXR), their lags and the lag of dependent variable is used as regressors. From among all variables lag12 of PGC ang lag12 of inflation are important and is used for constructing the tree. Three terminals nodes are used and the residual mean deviance is 0.00046 which is used as mean square error.

4.8. Adaptive LASSO using Inflation as Dependent Variable

Table 4.40 Adaptive LASSO Results of Inflation

	Variable	**Adaptive LASSO**
Lag3 CIC	Lag3 currency in circulation	0.1983
Lag2 M2	Lag2 money supply	-0.0064
Lag10 CIC	Lag10 currency in circulation	0.2106

Above Table 4.40 shows that inflation is used as dependent variable whereas money supply (M2), currency in circulation (CIC), real effective exchange rate (REER), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) and export of rice (EXR), their lags and the lags of dependent variable are used as regressors. From among all variables lag3 and lag10 of CIC and lag2 of M2 are significant and important variables and the mean square error is 0.000039.

4.8.1 Adaptive LASSO using M2 as Dependent Variable

Here money supply (M2) is used as dependent variable whereas currency in circulatio n (CIC), inflation, REER (REER), worker remittances (WR), natural gas (NG), electri city (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), pro duction of cement (PC) and export of rice (EXR), their lags and the lags of dependent variable are used as regressors. From among all variable there is no one variable signi ficant and important variable. The mean square error is 0.00091.

4.8.2 Adaptive LASSO using REER as Dependent Variable

Table 4.41 Adaptive LASSO Results of Real Effective Exchange Rate

	Variables	**Adaptive LASSO**
Lag4 CIC	Lag4 currency in circulation	0.2644
Lag8CIC	Lag8 currency in circulation	-0.1465
Lag9 CIC	Lag9 currency in circulation	-0.0176
Lag10 NG	Lag10 natural gas	-0.0109
Lag9 PLS	Lag11 production of lime stone	0.2901
Lag9 PGC	Lag9 production of	-0.0114
Lag10 WR	Lag10 worker remittances	0.0125
Lag11 PLS	Lag11 of production of lime stone	-0.0121

Above Table 4.43 shows that real effective exchange rate (REER) is used as dependen t variable whereas money supply (M2), inflation, currency in circulation (CIC), worke r remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG) , production of lime stone (PLS), production of rock salt (PRS), production of crude o il (PCO), production of good coal (PGC), production of cement (PC) and export of ric e (EXR), their lags and the lags of dependent variable are used as regressors. From am ong all variables lag4, 8, 9, lag10 NG, lag 9, 11 PLS lag9 PGC and lag10 WR are imp ortant and significant variables. The mean square error is 0.00013.

4.8.3 Adaptive LASSO using Currency in Circulation as Dependent Variable

Table 4.42 Adaptive LASSO Results of Currency in Circulation

	Variables	**Adaptive LASSO**
Lag2 CIC	Lag2 currency in circulation	-0.5619
Lag3 CIC	Lag3 currency in circulation	-0.1756
Lag4 REER	Lag4 REER	0.7278

Above Table 4.44 shows that currency in circulation (CIC) is used as dependent varia ble whereas money supply (M2), inflation, REER (REER), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) and export of rice (EXR), their lags a nd the lags of dependent variable are used as regressors. From among all variables lag 2, lag3 CIC and lag4 REER are significant and important variables. The mean square error is 0.0016.

4.8.4 Boosting for Inflation Using Information Set i = 14

4.8.4.1 Summary of Boosting

Variables are used in boosted regression tree inflation as dependent variables whereas money supply (M2), real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement

(PC) and export of rice (EXR), their lags and the lags of dependent variable are used as regressors. A gradient boosted model, with gaussian loss function 5000 iteration were performed and there are 47 regressors are used of which 47 had non zero influence.

4.8.5 Bagging using Inflation as Dependent Variable Information Set i=14

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 14.

Table 4.43 Bagging Results of Inflation for Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	167
Mean of square residual	0.000047
% variation explained	14.49

Variable use in the bagging method are money supply ,REER and inflation used as a dependent variable. Mean of square residual is consider as error rate of the model. Error rate of bagging technique on the train data set is greater than RT so bagging technique give not better prediction than the RT.

b) On Test Data

Table 4.44 Bagging Results of Inflation for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	167
Mean of square residual	0.000038
% variation explained	16.65

Again we apply bagging method on test data set results show that the mean square residual of test data set is less than train data set that indicates that there is no overfitting problem.

4.8.6 Random Forest Using Inflation as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 14.

Table 4.45 Random Forest Results of Inflation for Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	55
Mean of square residual	0.000046
% variation explained	16.1

In this case we use inflation as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and Boosting .

b) On Test Data

Table 4.46 Random Forest Results of Inflation for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	55
Mean of square residual	0.000036
% variation explained	9.06

We apply random forest on test data set then results indicate that mean square of residual is 0.000036 which is less than train data set so this results indicates that random forest give better prediction other the ensemble method and there is no evidence of overfitting problem.

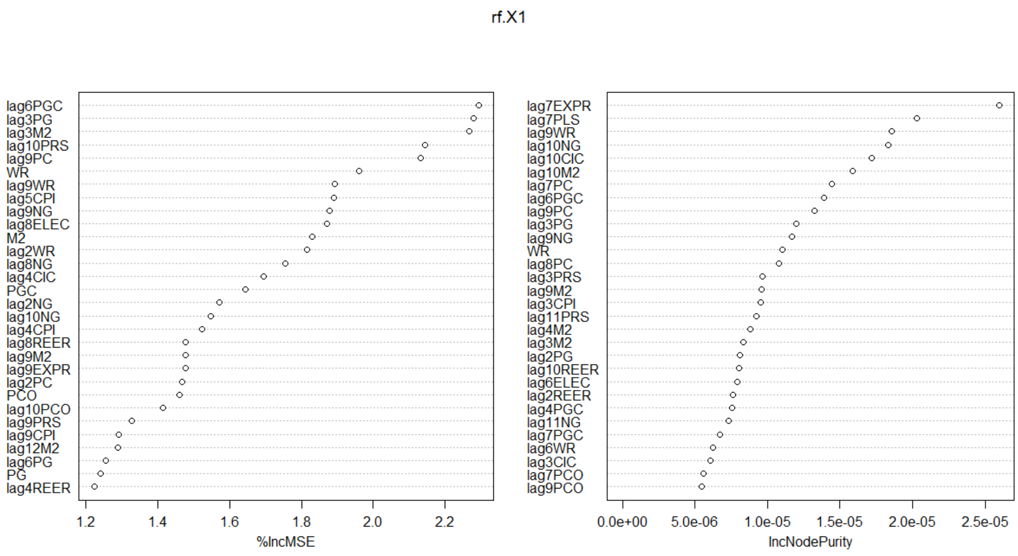

Figure 4 28: Important Variables

The Above Figure 4.28 shows the important of variables in random forest that how effect the node purity and MSE. The most important variables are lag6 of PGC, lag3 of PG and so on. The least important variable are PG and lag4 of REER.

4.8.7 Bagging Using Money Supply as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 14.

Table 4.47 Bagging Results of Money Supply for Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	147
Mean of square residual	0.3394
% variation explained	39.76

In this case we use money supply as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and boosting .

b) On Test Data

Table 4.48 Bagging Results of Money Supply for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	147
Mean of square residual	0.00019
% variation explained	53.61

Again we apply bagging method on test data set results show that the mean square residual of test data set is less than train data set that indicates that there is no overfitting problem.

4.8.8 Random Forest Using Money Supply as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 14.

Table 4.49 Random Forest Results of Money Supply for Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	55
Mean of square residual	0.2731
% variation explained	12.49

In this case we use inflation as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and Boosting .

b) On Test Data

Table 4.50 Random Forest Results of Money Supply for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	55
Mean of square residual	0.00021
% variation explained	49.85

We apply random forest on test data set then results indicate that mean square of residual is 0.00021 which is less than train data set so this results indicates that random forest give better prediction other the ensemble method and there is no evidence of overfitting problem.

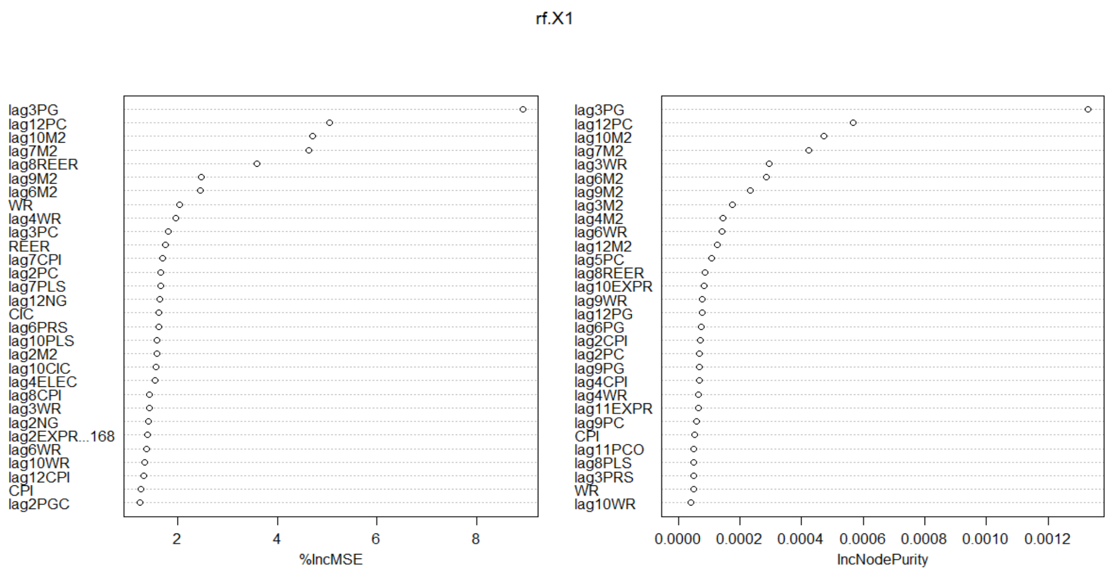

Figure 4 29: Important Variables

The Above Figure 4.29 shows the important of variables in random forest that how eff ect the node purity and MSE. The most important variables are lag3 of PG, then lag12 of PC and so on. The least important variables are inflation and lag2 of PGC.

