Global Journal of Science Frontier Research: F
Mathematics and Decision Sciences
Volume 18 Issue 4 Version 1.0 Year 2018
Type : Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

Can A Decision Tree Forecast Real Economic Growth from Relative Depth of Financial Sector in Nigeria?

By Alabi Nurudeen Olawale & Bada Olatunbosun

Abstract- We employed a decision tree statistical learning method which is lately gaining wide usage in the field of econometrics to establish the relationships between real gross domestic products growth rate and financial depth indicators such as stock market turnover ratio, credit to private sector (CPS) and broad money supply (M2) relative to gross domestic product (GDP) in Nigeria between 1981 to 2016. The data was divided into training and test datasets. The former was used to train the decision tree while the later was used to test the performance of the fitted decision tree model. Recursive binary splitting produced a fitted tree with nine nodes (leaves). This tree was pruned using cost complexity pruning procedure which uses a tuning parameter α to control the tradeoff between the tree complexity and overfitting the data. Pruning produced a tree with four terminal nodes and improved predictability in terms of lower model MSE on test dataset and interpretability. Bagging and Random Forest procedure were employed to further improve the performance of the model by aggregating bootstrapped training samples in order to reduce the variance.

Keywords: decision tree, recursive binary splitting, cost complexity pruning, bagging, random forest, financial depth, stock market liquidity.

GJSFR-F Classification: FOR Code: MSC 2010: 62P05

CANADECISIONTREEFORECASTREALECONOMICGROWTHFROMRELATIVEDEPTHOFFINANCIALSECTORINNIGERIA?

Strictly as per the compliance and regulations of:

Can A Decision Tree Forecast Real Economic Growth from Relative Depth of Financial Sector in Nigeria?

Alabi Nurudeen Olawale [α] & Bada Olatunbosun [σ]

Abstract- We employed a decision tree statistical learning method which is lately gaining wide usage in the field of econometrics to establish the relationships between real gross domestic products growth rate and financial depth indicators such as stock market turnover ratio, credit to private sector (CPS) and broad money supply (M2) relative to gross domestic product (GDP) in Nigeria between 1981 to 2016. The data was divided into training and test datasets. The former was used to train the decision tree while the later was used to test the performance of the fitted decision tree model. Recursive binary splitting produced a fitted tree with nine nodes (leaves). This tree was pruned using cost complexity pruning procedure which uses a tuning parameter α to control the tradeoff between the tree complexity and overfitting the data. Pruning produced a tree with four terminal nodes and improved predictability in terms of lower model MSE on test dataset and interpretability. Bagging and Random Forest procedure were employed to further improve the performance of the model by aggregating bootstrapped training samples in order to reduce the variance. These resulted in lower model MSEs on the test dataset. The regression tree model reveals that stock market turnover ratio and broad money supply relative to GDP are the most important financial depth measures in the real economic growth model. However, real GDP growth rate rises with stock market turnover ratio but dips with values of broad money supply relative to GDP between 10 per cent and 15 per cent. The real GDP growth rate stability threshold for stock market turnover ratio and broad money supply (M2) relative to gross domestic product are 10 per cent and 20 per cent respectively. R programming language was used throughout the paper.

Keywords: decision tree, recursive binary splitting, cost complexity pruning, bagging, random forest, financial depth, stock market liquidity.

I. Introduction

A decision tree is a supervised learning method with associated learning algorithms that analyze data used for classification and regression analysis. According to [1], amongst the learning methods in the Data Mining field, decision tree is considered to be closest to being "off-the-shelf" methods in terms of data handling, robustness to outliers in the regressor space, insensitivity to monotone transformations of inputs, computational scalability, and ability to deal with unimportant regressors. However, despite all these good characteristics possessed by decision trees, they perform poorly in their ability to extract linear combinations of the regressor space and in terms of predictive power. Off-the-shelf learning methods do not require time costly tuning of the learning procedure and preprocessing of data [1]. Hence these methods are applied directly on our real GDP growth rate data. In this current work, we study the relevance of this increasingly accepted modeling technique to the forecast of real economic growth

Author α: Department of Mathematics and Statistics, Federal Polytechnic Ilaro, Ogun State, Nigeria.
Author σ: Department of Statistics, Auchi Polytechnic, Auchi, Edo State, Nigeria. e-mail: bosunbada@auchipolyedu.ng

Ref

1. Hastie Trevor, Tibshirani Robert, Friedman Jerome (2008), *Elements of Statistical Learning, Data Mining, Inference and Prediction*, Second Edition, Springer, Stanford, California.

using three major measures of financial depth in Nigeria between 1981 to 2016. In addition, our interest in this learning method is directed at the interpretability of the relationships between the macroeconomic variables under study. These financial depth variables are functions of credit to private sector and broad money supply relative to gross domestic products (gdp) and Nigerian Stock market turnover ratio. Recursive Binary Splitting procedure was used to construct an unpruned tree. We then improve the interpretability and predictive power of this method by using the cost complexity pruning, bagging and random forest. Cost complexity pruning is quite efficient because it uses a nonnegative tuning parameter α to select a sequence of trees from a large number of subtrees resulting from splitting of the original tree. Bagging and Random Forest are better than cost complexity pruning in that they potentially increase the predictive power of our model. These concepts will be discussed later in the paper.