4.8.9 Bagging Using REER as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 4.

Table 4.51 Bagging Results of Real Effective Exchange Rate for Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	167
Mean of square residual	0.000272
% variation explained	3.54

In this case we use REER as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and Boosting .

b) On Test Data

Table 4.52 Bagging Results of Real Effective Exchange Rate for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	167
Mean of square residual	0.00014
% variation explained	35.62

Again we apply bagging method on test data set results show that the mean square residual of test data set is less than train data set that indicates that there is no overfitting problem.

4.8.10 Random Forest Using REER as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 14.

Table 4.53 Random Forest Results of Real Effective Exchange Rate for Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	55
Mean of square residual	0.00027
% variation explained	1.93

In this case we use inflation as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and Boosting .

b) On Test Data

Table 4.54 Random Forest Results of Real Effective Exchange Rate for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	55
Mean of square residual	0.00013
% variation explained	24.67

We apply random forest on test data set then results indicate that mean square of residual is 0.00013 which is less than train data set so this results indicates that random forest give better prediction other the ensemble method and there is no evidence of overfitting problem.

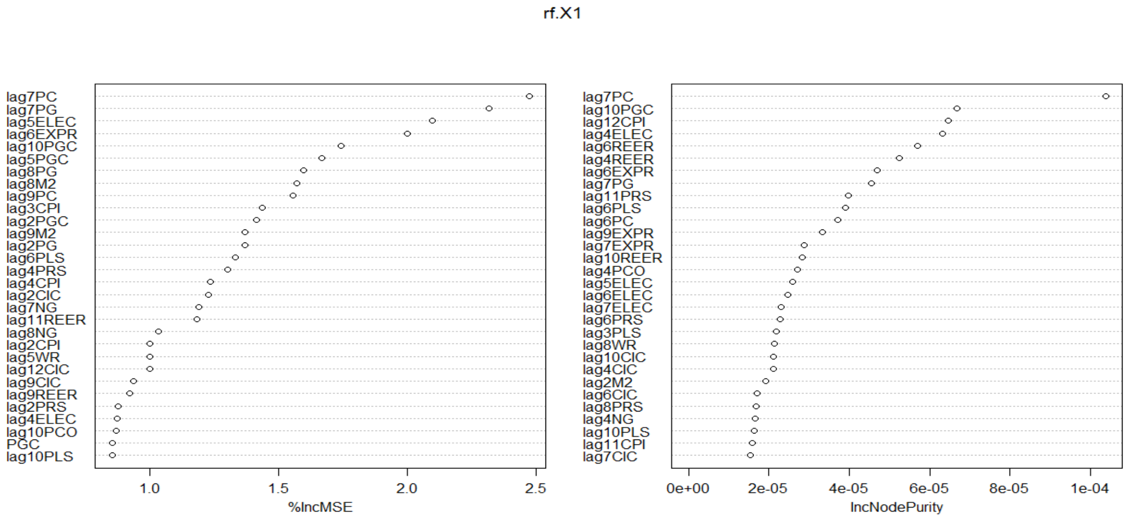

Figure 4 30: Important Variables

The Above Plot 4.30 shows the important of variables in random forest that how effec t the node purity and MSE. The most important variables are lag7 of PC, then lag7 of PG and so on. The least important variables are PGC and lag10 of PLS.

4.8.11 Bagging Using Currency in Circulation as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 14.

Table 4.55 Bagging Results of Currency in Circulation for Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	167
Mean of square residual	0.1041
% variation explained	6.42

In this case we use currency in circulation as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and Boosting .

b) On Test Data

Table 4.56 Bagging Results of Currency in Circulation for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	167
Mean of square residual	0.000091
% variation explained	5.62

Again we apply bagging method on test data set results show that the mean square residual of test data set is less than train data set that indicates that there is no overfitting problem.

4.8.12 Random Forest Using Currency in Circulation as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 14.

Table 4.57 Random Forest Results of Currency in Circulation for Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	55
Mean of square residual	0.1031
% variation explained	7.32

In this case we use inflation as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and Boosting .

b) On Test Data

Table 4.58 Random Forest Results of Currency in Circulation for Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	55
Mean of square residual	0.00102
% variation explained	5.96

We apply random forest on test data set then results indicate that mean square of residual is 0.00102 which is less than train data set so this results indicates that random forest give better prediction other the ensemble method and there is no evidence of overfitting problem.

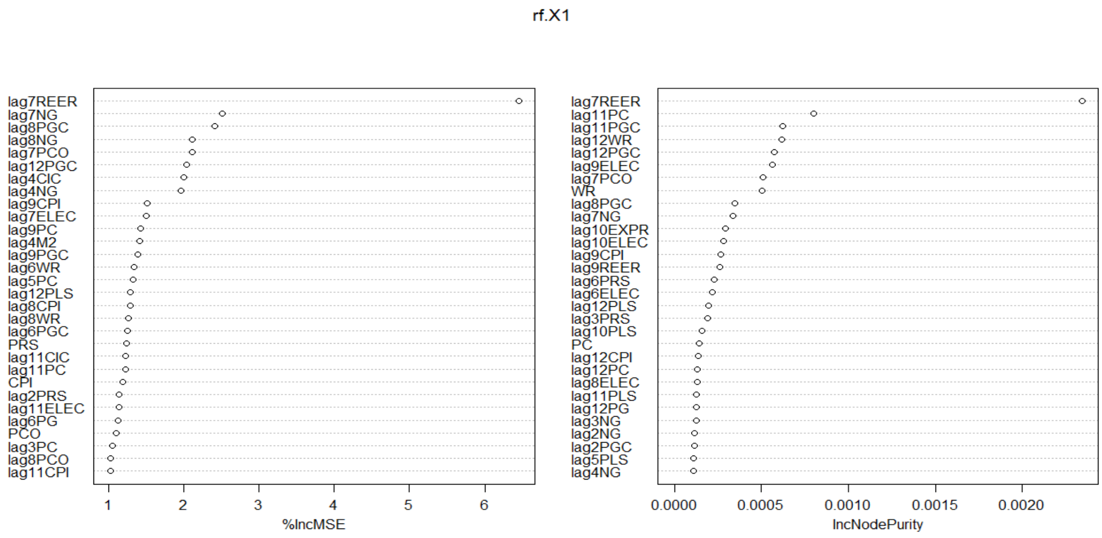

Figure 4 31: Important Variables

The Above Plot 4.31 shows the important of variables in random forest that how effect the node purity and MSE. The most important variables are lag7 of REER, then lag7 of NG and so on. The least important variables are lag8 of PCO and lag11 of inflation.

4.8.13 Summary of Boosting Using Inflation as Dependent Variable

Variables are used in the boosted RT inflation is used as dependent variable whereas money supply (M2), real effective exchange rate (REER), currency in circulation

(CIC), worker remittances (WR), natural gas (NG), electricity(ELEC), production of gypsum (PG), production of lime stone(PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC), export of rice(EXPR), their lags and the lags of dependent variable is used as regressors. A gradient boosted model, with gaussian loss function 5000 iteration were performed and there are 167 regressors are used of which 167 had non zero influence.

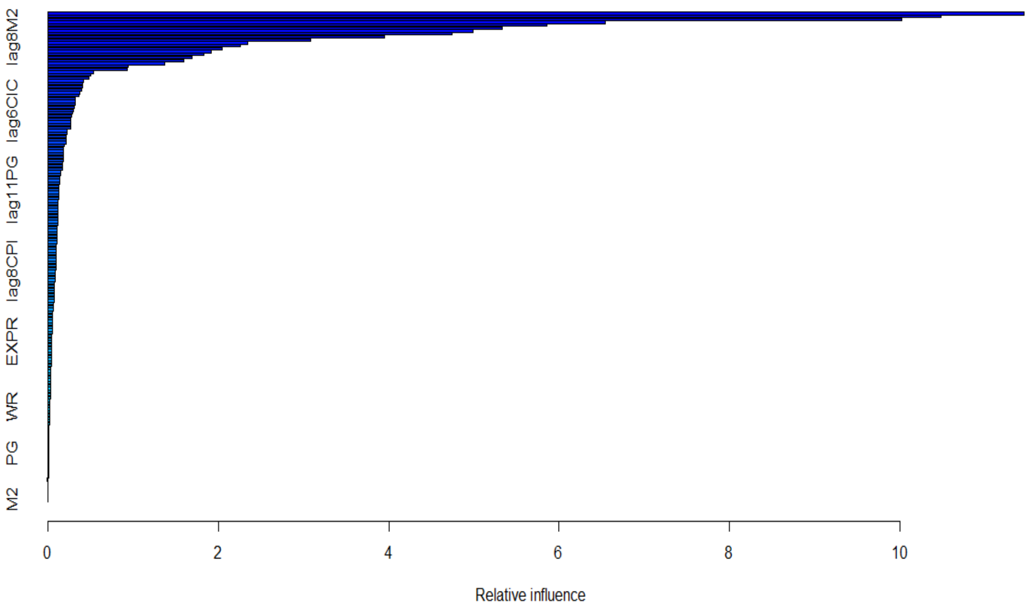

Figure 4 32: Boosted Regression Tree

The Above Figure 4.32 shows the height of the bar relative importance of the variables. First or the highest bar shows, lag8 of M2 is the most significant variable and influential variable, lag6 of CIC is second significant and important variable showed by the second highest bar and so on. The mean square error is 0.00002.

4.8.14 Summary of Boosting Using Money Supply as Dependent Variable

Variables are used in the boosted regression tree money supply is used as dependent variable whereas inflation, real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) and export of rice (EXR), their lags and the lags of dependent variable is used as regressors. A gradient boosted model, with gaussian loss function 5000 iteration were performed and there are 167 regressors are used of which 167 had non zero influence.