Financial development in Nigeria has witnessed series of reforms since 1986. Figure 1 shows five significant periods, which may be associated with periods the financial sector witnessed influx of institutions (including banking and nonbanking), changes in ownership and agency structure, depth and mode of operations. These structural reforms were envisioned to make the financial sector adapt to the fast changes in global financial system, more competitive, strong and reliable. Between 1986 and 2004, the Nigerian government through the Central Bank liberalized the banking sector by deregulating the sector, enhancing financial inclusion, influencing savings, investment and consumption through interest rates and credit control designed primarily for monetary and price stability. Furthermore, the state-owned banks were privatized. During this period, financial depth improved but far below the desired levels[1]. The banking crisis of 2004 and 2009 revealed weaknesses which brought about massive regulatory interventions and reinstatement of financial controls such as deleveraging of the banking sector, expansion of capital market, insurance market reform, creation of pension fund system, facilitation of bond issuance and establishment of long-term institutional investors. Institutional investors are pivotal to Nigeria financial system as their "*buy-and-hold*" strategies contribute in no small measure to low liquidity in the system[2]. The 2004 banking sector reform and consolidation led to significant rise in financial depth[3] with M2/GDP rising from 18.1 per cent of GDP in 2005 to 38 per cent in 2009, CPS/GDP rising from 12.6 per cent of GDP to 36.7 per cent of GDP, Stock market turnover ratio increasing by 0.68 per cent during the same period. A major increase in market liquidity was recorded in 2008 due to global financial crisis and Real economic growth rose marginally by 0.45 per cent during this period.

Notes

[1] M2/GDP, CPS/GDP, Stock market turnover ratio and real economic growth rate averaged geometrically 13.9 per cent, 9.01 per cent, 3.71 per cent and 3.99 per cent respectively.

[2] Whilst the total asset of pension funds accounted for about 5 per cent of GDP, the insurance companies accounted for less than 1 per cent of GDP

[3] Stock market has grown by more than 16 per cent of GDP since 2000, Liquidity rose marginally by 0.74 per cent and M2 averaged geometrically 20.81 per cent of GDP since 2000. Highest M2/GDP and CPS/GDP were recorded in 2009. In 2014, outstanding credit to private sector accounted for 22.4 per cent of GDP with emerging markets averaging over 50 per cent in 2014.

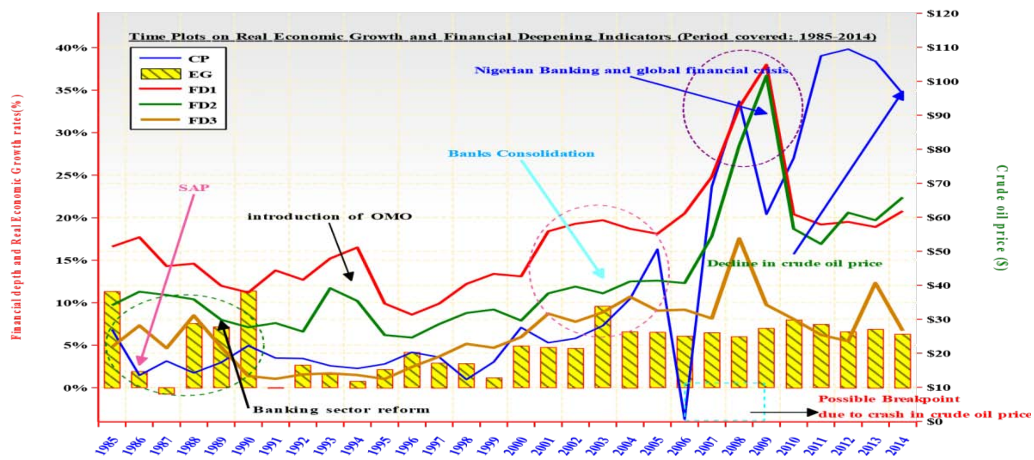

Figure 1: Time plot of Real Economic Growth rate (reg) versus M2/GDP (fd_1), CPS/GDP (fd_2) and Stock Market Turnover ratio (fd_3). This plot shows the five significant periods associated with periods the financial sector witnessed influx of institutions (including banking and nonbanking), changes in ownership and agency structure, depth and mode of operations

The banking sector remained dominant by the end of 2014, 21 major banks remain standing out of 89 major banks during pre-consolidation era[4]. Financial development has become an increasingly attractive topic to researchers globally in the last decade especially with the occurrence of the 2008 global financial crisis. The reasons for this recent interest in the topic stems from the believe that a well-structured financial development provides resilience and ensures economic growth needed to avert the effects of such crisis. By definition, financial development is a combination of depth, access and efficiency of financial system which provides opportunity for savings mobilization required for investment purposes, promotion of efficient information propagation, improvement of resource allocation and management of risks. Depth refers to size and liquidity of financial markets, access involves the ability of individuals to receive financial services and efficiency refers to ability of institutions to provide these services at reduced cost with persistent income and the level of activity in the capital market [2]. Depth of financial market gives rise to financial deepening, which may imply excessive financial development influenced by worsen regulatory framework. Although financial development enhances resilience and economic growth, concession between growth and stability due to financial deepening can emerge. A crucial question is "what are the limits to which financial development boosts long-term positive economic growth or limits of financial depth sustainable for enhancement of long-term positive economic growth in advanced and emerging economies?" [2] came up with different thresholds for stability and growth driven financial development for both the advanced and emerging economies. They based their work on newly developed financial development index, which combines the financial institution and financial market developments characteristics [2].