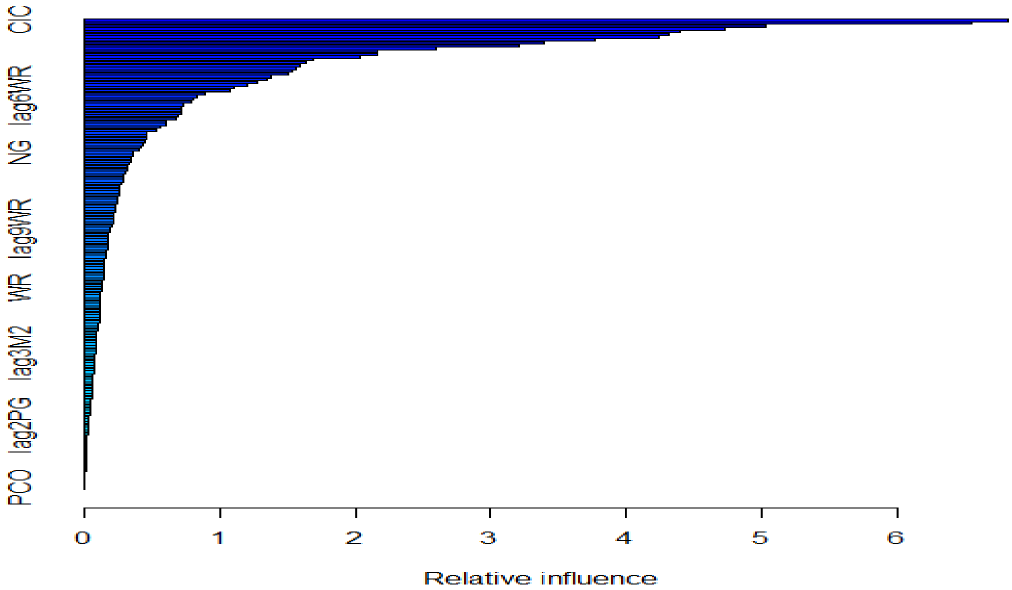

Figure 4 33: Boosted Regression Tree

The Above Figure 4.33 shows the highest of the relative influence of the highest bar o f the variables. CIC is the significant variable showed by the first highest bar. Second is the lag6 WR showed by second highest bar. The last bar showed least significant im portant. The mean square error is 0.0004.

4.8.15 Boosting Using REER as Dependent Variable

Variables are used in the boosted regression tree REER is used as dependent variable whereas inflation, money supply (M2), real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) ,their lags and the lags of dependent variable is used as regressors. A gradient boosted model, with gaussian loss function 5000 iteration were performed and there are 167 regressors are used of which 167 had non zero influence.

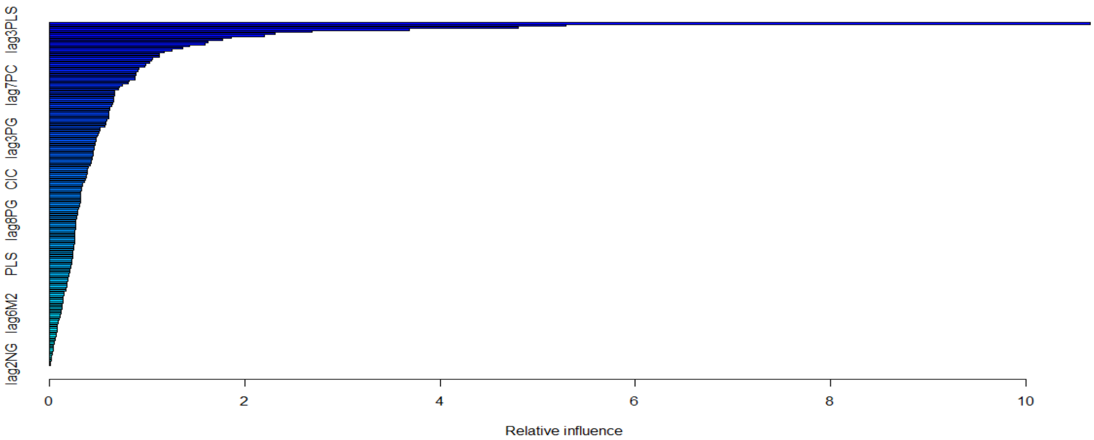

Figure 4 34: Boosted Regression Tree

The Above Figure 4.35 shows the relative influence of the variables. The first highest bar show that lag3 of PLS is the most important variable. Second highest bar shows lag7 of PC is second significant variable and so on. The last bars shows that those variables are least important. The mean square error is 0.00001

4.8.16 Summary Boosting Regression Tree Using Currency in Circulation as Dependent Variable

Variables are used in the boosted regression tree currency in circulation (CIC) is used as dependent variable whereas inflation, money supply(M2), real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR), their lags and the lags of dependent variable is used as regressors. A gradient boosted model, with gaussian loss function 5000 iteration were performed and there are 167 regressors are used of which 167 had non zero influence.

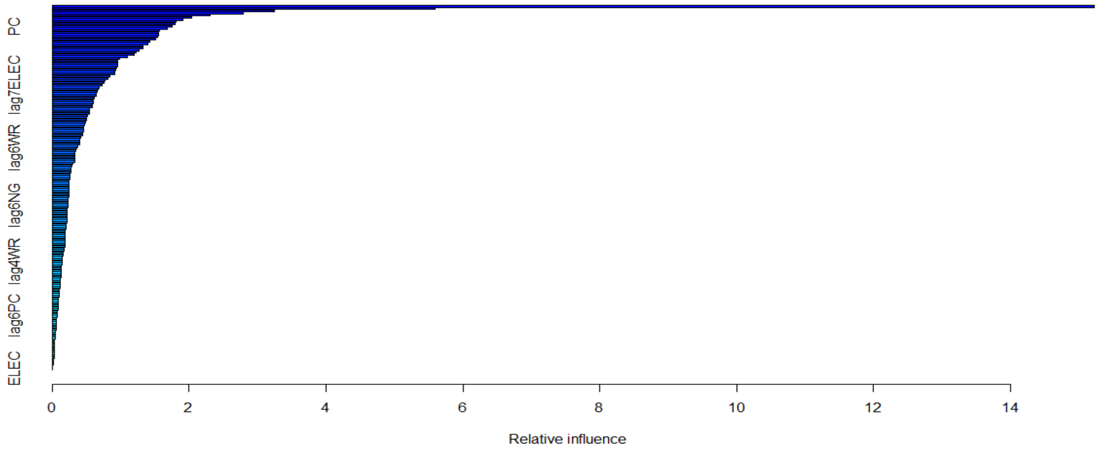

Figure 4 35: Boosted Regression tree

The Above Figure 4.36 shows the relative influence of variables. The first highest bar shows that PC is most significant variable. Second bar shows that lag REER is second significant variable and so on. Last bars shows the least significant variables. The mean square error is 0.00005.

4.9 Regression Tree for information set i = 31

4.9.1 Using Inflation as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 31.

Table 4 59: Regression Tree Result of Inflation on Train Data

Number of terminal nodes			27		
Residual mean deviance			0.000015 = 0.004016 / 267		
Distribution of residuals					
Min -0.0097	1st Qu -0.0023	Median -0.00032	Mean 0.000	3rd Qu 0.0022	Max 0.0114
Variables: lag12WPIF, lag10NG, lag3PC, lag5ERM, lag8PRS, lag9ERC, lag6WR, lag4PLS, lag9ELEC, lag12INL, lag4INL, lag11ERM, lag10CIC, lag11INL, WPIF, WPIM, lag6PGC, lag3M2, lag2WPIF, lag2CRUA, lag7PC, lag12PG, lag3ELEC, lag9NEER, lag12PLS					

The Above Table 4.59 shows that inflation is used as dependent variables whereas money supply (M2), real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM) their lags and the lags of dependent variable is used as regressors. From among all variables lag12 WPIF, lag10 of NG, lag3PC, lag5ERM, lag8PRS, lag9ERC, lag6WR, lag4PLS, lag9ELEC, lag12INL", lag4INL, lag11ERM, lag10CIC, lag11INL, WPIF, WPIM, lag6PGC, lag3M2, lag2WPIF, lag2CRUA, lag7PC, lag12PG, lag3ELEC, lag9NEER and lag12PLS are the important variable and is used for constructing the tree. Here 27 terminal nodes are used and the residual mean deviance is 0.000015 which is used as mean square error.

4.9.1.1 Cross validation

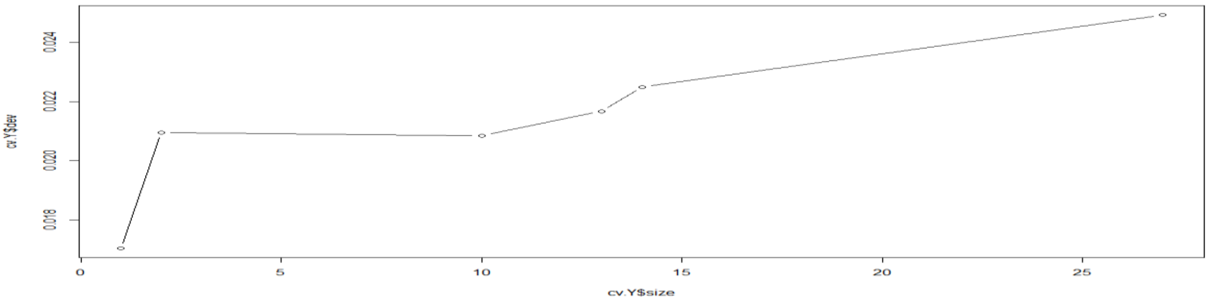

Figure 4 36: Cross Validation

Cross validation Figure 4.37 is used for the size of tree against the deviance. Here the minimum deviance is 2.

Figure 4 37: Regression Tree

The Above Plot 4.38 shows that when lag12 of WPIF is less than 0.0061 the predicted inflation is 0.0046 and lag12 of WPIF is greater than 0.0061 the predicted inflation is 0.0089.

b) On Test Data

Table 4.60: Regression Tree Result of Inflation on Test Data

Variables actually used in tree construction: "lag3PG" "lag3IMP"					
Number of terminal nodes		3			
Residual mean deviance		0.000014 = 0.0002091 / 15			
Distribution of residuals					
Min -0.0056	1st Qu -0.0023	Median -0.0002	Mean 0.000	3rd Qu 0.0014	Max 0.0079

The Above Table 4.60 shows that inflation is used as dependent variable whereas money supply (M2), real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM), their lags and the lags of dependent variable is used as regressors. From among all variables lag3 of PG and lag3 of IMP are the important variable and is used for constructing the tree. The three terminal nodes are used and the mean deviance is 0.000014 which is used as mean square error.