Prior to their work, a rich and diverse literature on the relationships between financial development and economic growth exists for both the advanced and emerging markets. The list is quite exhaustive but the earliest works were carried out by [3]; [4];

[4] Total assets of major banks accounted for more than 27 per cent of GDP in 2014. Branches have grown by more than 300 per cent since 1985, opening up access to well-developed branch networks, more than 7 branches per 100,000 adults and nearly 20 ATM per 100,000 adults.

Ref

3. King Robert, and Ross Levine. (1993), "Finance and Growth: Schumpeter Might Be Right." *The Quarterly Journal of Economics*, 108 (3): 717–37.

[5]. They applied various econometric techniques such as causality analysis amongst financial depth indicators like size of banking system relative to gross domestic products (GDP), stock market depth, and instrumental variables. They concluded that the past and current sizes of banking system relative to GDP could predict future values of economic growth by establishing unidirectional causality from these financial depth indicators to economic growth. [6], through Johansen vector error correction model (VECM) studied the long run causal relationships between financial depth indicators such as value added ratio and ratio of total private credit extension to GDP. They concluded that financial depth has significant effects on economic growth and credit rationing is prevalent in South Africa with firms with extensive reliance on internally generated income for operational requirements. Other works include [7] who employed panel data analysis in which financial depth indicators were instruments and controlling for other determinants of growth. [8] through correlation analysis found strong positive correlation between financial depth and real economic activity such as economic growth. Their work emphasized that countries with solid and well-established regulatory framework suffer less from shocks on inflation than countries with faulty regulatory structure. [9] established the existence of threshold cointegration between financial depth and economic growth in Taiwan and pointed out that there exists a positive and significant financial impact on economic growth. They concluded that financial depth could increase economic growth in Taiwan. Another study by [10] tried answering the question of how important financial development is to economic growth through a costly state verification model. Their resolve was that financial intermediation is crucial to economic development in the United States and a cross-country data. Furthermore, emphasis was that 29 per cent of US economic growth could be associated with technological improvement in financial intermediation. [11] analyzed the impact of financial deepening on economic growth in Turkey and found a strong negative relationship between financial deepening and economic growth. They summed up that financial development does not always lead to positive economic growth. In Nigeria, numerous monetary aggregates are calculated relative to GDP and used to measure the level of financial development/innovations. However, this current work focuses on three widely reported indicators; stock market liquidity/turnover ratio, M2/GDP and CPS/GDP. By the end of 2014, real economic growth in Nigeria stood at 6.30 per cent, broad money supply (M2) was 20.78 per cent of the GDP. Similarly, credit to private sector (CPS) relative to GDP stood at 22.40 per cent. However, the annualized economic growth rate in the last decade was 6.69 which made growth rather normally distributed. The country's financial development has gone through various phases of evolutions. Over time, several research works have emerged looking in-depth at the impact of financial deepening and economic growth in Nigeria. Prominent is the work of [12] which examined the relationship between real economic growth and three measures of development; M2/GDP, real interest rate and CPS/GDP. The study concluded that out of the financial depth indicators, only real interest rate is negatively related to real economic growth. The postulated model was statistically insignificant despite a high coefficient of determination. By implication the research failed to determine the data generating process (*d.g.p*, henceforth) underlying the time series, which is necessary to establish the order of integration of the series. Failure to do this may indicate that the estimated coefficients are mere contemporaneous correlations. Nevertheless, the study is important to this current work as it emphasized that financial development does not always result in positive economic growth in Nigeria, which is in line with conclusions made in previous studies by [11] in Turkey. [13]; [14] conducted Johansen test of

Ref

7. Beck, Thorsten and Ross Levine (2004), "Stock Markets, Banks and Growth: Panel Evidence." *Journal of Banking and Finance* 28 (3): 423–42.

cointegration and time varying granger causality respectively using data between 1960-2012 on poverty level, financial development and economic growth. The former found no evidence of long run co-movement between M2/GDP, per capita consumption and real GDP per capita. However, the later by introducing structural break in the granger causality model of M2/GDP and real economic growth between 1961-2012 found predictive powers from financial depth to real economic growth and vice versa during different time periods. Another research by [15] concluded that CPS is insignificant in promoting higher economic growth in Nigeria. This paper focuses on the effects of three major financial depth indicators; M2/GDP, Stock market liquidity and CPS/GDP on real economic growth in Nigeria. This will be achieved by recursive binary splitting, cost complexity pruning, bagging and random forest. The data employed are yearly observations from 1985 to 2014 of real economic growth (**reg**), broad money supply (M2) relative to GDP ($\mathbf{fd}_1$), credit to private sector (CPS) relative to GDP ($\mathbf{fd}_2$) and stock market turnover ratio ($\mathbf{fd}_3$). The source of all time-series data is the Central Bank of Nigeria (CBN) database.

Ref

15. Oriavwote Victor E., Eshenake Samuel J., (2014), "An empirical assessment of financial sector development and economic growth in Nigeria", *International Review of Management and Business Research*, Vol. 3 (1).