4.9.2 Bagging Using Inflation as Dependent Variable Information Set i=31

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 31.

Table 4.61: Bagging Result on Inflation on Train Data

Type of Random Forest	**Regression**
No. of trees	500
No. of variables tried at each split	371
Mean of square residual	0.000046
% variation explained	15.16

In this case we use inflation as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and Boosting .

b) On Test Data

Table 4.62: Bagging Result of Inflation on Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	371
Mean of square residual	0.000043
% variation explained	30.17

Again we apply bagging method on test data set results show that the mean square residual of test data set is less than train data set that indicates that there is no overfitting problem.

4.9.3 Random Forest Using Inflation as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 31.

Table 4.63: Random Forest Result of Inflation on Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	125
Mean of square residual	0.000046
% variation explained	16.26

In this case we use inflation as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and Boosting .

b) On Test Data

Table 4.64: Random Forest Result of Inflation on test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	125
Mean of square residual	0.000036
% variation explained	11.16

We apply random forest on test data set then results indicate that mean square of residual is 0.000036 which is less than train data set so this results indicates that random forest give better prediction other the ensemble method and there is no evidence of overfitting problem.

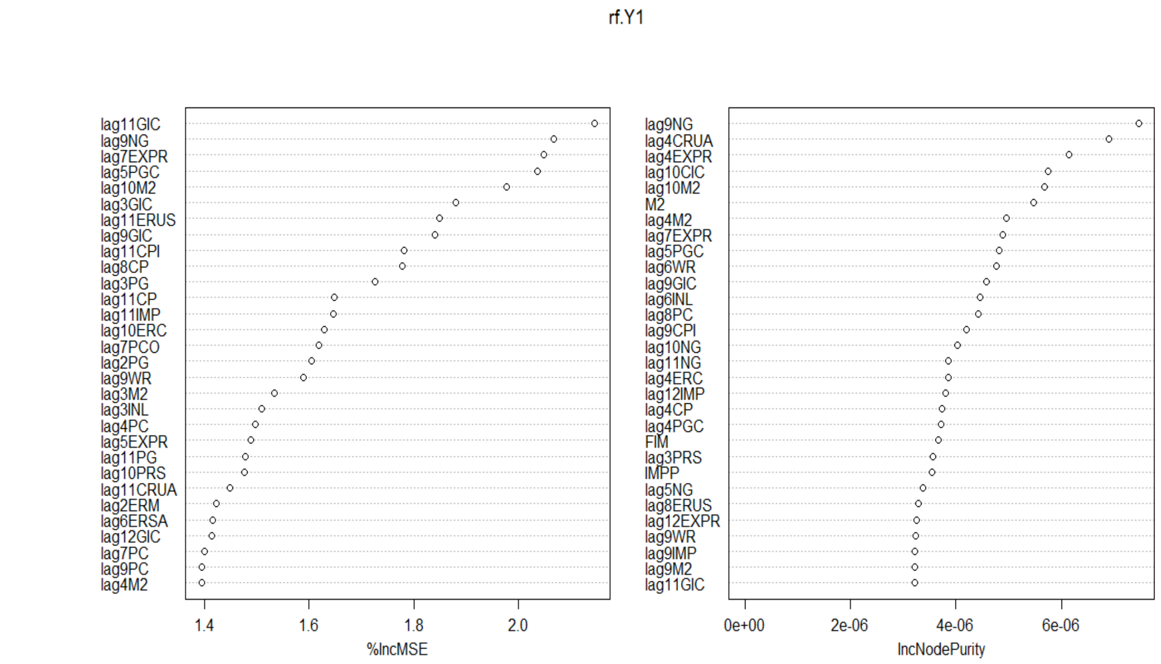

Figure 4.38: Important Variables

The Above Figure 4.38 shows the important of variables in random forest that how effect the node purity and MSE. The most important variables are lag11 of GIC, then lag9 of natural gas, then lag7 EXPR and so on. The least important variables are money supply, lag9 of production of cement and lag4 of money supply.

4.9.4 Summary of Boosted Regression Tree Using Inflation as Dependent Variable

The Above Table shows that inflation is used as dependent variable money supply (M2), real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with

SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM), their lags and the lags of dependent variable are used as regressors. A gradient boosted model, with gaussian loss function 5000 iteration were performed and there are 371 regressors are used of which 371 had non zero influence.

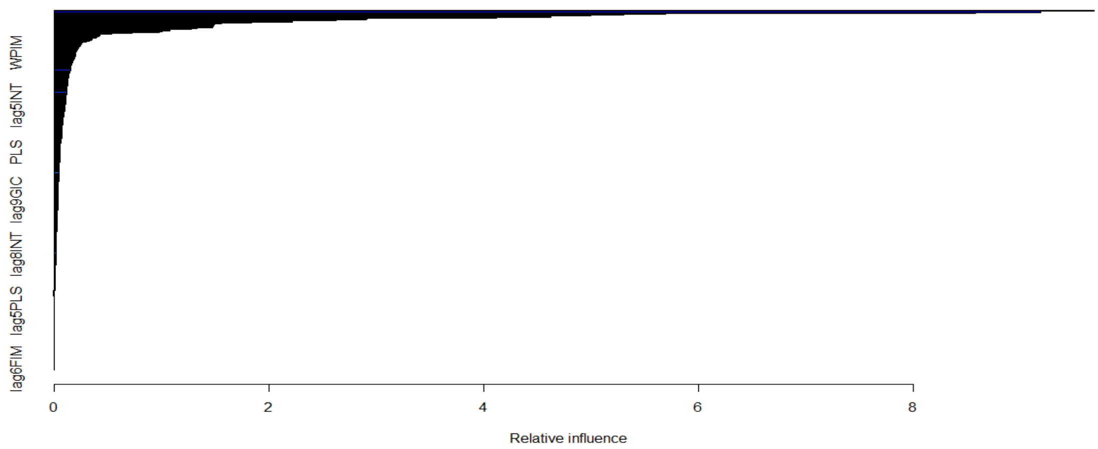

Figure 4.39: Boosted Regression Tree

The Above Figure 4.40 shows the relative influence of the variable. The highest bar shows that WPIF is the first significant variable, then the lag5 of INT is the second highest and significant variable and so on.

4.9.5 Using Money Supply as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 31.

Table 4.65: Regression Tree Result of Money Supply on Train Data

Number of terminal nodes	4
Residual mean deviance	0.1612 = 46.76 / 290
Distribution of residuals	

Min	1st Qu	Median	Mean	3rd Qu	Max
-5.074	-0.0119	-0.0019	0.000	0.0134	1.8610
Variables : lag11BSBP, lag2M2, lag5INT					

The Above Table 4.65 shows that money supply (M2) is used as dependent whereas inflation, real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM), their lags and the lags of dependent variable are used as regressors. From among all variables lag11 of BSBP, lag2 of M2 and lag5 of INT are the important variables and is used for constructing the tree. Four terminals nodes are used and the mean deviance is 0.1612 which is used for mean square error.

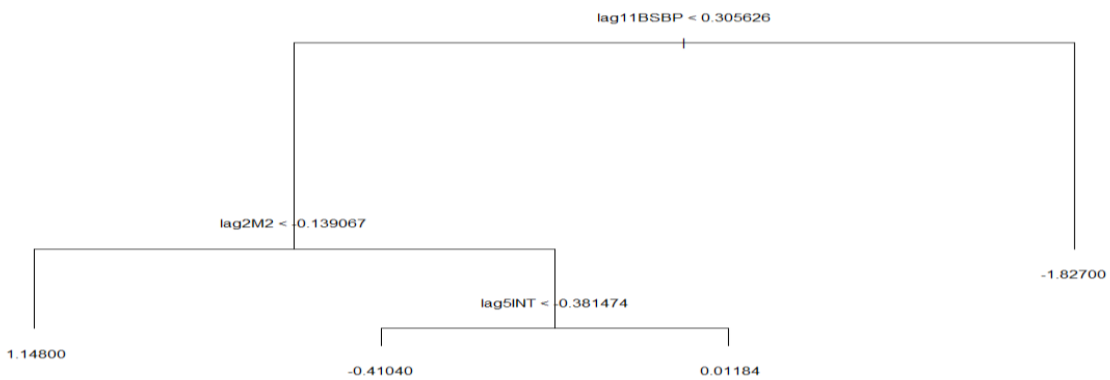

Figure 4.40: Regression Tree

The Above Figure 4.40 shows that when lag11 of BSBP is less than 0.306 and lag2 of M2 is less than -0.1391 the predicted money supply is 1.148 and lag2 of M2 is greater than -0.139 and lag5 of INT is less than 0.3815 the predicted money supply is -0.4104 and lag5 of INT is greater than 0.3815 the predicted money supply is 0.0118. On other hand when lag11 of BSBP the predicted money supply is -1.827.

b) On Test Data

Table 4.66: Regression Tree Result of Money Supply on Test Data

Number of terminal nodes			3		
Residual mean deviance			0.000065 = 0.0009732 / 15		
Distribution of residuals					
Min -0.0173	1st Qu -0.0029	Median 0.0002	Mean 0.000	3rd Qu 0.0028	Max 0.0218
Variables: lag3PG, lag5PC					

The Above Table 4.66 shows that that money supply (M2) is used as dependent variable whereas inflation, real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM), their lags and the lags of dependent variable are used as regressors. From among all variables lag3 of PG and lag5 of PC are the important variables and is used for constructing the tree. Three terminal nodes are used and the residual mean deviance is 0.000065 which is used as mean square error.

4.9.6 Bagging Using M2 as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 31.

Table 4.67: Bagging Result of Money Supply on Train Data

Type of random forest	Regression
No. of trees	500
No. of variables tried at each split	371
Mean of square residual	0.2489
% variation explained	2.5

In this case we use money supply as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and Boosting .

b) On Test Data

Table 4.68: Bagging Result of Money Supply in Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	125
Mean of square residual	0.00024
% variation explained	41.35

Again we apply bagging method on test data set results show that the mean square residual of test data set is less than train data set that indicates that there is no overfitting problem.