II. Methods, Empirical Analyses and Results

The four macroeconomic variables included in this paper are real GDP growth rate (**reg**), M2/GDP ($\mathbf{fd}_1$), CPS/GDP ($\mathbf{fd}_2$) as proxies for financial deepening and turnover ratio as proxy for stock market liquidity ($\mathbf{fd}_3$). We divided the quarterly data between 1981 to 2016 into two. One part (training dataset) was used to grow our decision tree and the second half (test dataset) to test the performance of the decision tree model. As earlier mentioned in this paper, decision trees either for classification or regression is considered to be "*off-the-shelf*" statistical learning method which presently is gaining acceptance in the field of econometrics. This method derive motivation from a tree analogy grown with many branches and leaves. The leaves are referred to as the *terminal nodes* and the points along which the regressor space is split are called the *internal nodes*. The internal and terminal nodes are linked by *branches*. We draw our inspiration from the procedure outlined by [1] which involves dividing the regressor space $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3]$ where $\boldsymbol{x}_1=\mathbf{fd}_1$, $\boldsymbol{x}_2=\mathbf{fd}_2$, $\boldsymbol{x}_3=\mathbf{fd}_3$, into j distinct and disjoint regions $R_1, R_2, \ldots R_j$. The prediction is then done by determining the region R_j in which an observation falls into and using the mean of the response values of the training observations in that region R_j as the predicted value for that observation. The regions are determined such that the residual sum of squares given in equation 1.0 is minimized:

$$RSS = \sum_{j=1}^{J} \sum_{i \in Rj} (y_i - \bar{y}_{Rj})^2 \qquad 1.0$$

Where $\bar{y}_{Rj}$ is the mean response (real GDP growth rate, **reg**) for training observations in the j^{th} region. Generally speaking, it may be numerically infeasible to generate every partition of the regression space into J regions.

a) Recursive Binary Splitting on the fitted Real GDP Regression tree model

We adopted the *recursive binary splitting* (RBS, hereon) which start with single region at the top of the tree and gradually perform splitting in an optimal fashion at each step of the tree construction. The RBS procedure select a regressor $\boldsymbol{x}_j$ in $\boldsymbol{X}$ and a cutoff c such that the regressor space is split into two regions.

$$R_1 = \{X \mid x_j < c\} \text{ and } R_2 = \{X \mid x_j \geq c\}$$

With the ultimate aim of reducing the RSS in equation 1.0. this implies that we try to generate the value of *j* and *c*, which minimize equation 1.1

$$\sum_{i:x_i \in R_1(j,c)} (y_i - \bar{y}_{R_1})^2 + \sum_{i:x_i \in R_2(j,c)} (y_i - \bar{y}_{R_2})^2 \qquad 1.1$$

Where $\bar{y}_{R_1}$ is the mean response (real GDP growth rate, reg) for training observations in the $R_1(j, c)$ and $\bar{y}_{R_2}$ is the mean response for training observations in the $R_2(j, c)$ regions. This process was repeated in the subsequent steps minimizing RSS in each step. This resulted in the tree in Figure 2 with the value of *j* and *c* that minimizes the RSS in equation 1.1 are 3 and 7.585 per cent respectively. That is stock market turnover ratio (fd_3) is the regressor at the top of the tree used for the initial split such that

$$R_1 = \{X \mid fd_3 < 7.585\} \text{ and } R_2 = \{X \mid fd_3 \geq 7.585\}$$

The value of RSS = 362.9, MSE = 5.76 and the number of terminal nodes = 9. The splitting was terminated as soon as we have not more than 5 observations in each region Table 1. The unpruned tree used up about 88 per cent of the training observations.

Table 1: Analysis of Recursive Binary Splitting on Real GDP Growth Regression Tree Model

Node	Split	Number of observation	RSS	Predicted Response
1	root	72	1472	3.875
2	$fd_3 < 7.585$	47	877.3	2.189
4	$fd_2 < 8.770$	21	253.9	4.359
8	$fd_1 < 12.420$	12	166.9	5.381
16	$fd_1 < 10.795$	6	1.583	3.0380*
17	$fd_1 > 10.795$	6	99.480	7.7230*
9	$fd_1 > 12.420$	9	57.760	2.9970*
5	$fd_2 < 8.770$	26	444.6	0.4358
10	$fd_1 < 17.290$	17	225.5	-1.5580
20	$fd_3 < 5.380$	12	10.120	0.3708*
21	$fd_3 > 5.380$	5	63.560	6.1880*
11	$fd_1 > 17.290$	9	23.82	4.2020*
3	$fd_3 > 7.585$	25	209.6	7.0460
6	$fd_3 < 10.875$	15	89.940	8.3050
12	$fd_1 < 18.560$	5	3.439	5.892*
13	$fd_1 > 18.560$	10	42.840	9.511*
7	$fd_3 > 10.875$	10	60.280	5.159*

Terminal node (leave)* *Source: Computations using **R language*

Notes

Notes

Recursive Binary Splitting of Regression Tree on Real GDP growth rate and Financial Depth Regression Space