4.9.7 Random Forest Using M2 as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 31.

Table 4.69: Random Forest Result of Money Supply in Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	124
Mean of square residual	0.2486
% variation explained	2.35

In this case we use inflation as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and Boosting .

b) On Test Data

Table 4.70: Random Forest Result of Money Supply in Test Data

Type of random forest	Regression
No. of trees	500
No. of variables tried at each split	124
Mean of square residual	0.00028
% variation explained	31.85

We apply random forest on test data set then results indicate that mean square of residual is 0.000028 which is less than train data set so this results indicates that random forest give better prediction other the ensemble method and there is no evidence of overfitting problem.

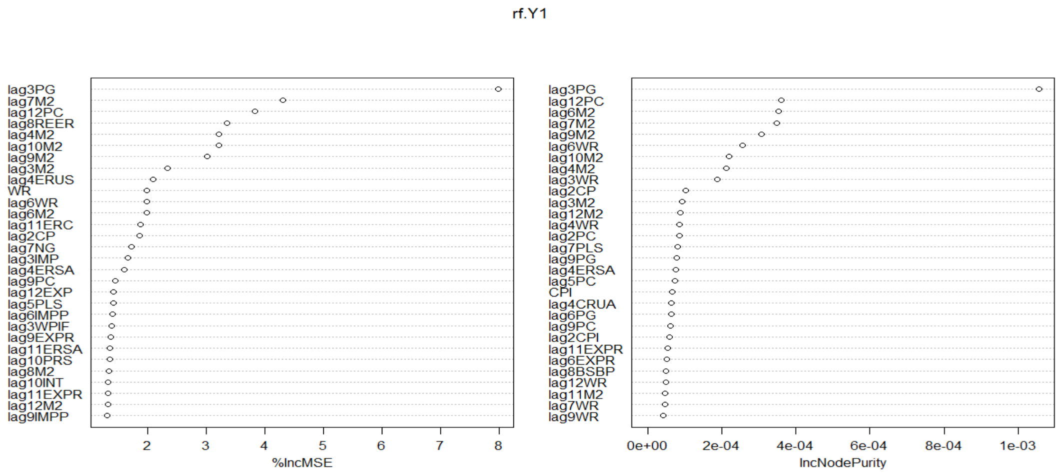

Figure 4.41: Important Variables

The Above Figure 4.42 shows the important of variables in random forest that how effect the node purity and MSE. The most important variable is lag3 of PG, then lag7 of M2, lag12 of PC and so on.

4.9.8 Summary of Boosted Regression Tree Using Money Supply as Dependent Variable

The Above Table shows that money supply (M2) is used as dependent variable whereas inflation, real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective exchange rate (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia

(ERM), their lags and the lags of dependent variable are used as regressors. A gradient boosted model, with gaussian loss function 5000 iteration were performed and there are 371 regressors are used of which 371 had non zero influence. The mean square error is 0.00002

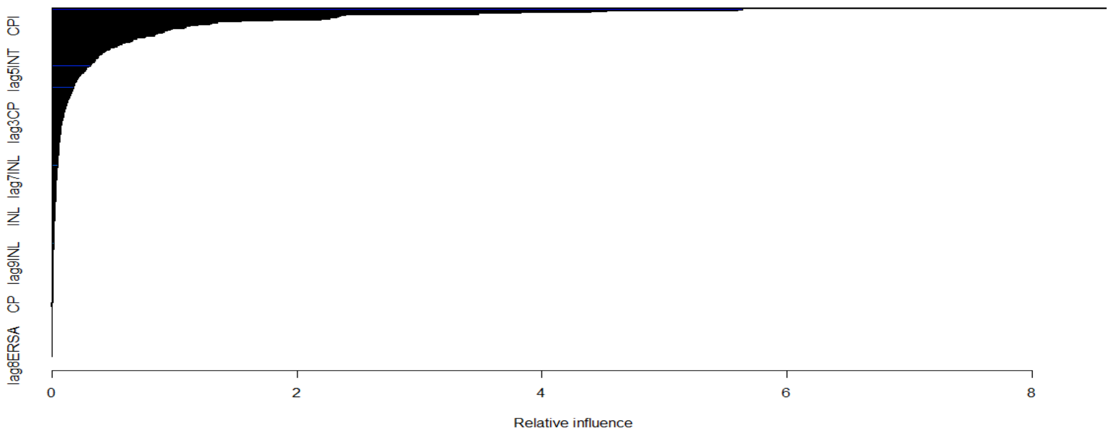

Figure 4.42: Boosted Regression Tree

The Above Figure 4.43 shows the relative influence of the variables. The first highest bar shows that inflation is the most significant and important variable. Second highest bar shows that lag5 INT is the second important and significant variable and so on. The mean square error is 0.00036.

4.9.9 Using REER is Used as Dependent Variable

Table 4.71: Random Forest Results of Real Effective Exchange Rate for Train Data

Number of terminal nodes			9		
Residual mean deviance			0.000033 = 0.009426 / 285		
Distribution of residuals					
Min -0.0185	1st Qu -0.0039	Median -0.00006	Mean 0.000	3rd Qu 0.0037	Max 0.0221
Variables: NEER, lag2CPI					

The Above Table 4.71 shows that REER is used as dependent variable whereas inflation, money supply (M2), currency in circulation (CIC), worker remittances

(WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM), their lags and the lags of dependent variable is used as regressors. From among all variables NEER and lag2 of CPI are important variables and is used for constructing the tree. Nine terminal nodes are used and the residual mean deviance is 0.000033 which is used as mean square error.

4.9.9.1 Cross Validation

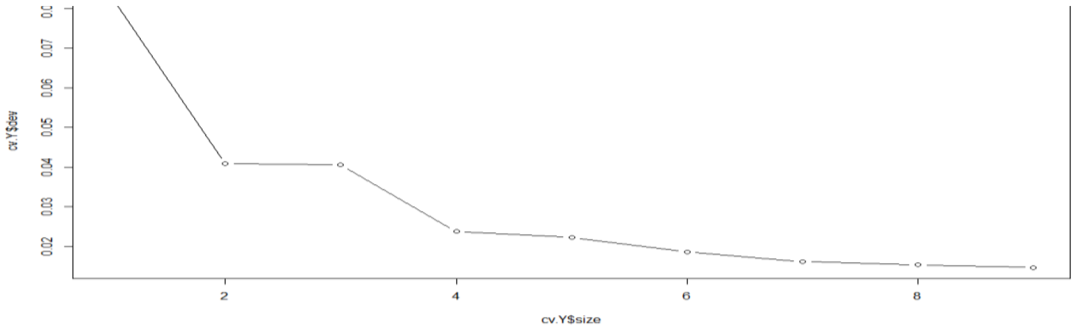

Figure 4.43: Cross Validation

The cross validation is the size of RT among the deviance and the minimum deviance is 9.

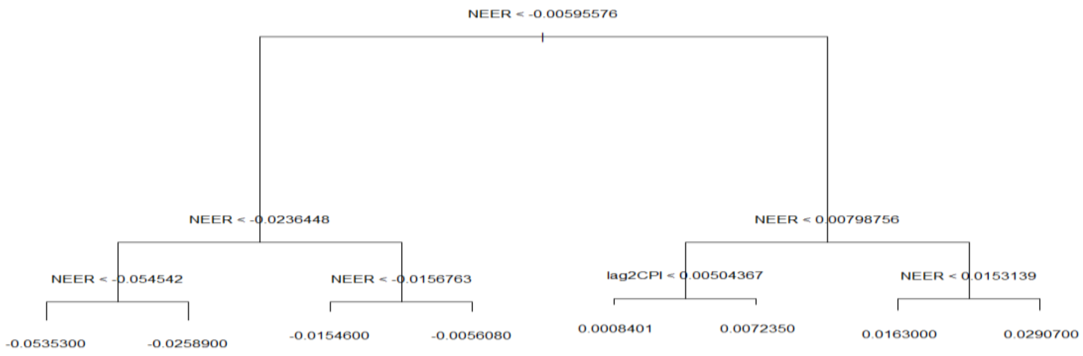

Figure 4.44: Regression Tree

The Above Plot 4.44 shows that when NEER is less than -0.0059 and NEER is less than -0.0236 and NEER is less than -0.0545 the predicted REER is -0.0535 and NEER is greater than -0.0545 the predicted REER is -0.0359 and NEER is greater than -0.0236and NEER is less than -0.01568 the predicted NEER is -0.01546 and NEER is greater than -0.0157 the predicted REER is -0.0056.

On the other hand when NEER is greater than -0.0059 and NEER is less than 0.0079 and lag2 of inflation is less than 0.0050 the predicted REER is 0.00084and lag2 of inflation is greater than 0.0050 the predicted REER is 0.0073. when NEER is greater than 0.0079 and NEER is less than 0.0153 the predicted REER is 0.0163 and NEER is greater than 0.0153 the predicted REER is 0.0291.

b) On Test Data

Table 4.72: Regression Tree Result of Regression Tree in Real Effective Exchange Rate

Number of terminal nodes			3		
Residual mean deviance			0.000007 = 9.844e-05 / 15		
Distribution of residuals					
Min -0.0048	1^{st} Qu -0.0014	Median -0.0006	Mean 0.000	3^{rd} Qu 0.0019	Max 0.0039
Variables: NEER, "lag6PRS					

The Above Table 4.74 shows that real effective exchange rate (REER) is used as dependent variable whereas inflation, money supply (M2), real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas

(NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM), their lags and the lags of dependent variable are used as regressors. From among all variables NEER and lag6 of PRS are important variables and is used for constructing the tree. Three terminals nodes are used and the mean deviance is 0.000007 which is used as mean square error.