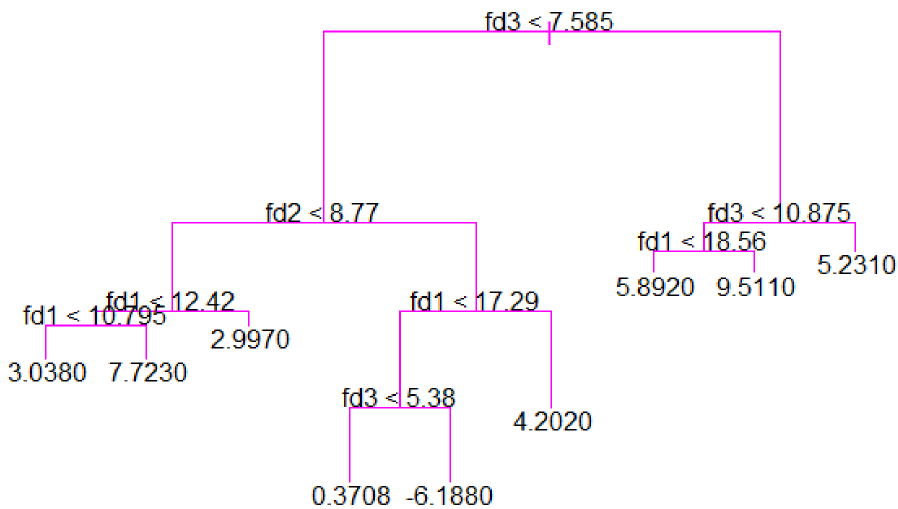

Figure 2: Unpruned Regression tree with 9 terminal nodes (leaves) and 8 internal nodes. This regression tree model shows that $\mathbf{fd}_3$ which is the measure of stock market liquidity is the most important regressor in the model since it minimizes the RSS in equation 1.0. Splitting was done in such a way that at each step, one of the regressor is selected. At a given internal node, the left-hand branch is represented by $X_j < q_k$ resulting from the split and $X_j > q_k$ indicates the right hand branch. At the top of the tree, the split resulted in two branches in which the left hand branch corresponds to $\mathbf{fd}_3 < 7.585$ per cent and the right hand branch corresponds to $\mathbf{fd}_3 \geq 7.585$ per cent. Splitting ensures simplicity and ease of interpretability of the regression tree model. **Source**: Computations using **R** language.

This procedure of RBS resulting in **Figure** 2 is quite simple which may have produced good predictions on the training dataset but eventually overfitting the real GDP growth and financial depth data. One snag is that this might lead to a very poor performance of the regression tree model on the test dataset. Consequently, in order to achieve a good test performance, we generated a smaller tree that contain fewer splits using *cost complexity pruning*. With this method, we are able to achieve a regression tree model with lower variance but slightly higher bias. The cost complexity pruning involves a nonnegative tuning parameter α such that for every value of this positive tuning parameter, there exists a subtree $T \subset T_0$ for which equation 1.2 is as small as possible.

$$\sum_{m=1}^{|T|} \sum_{i:x_i \in R_m} (y_i - \bar{y}_{R_m})^2 + \alpha |T| \qquad — \qquad 1.2$$

The quantity —T— is the number of terminal nodes or leaves of the regression tree T, R_m is the regions relating to the m^{th} terminal node and $\bar{y}_{Rm}$ is the mean of the training dataset's real GDP growth rate (**reg**) corresponding to the region R_m. Increasing the value of α from zero in equation 1.2 ensures that the tree branches get pruned and controls the tradeoff between the complexity of the regression subtree and the goodness of fit. We employed the 10-fold cross validation to determine the value of the positive tuning parameter and a most complex tree. We computed a tuning parameter $\alpha = 151.829$ with an MSE value of 1302.169 using cross validation Table 2 shows the result of the cost complexity pruning for various values of α at each split. We set the value of the best number of terminal nodes to 3 which is equivalent to the number of regressors in the regressor space for the tree pruning. This pruned tree was generated from a large tree on the training dataset containing 72 observations and varying the nonnegative tuning parameter α in equation 1.2 which resulted in 7 subtrees.

Notes

Table 2: *Cost Complexity Pruning on Regression Tree Model*

Node	Split	Number of observation	RSS	Predicted Response
1	root	72	1472	3.875
2	$fd_3 < 7.585$	47	877.3	2.189
4	$fd_2 < 8.770$	21	253.9	4.359*
5	$fd_2 > 8.770$	26	444.6	0.4358
10	$fd_1 < 17.29$	17	225.5	-1.5580*
11	$fd_1 > 17.29$	9	23.82	4.2020*
3	$fd_3 < 7.585$	25	209.6	7.0460*

*Terminal node (leave) *Source: Computations using R language*

Cost Complexity Pruning of Regression Tree on Real GDP growth rate and Financial Depth Regression Space