4.9.10 Bagging Using REER as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 31.

Table 4.73: Bagging Result of Real Effective Exchange Rate in Train Data

Type of random forest	Regression
No. of trees	500
No. of variables tried at each split	371
Mean of square residual	0.000037
% variation explained	86.92

In this case we use REER as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and Boosting .

b) On Test Data

Table 4.74: Bagging Result of Real Effective Exchange Rate in Test Data

Type of random forest	Regression
No. of trees	500
No. of variables tried at each split	371
Mean of square residual	0.000025
% variation explained	21.39

Again we apply bagging method on test data set results show that the mean square residual of test data set is less than train data set that indicates that there is no overfitting problem.

4.9.11 Random Forest Using REER as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 31.

Table 4.75: Random Forest Result of Real Effective Exchange Rate in Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	125
Mean of square residual	0.000064
% variation explained	77.46

In this case we use inflation as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and Boosting.

b) On Test Data

Table 4.76: Random Forest Result of Real Effective Exchange Rate in Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	125
Mean of square residual	0.000082
% variation explained	19.3

We apply random forest on test data set then results indicate that mean square of residual is 0.000082 which is less than train data set so this results indicates that random forest give better prediction other the ensemble method and there is no evidence of overfitting problem.

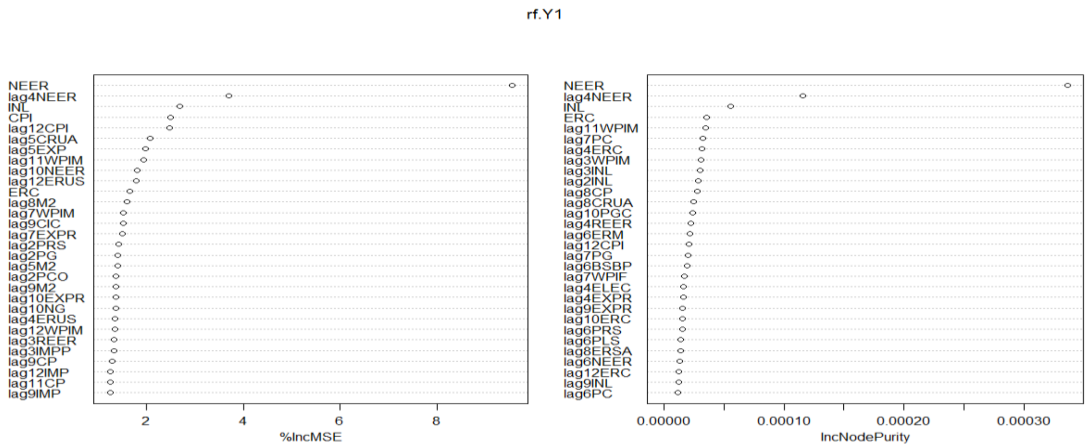

Figure 4.45: Important Variables

The Above Figure 4.45 shows the important of variables in random forest that how effect the node purity and MSE. The most important variable is NEER, then lag4 of NEER, INL and so on.

4.9.12 Summary of Boosted Regression Tree Using REER as Dependent Variable

The Above Table shows that real effective exchange rate (REER) is used as dependent variable whereas inflation, money supply (M2), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM), their lags and the lags of dependent variable are used as regressors. A gradient boosted model, with gaussian loss function 5000 iteration were performed and there are 371 regressors are used of which 371 had non zero influence. The mean square error is 0.00002

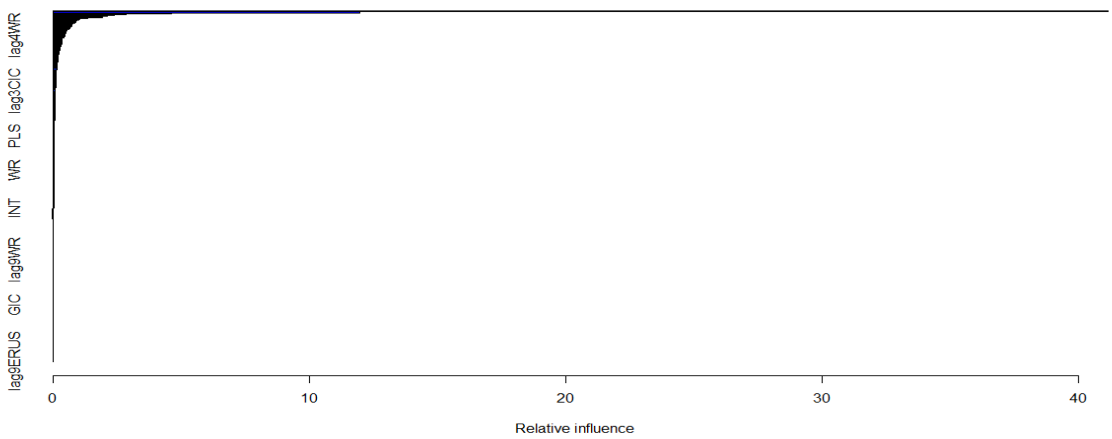

Figure 4.46: Cross Validation

The Above Figure 4.46 shows the relative influence of the variables. The first highest bar shows that lag4 WR is the most significant and important variable. Second highest bar shows that lag3 CIC is the second important and significant variable and so on. The mean square error is 0.000011.

4.9.13 Using Currency in Circulation as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 31.

Table 4.77: Regression Tree Result of Currency in Circulation in Train Data

Number of terminal nodes			6		
Residual mean deviance			0.07018 = 20.21 / 288		
Distribution of residuals					
Min -1.750	1st Qu -0.0233	Median -0.00175	Mean 0.000	3rd Qu 0.02190	Max 1.833
Variables: lag2CIC, lag3EXPR, lag6CP, lag3CP					

The Above Table 4.77 shows that currency in circulation is used as dependent variable whereas inflation, money supply (M2), real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM), their lags and the lags of dependent variable are used as regressors. From among all variables lag2 of CIC, lag3 of EXPR, lag 6 and 3of CP are the important variables and is used for constructing the tree. Here six terminal nodes are used and the mean deviance is 0.0702 which is used as mean square error.

4.9.16.1 Cross Validation

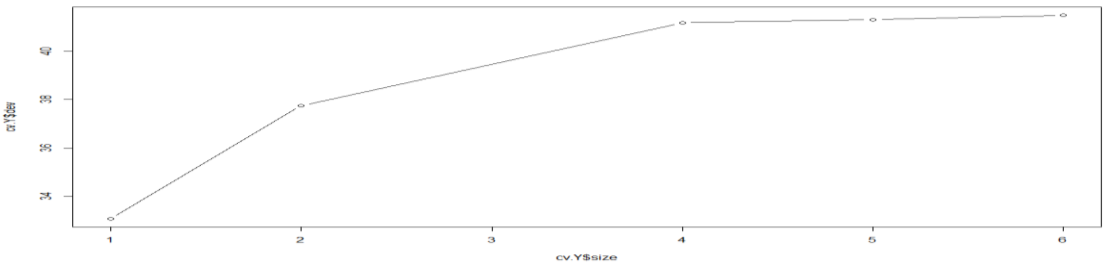

Figure 4.47: Cross Validation

Cross validation is used for the size of tree against the deviance and the minimum deviance is 1.

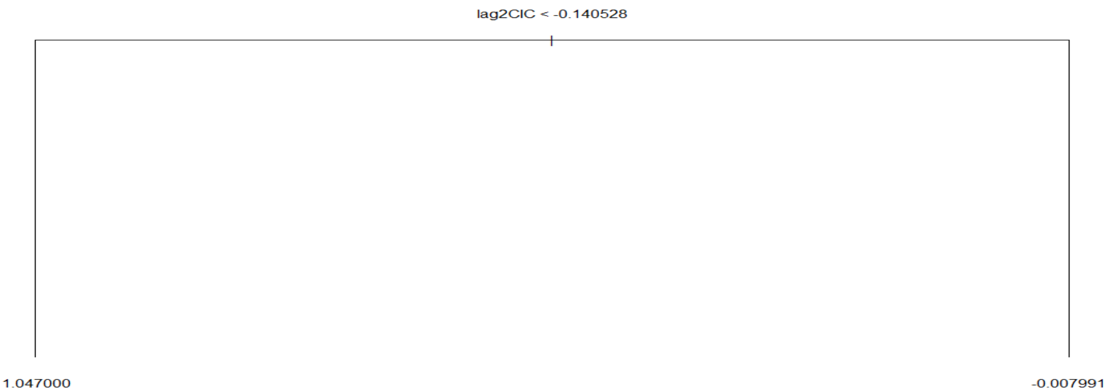

Figure 4.48: Regression Tree

The Above Figure 4.49 shows that when lag2 of CIC is less than -0.1405 the predicted currency in circulation is 1.047 and lag2 of CIC is greater than -0.1405 the predicted currency in circulation is -0.0079.

b) On Test Data

Table 4.78: Regression Tree Result Currency in Circulation in Test Data

Number of terminal nodes			3		
Residual mean deviance			0.00046 = 0.006828 / 15		
Distribution of residuals					
Min -0.0348	1st Qu -0.0081	Median 0.0009	Mean 0.000	3rd Qu 0.0105	Max 0.0514
Variables: lag12PGC, IMP					

The Above Table 4.78 shows that currency in circulation (CIC) is used as dependent variable whereas inflation, money supply (M2), real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM), their lags and the lags of dependent variable are used as regressors. From among all variables lag12 of PGC and IMP are important variables and are used for constructing the tree. Three terminal nodes are used and the mean deviance is 0.00046 which is used as mean square error.

4.9.14 Bagging Using Currency in Circulation as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 31.

Table 4.79: Bagging Result of Currency in Circulation in Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	371
Mean of square residual	0.1164
% variation explained	4.63

In this case we use currency in circulation as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and Boosting .

b) On Test Data

Table 4.80: Bagging Result of Currency in Circulation in Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	371
Mean of square residual	0.00085
% variation explained	11.87

Again we apply bagging method on test data set results show that the mean square residual of test data set is less than train data set that indicates that there is no overfitting problem.

4.9.15 Random Forest Using Currency in Circulation as Dependent Variable

a) On Train Data

The data used for train data set 1990 M1-2014 M2. In first step we use the information set i= 31.