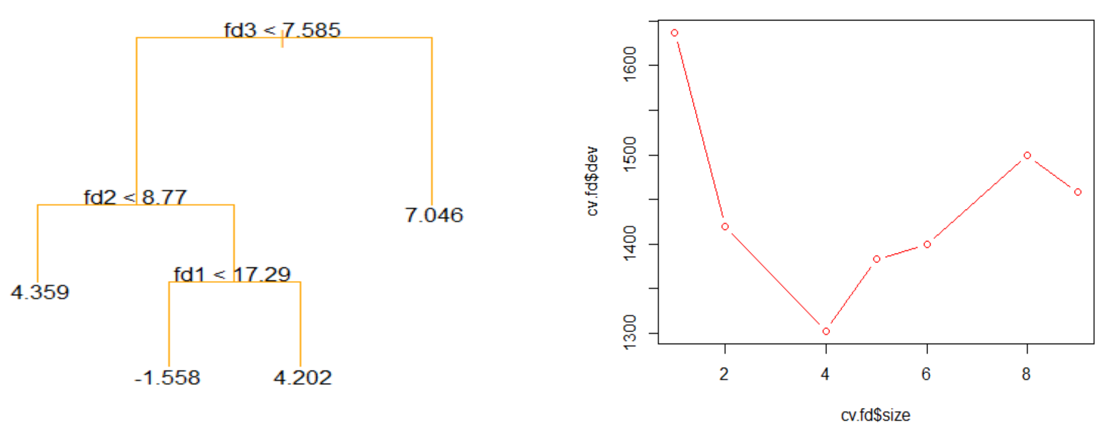

Figure 3: Analysis of cost complexity pruning on the fitted regression tree of Figure 2. Left panel: Pruned tree with 4 terminal nodes, 3 internal nodes. At a given internal the left-hand branch is represented by $X_j < q_k$ resulting from the split and $X_j > q_k$ indicates the right hand branch. At the top of the tree, the split resulted in two branches in which the left hand branch corresponds to $fd_3 < 7.585$ per cent and the right hand branch corresponds to $fd_3 \geq 7.585$ per cent. Right panel: The result of 10-fold cross validation showing the cross-validation error (cv.fd$dev) as a function of the terminal nodes (cv.fd$size). It indicates that the CV error dipped at terminal node value of 4 with MSE = 1,302.169. Source: Computations using R language

Notes

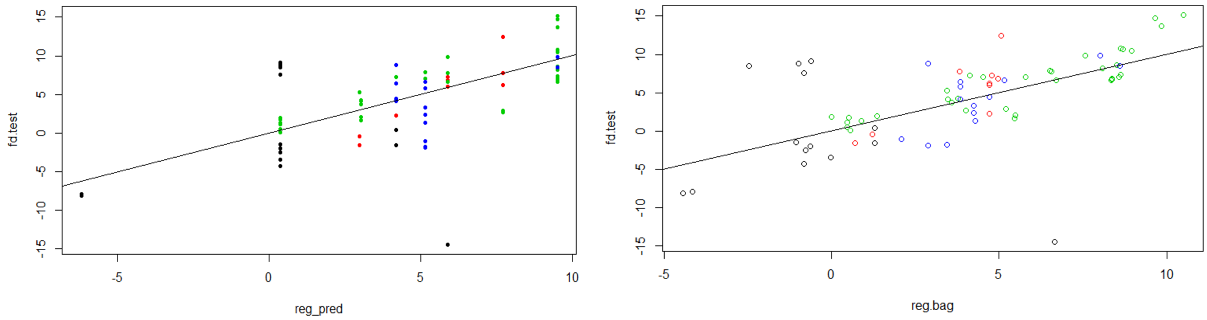

Figure 4: Left panel: The scatter plot on the residuals of the optimally-pruned regression tree using the test dataset. Right Panel: The scatter plot on the residuals of the aggregated bootstrapping on G = 200 regression trees using the test dataset (i.e. Bagging procedure). The straight line through the dotted points is the trend line. Source: Computation using R language

The pruned tree in Figure 3 reveals that out of the three regressors, stock market turnover ratio (fd_3) is the most important in predicting real economic growth. Using 25 training observations, given that fd_3 is greater or equal to 7.585 per cent, the mean of the observations on the real economic growth is 7.046 per cent which represent the first terminal node. The fitted regression was further split by credit to private sector (fd_2) given the stock market turnover ratio is less than 7.585 per cent. At this step, if fd_2 less than 8.77 per cent, a total of 21 training observations produced a mean response of 4.359 per cent. Otherwise another split was carried out which resulted in two terminal nodes. These two terminal nodes are -1.558 per cent if broad money supply relative to GDP (fd_1) is less than 17.29 per cent and 4.404 per cent otherwise. A total of 17 and 9 training observations were involved in calculating these two mean responses of the reg. using the result of this pruned tree, we predicted the real GDP growth rate for the 72 test observations to determine its performance. The mean square error (MSE) of the model using test dataset is 17.685 compared to the MSE of 34.785 using the train dataset.

b) Bagging and Random Forest on the fitted Real GDP Regression tree model

Decision trees are known to present some setbacks in terms of predictive accuracy (high variance) and lack of robustness to changes in data. Therefore, we employed bagging and random forests to improve the accuracy of the predictions generated by the regression tree model. Generally, bagging is a bootstrap aggregation procedure used in reducing the variance of a statistical learning method such as our fitted regression tree. It has been shown that averaging a set of observations reduces the variance. Bootstrapping involves taking single training dataset, generate $B = 400{,}000$ different bootstrapped regression trees and then trained our model on the $400{,}000^{th}$ bootstrapped regression tree. Using $G = 400{,}000$ subtrees, we computed

$$\hat{f}^{1}(x), \hat{f}^{2}(x)\ldots\ldots\ldots\ldots\ldots\hat{f}^{400,000}(x)$$

and an average prediction which gives

$$\hat{f}_{bag}(x) = \frac{1}{400{,}000}\sum_{g=1}^{400{,}000}\hat{f}^{*g}(x) \qquad 1.3$$

Because these trees are grown deep and unpruned, they possess higher variance but low bias. Hence averaging using equation 1.3 reduces the variance. The MSE of the bagging is 13.409 with 34.41 per cent explained variance. The performance of the regression tree model was verified using the test dataset. The MSE of the model using the test dataset is 18.292. In order to further reduce the test MSE of the regression model, we reduced the number of trees G to 200 trees, the training MSE becomes 13.410 and test MSE equals 17.653 which is slightly better than the optimally-pruned single tree. The right panel of Figure 4 shows residual plot on the bagging procedure. Bagging removes any simple structure in the regression tree model due to extremely large number of trees to be bootstrapped. Thus, our bagged tree on reg model stops to be a tree. Another problem with bagging is that it gives priority to strong and moderately strong regressors at the top split during the bootstrapping and aggregation process.

Subsequently, we employed another powerful method called the *Random Forest*. Random Forest is an improvement over the bagging method in the sense that it reduces the extent of correlation between the sampled trees. The procedure on Random Forest like bagging uses regression trees as foundation. We constructed a large number of trees on bootstrapped training samples using a random sample of m ¡ p predictors at every split in the regression tree. Specifically, we set $m = p/3$ and since $p = 3$ regressors in our regression tree model, we use $m = 1$ regressor at every split in the tree. This procedure ensures that no strong or moderately strong regressor is given priority over weaker ones by considering only a subset of the regressors which eliminates high correlations. by this method, we found that two-third of the splits might not involve the strong predictors thereby giving the other regressors a chance of being involved in the tree building process. A total of $G = 500$ trees was used for the bootstrapping process. This regression tree model produced a training dataset MSE value of 11.629 and 43.12 per cent explained variation. Furthermore, the test dataset MSE was 16.41 which is an improvement over bagged regression model.

Table 2: Analysis of regressor importance

Regressor	Accuracy (MSE)	Reduction in RSS
fd_2	9.180	425.052
fd_1	16.615	438.143
fd_3	10.347	450.300

Source: Computations using R language

Notes

Notes

Plots on measures of regressor importance

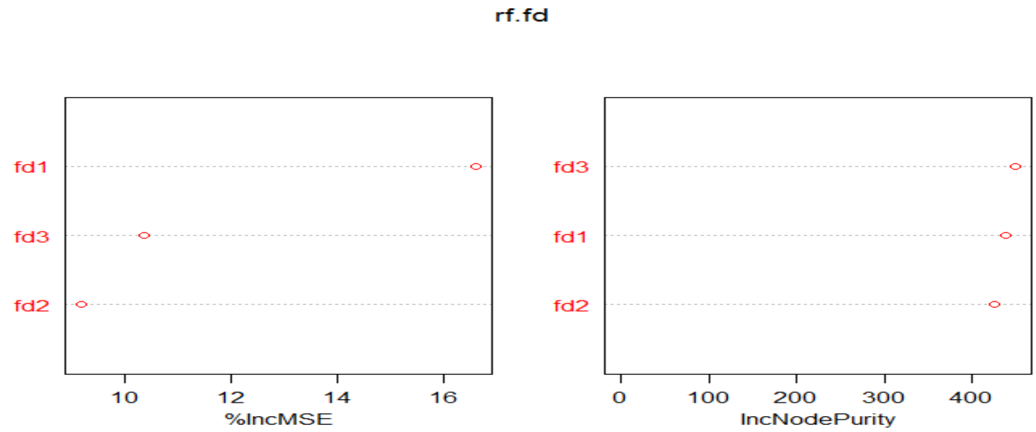

Figure 5: Left panel: Plot on the mean reduction in accuracy as measured by MSE in predictions on out of bag samples when a corresponding regressor is excluded from the model. In bagging, the bagged tree utilizes about two-third of observations. The remaining one-third of the observations not used up in the bagging procedure are referred to as out-of-bag (OOB) observations. These OOB observations are used to predict the response (reg) for the i^{th} observations using each of the trees in which this observation is an OOB. Right panel: Plot on the decrease in node impurity that results from splits over that regressor averaged over all trees. Node impurity is measured by the RSS. This plot indicates that on 500 trees, the RSS decreased by 450.300 if the split is done over fd_3 at the top of the tree. Source: Computation using R language

Partial Dependence Plots on Stock Market Turnover Ratio and Broad Money Supply Relative to GDP