Table 4.81: Random Forest Result of Currency in Circulation on Train Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	124
Mean of square residual	0.11.06
% variation explained	0.58

In this case we use inflation as a dependent variable when we use random forest on train data set then results indicates that the error of this model is less than other ensemble methods such as bagging and Boosting .

b)On Test Data

Table 4.82: Random Forest Result of Currency in Circulation in Test Data

Type of random forest	**Regression**
No. of trees	500
No. of variables tried at each split	124
Mean of square residual	0.0010
% variation explained	5.88

We apply random forest on test data set then results indicate that mean square of residual is 0.00002 which is less than train data set so this results indicates that random forest give better prediction other the ensemble method and there is no evidence of overfitting problem.

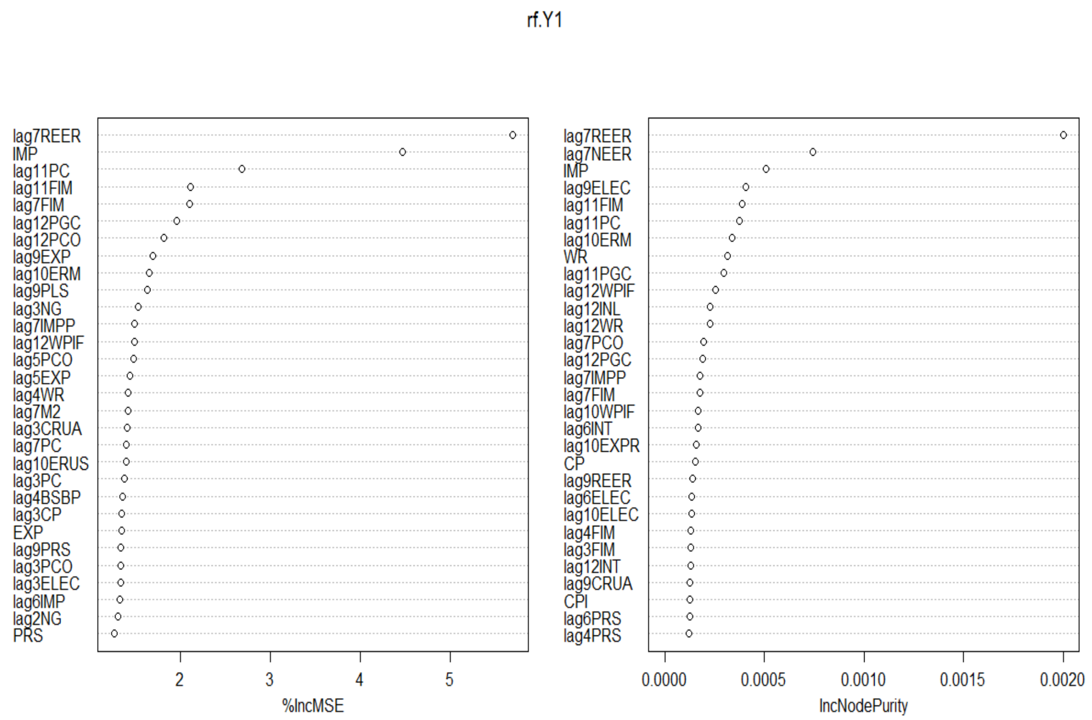

Figure 4 49: Important Variables

The Above Figure 4.49 shows the important of variables in random forest that how effect the node purity and MSE. The most important variable is lag7 REER, then IMP, lag11 of PC and so on.

4.9.16 Summary of Boosted RT Using Currency in Circulation as Dependent Variable

The Above Table shows that currency in circulation (CIC) is used as dependent variable whereas inflation, real effective exchange rate (REER), money supply (M2), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM), their lags and the lags of dependent variable are used as regressors. A gradient boosted model, with gaussian loss function 5000 iteration were performed and there are 371 regressors are used of which 371 had non zero influence. The mean square error is 0.00002.

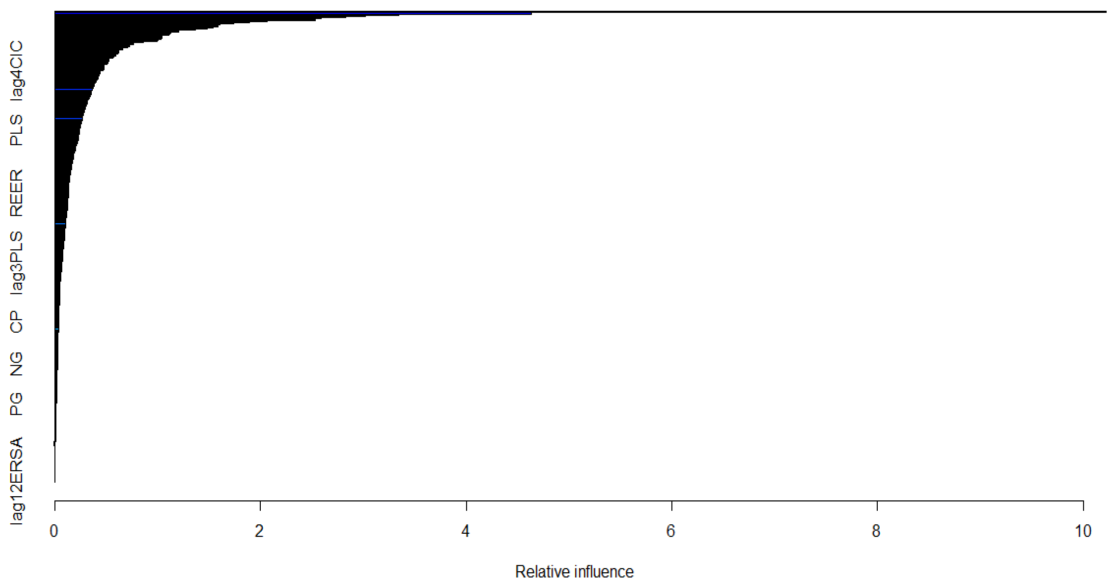

Figure 4 50: Boosted Regression Tree

The Above Figure 4.51 shows the relative influence of the variables. The first highest bar shows that lag4 CIC is the most significant and important variable. Second highest bar shows that PLS is the second important and significant variable and so on. The mean square error is 0.000042.

4.9.17 Adaptive LASSO Using Inflation as Dependent Variable

Table 4.83: Adaptive LASSO Result of Currency in Circulation

	Variable	Adaptive LASSO

Lag12 CPI	Lag12 of Inflation	0.9680

The Above Table 4.83 shows that inflation is used as dependent variable whereas money supply (M2), real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM), their lags and the lags of dependent variables used as regressors. From among all variables lag12 CIC is significant and important variable. The mean square error is 0.00000007.

4.9.18 Adaptive LASSO Using Money Supply (M2) as Dependent Variable

Table 4.84: Adaptive LASSO Result of Currency in Circulation

	Variable	**Adaptive LASSO**
Lag12 M2	Lag12 money supply	0.9649

The Above Table 4.84 shows that money supply (M2) is used as dependent variable whereas inflation, real effective exchange rate (REER), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM), their lags and the lags of dependent variables. From among all variables lag12 M2 is significant and important variable. The mean square error is 0.0000009.

4.9.19 Adaptive LASSO Using REER as Dependent Variable

Table 4 85: Adaptive LASSO Result of Currency in Circulation

	Variable	Adaptive LASSO
Lag12 REER	Lag12 REER	0.9680

The Above Table 4.85 shows that real effective exchange rate (REER) is used as dependent variable whereas inflation, money supply (M2), currency in circulation (CIC), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM), their lags and the lags of dependent variables. From among all variables only lag12 REER is significant and important variable. The mean square error is 0.0000003.

4.9.20 Adaptive LASSO Using Currency in Circulation as Dependent Variable

Table 4 86: Adaptive LASSO Result of Currency in Circulation

	Variables	**Adaptive LASSO**
Lag12 CIC	Lag12 currency in circulation	0.9648881

The Above Table 4.86 shows that currency in circulation (CIC) is used as dependent variable whereas inflation, real effective exchange rate (REER), money supply (M2), worker remittances (WR), natural gas (NG), electricity (ELEC), production of gypsum (PG), production of lime stone (PLS), production of rock salt (PRS), production of crude oil (PCO), production of good coal (PGC), production of cement (PC) , export of rice (EXR) , import (IMP), export (EXP), cash in Pakistan (CP), International Liquidity Gold Holdings National Valuation (INL), International Liquidity Total Reserves excluding Gold Foreign Exchange (INT), import of petroleum (IMP), balance with SBP (BSBP), Goods Value of Imports cost insurance freight(CIF), WPI food (WPIF), WPI manufacture (WPIM), financial interest rate (FIR), nominal effective XR (NEER), XR of china (EXC), XR of UAE (EXUA), XR of Saudi Arabia (ESA), XR of USA (ESUA) and XR of Malaysia (ERM), their lags and the lags of dependent variables. From among all variables only lag12 of lag12 CIC is important and significant variable. The mean square error is 0.000002.

4.9.21 Comparison of Mean Square Error of Forecasting in Multivariate Technique i=4, 14, 31

Table 4.87, 4.88 and 4.89 shows the forecast of four macroeconomic variables: CPI inflation, money supply, real effective exchange rate and currency in circulation. Each entry shows the mean square error (MSE), root mean square error (RMSE) and mean absolute error (MAE) of the forecast approach relative to the naïve benchmark. i present the number of variable in the information set. Our interest shows the value added to the forecasting key Pakistani macroeconomic variable by increasing the size of information set. Most entries in the table shows that the competition forecasting approach does not forecast the three targeted variables than the naïve approach. All competing method perform considerably better than the naïve benchmark in forecasting in REER. Interestingly the inflation and money supply also perform remarkably well for forecasting. In most case it seems that the best performance comes from using the smallest possible information set i=3.

Table 4.87: Comparison of All Methods

Mean Square Error (MSE)					
Models	I	Inflation	Money supply	REER	Currency in circulation
Regression Tree	4 14 31	0.000033 0.000031 0.000020	0.00010 0.00065 0.000075	0.000037 0.000035 0.00007	0.0072 0.0056 0.00044
Boosting	4 14 31	0.000033 0.000036 0.000030	0.000076 0.000054 0.000031	0.00043 0.00022 0.00015	0.0022 0.0014 0.0003
Bagging	4 14 31	0.000031 0.00038 0.000043	0.00026 0.00019 0.000025	0.000014 0.000014 0.000013	0.00085 0.000091 0.00085
Random Forest	4 14 31	0.000014 0.000011 0.000008	0.00026 0.00024 0.00001	0.000012 0.000013 0.0000082	0.0003 0.00021 0.0001
Adaptive LASSO	4 14 31	0.000020 0.000015 0.000013	0.00016 0.00019 0.00008	0.00015 0.00014 0.00008	0.0016 0.0048 0.00031
AR		0.000048	0.00014	0.0023	0.00026

Table 4.88 Comparison of All Methods

Root Mean Square Error (RMSE)					
Models	i	Inflation	Money supply	REER	Currency in circulation
Regression Tree	4	0.0054	0.01	0.0061	0.084
	14	0.0056	0.0254	0.0059	0.074
	31	0.0044	0.00866	0.0084	0.021
Boosting	4	0.0087	0.2903	0.0210	0.0446
	14	0.0086	0.2746	0.0209	0.0780
	31	0.0047	0.2741	0.0214	0.0462
Bagging	4	0.0056	0.016	0.0037	0.029
	14	0.0062	0.013	0.0037	0.0095
	31	0.0069	0.021	0.0035	0.029
Random Forest	4	0.0057	0.0047	0.0056	0.019
	14	0.0064	0.0021	0.0121	0.0213
	31	0.00027	0.00084	0.0004	0.0014
Adaptive LASSO	4	0.0057	----	0.0109	0.0816
	14	0.0063	0.0302	0.0111	0.0401
	31	0.00026	0.00099	0.00052	0.0012
AR		0.0062	0.4595	0.0107	0.0343

From among all techniques it shows that adaptive LASSO, Random forest and bagging get better performance. When the size of information set increases the forecast result perform good result.

Table 4.89 Comparison of All Methods

Mean Absolute Error (MAE)					
Models	I	Inflation	Money supply	REER	Currency in circulation
Regression Tree	4	0.0074	0.1351	0.0215	0.0346
	14	0.0067	0.1024	0.0159	0.0216
	31	0.0054	0.097	0.0123	0.0198
Boosting	4	0.0069	0.1254	0.0159	0.0320
	14	0.0067	0.1051	0.0158	0.5184
	31	0.0067	0.1079	0.0169	0.0340
Bagging	4	0.0066	0.121	0.0347	0.0401
	14	0.0063	0.0934	0.0234	0.0400
	31	0.0052	0.056	0.0178	0.0329
Random Forest	4	0.0034	0.0091	0.0045	0.0491
	14	0.0029	0.0065	0.0041	0.0317
	31	0.0004	0.00051	0.00030	0.00054
Adaptive LASSO	4	0.0045	---	0.0095	0.0594
	14	0.0052	0.0232	0.0099	0.0305
	31	0.00021	0.00076	0.00039	0.00090
AR		0.0053	0.0913	0.0095	0.0502

Table 4.90 Comparison of All Methods

Mean Absolute Percentage Error (MAPE)					
Models	I	Inflation	Money supply	REER	Currency in circulation
Regression Tree	4	0.74	13.51	2.15	3.46
	14	0.67	10.24	1.59	2.16
	31	0.54	9.7	1.23	1.98
Boosting	4	0.69	12.54	1.59	3.20
	14	0.67	10.51	1.58	51.84
	31	0.67	10.79	1.69	3.40
Bagging	4	0.66	12.1	3.47	4.01
	14	0.63	9.34	2.34	4.00
	31	0.52	5.6	1.78	3.29
Random Forest	4	0.34	0.91	0.45	4.91
	14	0.29	0.65	0.41	3.17
	31	0.04	0.051	0.030	0.054
Adaptive LASSO	4	0.45	---	0.95	5.94
	14	0.52	2.32	0.99	3.05
	31	0.021	0.07	0.039	0.090
AR		0.53	9.13	0.95	5.02

CHAPTER 5

CONCLUSION

In this study we compare the five machine learning technique to classify the important method for forecasting. In machine learning technique for multiple predictors that is Regression Tree, Bagging, Boosting, Random Forest, Adaptive LASSO and the benchmark method Auto regressive model. In univariate model the method is auto regressive model.

We had three set of information i-e, i = 4, 14 and 31. From among all the information set we conclude that when we have i= 4 variables the result showed that forecasting method is not doing well but we increase the information set from 4 to 14 the forecasting method become good performance. At the end of information set when the information set become 31 variable the result showed best forecasting from the multivariate techniques. The scope of benefit revealing from the consideration of range 20 to 31 variable for multiple predictors for forecasting of macroeconomic variables in case of Pakistan is deliberate from the result of this research. To classify the result

Also we conclude that from the comparison of all these techniques when we compare all the method Bagging, Random Forest, Adaptive LASSO and Boosting are the best technique for forecasting and for selection of important variables as compare to RT.

The prediction made through regression tree using complete set of observations i-e inflation, money supply, currency in circulation and real effective exchange rate are the dependent variables in all data set. In this study money supply result is more effective than currency in circulation. M2 is the dominant variable in all the estimation techniques. In regression Tree the tree become over fit in this case we using prune tree to remove the problem of over fitting in the data. In this case the result shows the importance of variable inflation and their lags 4, 5, M2 lag1, lag3 and currency in circulation are the important variables.

REFERENCES

Boivin, J., & Ng, S. (2006). Are more data always better for factor analysis?. Journal of Econometrics, 132(1), 169-194.

Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.

Carriero, A., Kapetanios, G., &Marcellino, M. (2009). Forecasting XRs with a large Bayesian VAR. International Journal of Forecasting, 25(2), 400-417.

De Mol, C., Giannone, D., &Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?. Journal of Econometrics, 146(2), 318-328.

Döpke, J., Fritsche, U., & Pierdzioch, C. (2015). Predicting recessions in Germany with boosted RTs (No. 5/2015). DEP (Socioeconomics) Discussion Papers, Macroeconomics and Finance Series.

Döpke, J., Fritsche, U., & Pierdzioch, C. (2017). Predicting recessions with boosted RTs. International Journal of Forecasting, 33(4), 745-759.

Döpke, J., Fritsche, U., & Pierdzioch, C. (2017). Predicting recessions with boosted RTs. International Journal of Forecasting, 33(4), 745-759.

Eickmeier, S., & Ng, T. (2009). Forecasting national activity using lots of international predictors: an application to New Zealand (No. 2009, 11). Deutsche Bundesbank.

ESTRUCH, V., FERRI, C., HERNANDEZ-ORALLO, J., & RAMIREZ-QUINTANA, M. J. (2004). Bagging decision multi-trees. Lecture notes in computer science, 41-51.

Exterkate, P., Groenen, P. J., Heij, C., & van Dijk, D. (2016). Nonlinear forecasting with many predictors using kernel ridge regression. International Journal of Forecasting, 32(3), 736-753.

Fattahi, S., Sohaili, K., Monkaresi, H., &Mehrabi, F. (2017). Modelling and Forecasting Recessions in Oil-Exporting Countries: The Case of Iran. International Journal of Economics and Financial Issues, 7(3), 569-574.

Hassani, H., & Silva, E. S. (2015). Forecasting with big data: A review. Annals of Data Science, 2(1), 5-19.

Huang, J., Ma, S., & Zhang, C. H. (2008). Adaptive Lasso for sparse high-dimensional regression models. Statistica Sinica, 1603-1618.

Hussain, A., Rahman, M., &Memon, J. A. (2016). Forecasting electricity consumption in Pakistan: The way forward. Energy Policy, 90, 73-80.

Inoue, A., & Kilian, L. (2008). How useful is bagging in forecasting economic time series? A case study of US consumer price inflation. Journal of the American Statistical Association, 103(482), 511-522.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R (Vol. 103). Springer Science & Business Media.

Jiang, B., Athanasopoulos, G., Hyndman, R. J., Panagiotelis, A., &Vahid, F. (2017). Macroeconomic forecasting for Australia using a large number of predictors. Monash Econometrics and Business Statistics Working Papers, 2, 17..

Kascha, C., &Trenkler, C. (2015). Forecasting VARs, model selection, and shrinkage (No. 15-07). Working Paper Series, Department of Economics, University of Mannheim.

Kim, H. H., & Swanson, N. R. (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. Journal of Econometrics, 178, 352-367.

Koop, G., & Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. International Economic Review, 53(3), 867-886.

Kuo, C., &Reitsch, A. (1995). Neural networks vs. conventional methods of forecasting. The Journal of Business Forecasting, 14(4), 17.

Li, J., & Chen, W. (2014). Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models. International Journal of Forecasting, 30(4), 996-1015.

Lockhart, r., taylor, j., tibshirani, r. J., & tibshirani, r. (2014). A significance test for the lasso1. The annals, 42(2), 413-468.

Pötscher, B. M., & Schneider, U. (2009). On the distribution of the adaptive LASSO estimator. Journal of Statistical Planning and Inference, 139(8), 2775-2790.

Pötscher, B. M., & Schneider, U. (2009). On the distribution of the adaptive LASSO estimator. *Journal of Statistical Planning and Inference*, *139*(8), 2775-2790.

Raviv, E., & van Dijk, D. (2014). Forecasting with many predictors: Allowing for non-linearity.

Schumacher, C. (2010). Factor forecasting using international targeted predictors: the case of German GDP. Economics Letters, 107(2), 95-98.

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. Journal of the American statistical association, 97(460), 1167-1179.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven- country data set. Journal of forecasting, 23(6), 405-430.

Stock, J. H., & Watson, M. W. (2006). Forecasting with many predictors. Handbook of economic forecasting, 1, 515-554.

Stock, J. H., & Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. Journal of Business & Economic Statistics, 30(4), 481-493.

Tahir, M. A. (2014). Analysing and Forecasting Output Gap and Inflation Using Bayesian Vector Autoregression (BVAR) Method: A Case of Pakistan. International Journal of Economics and Finance, 6(6), 257-266.

Zhang, H. H., & Lu, W. (2007). Adaptive Lasso for Cox's proportional hazards model. Biometrika, 94(3), 691-703.

Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American statistical association, 101(476), 1418-1429.