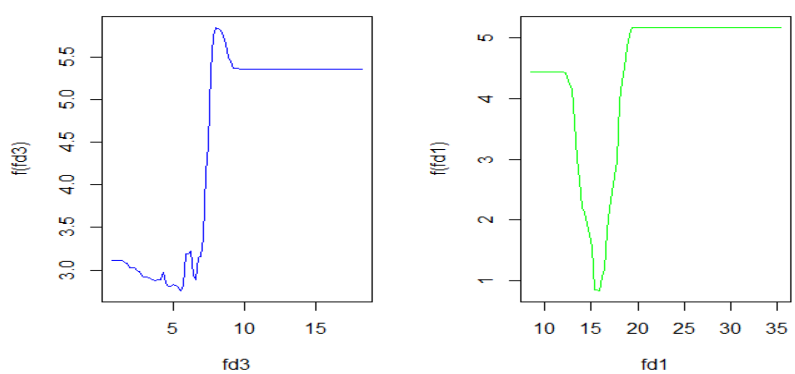

Figure 6: Plots showing the marginal effects of the fd_3 and fd_1 in the real GDP growth model. Left panel: This panel is the marginal effect of stock market turnover ratio (fd_3) on real GDP growth rate (reg). This plot shows that reg rises with increasing fd_3. reg rose sharply when fd_3 is about 6 per cent before stabilizing at values of fd_3 at 10 per cent or more. Right Panel: Marginal effect of the broad money supply relative to GDP (fd_1) on reg. Similar to the fd_3, this plot indicates a rising reg with increasing fd_1 and vice versa. Particularly, the reg dipped for values of fd_1 between 10 per cent and 15 per cent. The marginal effects of these regressors on reg were computed by integrating out other regressors. Source: Computations using R language

This procedure ensures that no strong or moderately strong regressor is given priority over weaker ones by considering only a subset of the regressors which eliminates high correlations. By this method, we found that two-third of the splits might not involve the strong predictors thereby giving the other regressors a chance of being involved in the tree building process. A total of $G = 500$ trees was used for the bootstrapping process. This regression tree model produced a training dataset MSE value of 11.629 and 43.12 per cent explained variation. Furthermore, the test dataset MSE was 16.41 which is an improvement over bagged regression model.

Notes

III. Conclusion

Decision tree for regression has been used to study the relationship between real gross domestic growth rate and three major financial depth indicators such as turnover ratio as proxy for stock market liquidity, credit to private sector (CPS) and broad money supply (M2) relative to gross domestic product (GDP) in Nigeria between 1981 to 2016. The analysis of the tree was done using the recursive binary splitting (RBS), cost complexity pruning, bagging and random forest. Recursive Binary Splitting produced a tree with nine leaves or terminal nodes. Though this single large tree produced a low MSE for the model on the training dataset, its performance on the test dataset was weak. Hence, we applied the cost complexity pruning (CCP) to trim down the number of terminal nodes (leaves) and improve the predictive capacity alongside the interpretability of our model. This model resulted in lower model MSE for the test dataset and 4 terminal nodes. Stock market turnover ratio minimizes the residual sum of squares (RSS). Therefore, this indicator was used as the initial splitting variable at the top of the tree. The CCP procedure was based on varied values of a tuning parameter which was used to control the tradeoff between the complexity of the model and it fitting the data adequately. We employed bagging and random forest to further improve the predictive capacity of the postulated regression tree model. Bagging and Random Forest are procedure that involve growing large number of separate trees which made plotting a tree impossible. However, the results of these two procedures indicate improved performance of the regression tree model on the test data set in terms of lower MSE. Partial dependence analysis of the regressors in terms of node purity indicates that the real GDP growth rate rises with stock market turnover ratio. It dipped for some values of broad money supply relative to GDP during the period under review. The fitted regression tree shows that stock market liquidity and broad money supply relative to GDP are the most important financial depth variables imparting the growth of real GDP in Nigeria during the period under review.

References Références Referencias

1. Hastie Trevor, Tibshirani Robert, Friedman Jerome (2008), *Elements of Statistical Learning, Data Mining, Inference and Prediction,* Second Edition, Springer, Stanford, California.
2. Ratna Sahay, Martin Čihák, Papa N'Diaye, Adolfo Barajas, Ran Bi, Diana Ayala, Yuan Gao, Annette Kyobe, Lam Nguyen, Christian Saborowski, Katsiaryna Svirydzenka, and Seyed Reza Yousefi (2015), *Rethinking Financial Deepening: Stability and Growth in Emerging Markets*, IMF Staff Discussion Note Number SDN/15/08.
3. King Robert, and Ross Levine. (1993), "Finance and Growth: Schumpeter Might Be Right." *The Quarterly Journal of Economics,* 108 (3): 717–37.

Notes

4. Levine, R. and Zervos, S., (1998), "Stock Markets, Banks and Economic Growth", *American Economic Review*, 88(3), pp. 537-58.
5. Levine, Ross, Norman Loayza, and Thorsten Beck. (2000), "Finance and the Sources of Growth." *Journal of Financial Economics* 58 (1/2): 261–300.
6. Kularatne Chandana (2001), *An examination of the impact of financial deepening on long-run economic growth: An application of a VECM structure to a middle-income country context*, Trade and Industrial Policy Strategies: 2001 Annual Forum.
7. Beck, Thorsten and Ross Levine (2004), "Stock Markets, Banks and Growth: Panel Evidence." *Journal of Banking and Finance* 28 (3): 423–42.
8. Berentsen Aleksander and Shi Shouyong (2008), *Financial Deepening Inflation and economic growth*, http://www.researchgate.net/publication/228424687_Financial_Deepening_Inflation_and_Economic_Growth. Retrieved on 20/06/2015.
9. Chang Shu-Chen, Wu Cheng-Hsien (2012), *The Relationship between Financial Deepening and Economic growth in Taiwan*, M. Zhu (Ed.): Business, Economics and Financial Sci., Manag., AISC 143, pp 205-210.
10. Greenwood Jeremy, Sanchez Juan M., Wang Cheng (2012), *Quantifying the Impact of Financial development on economic development", Review of Economic Dynamics*, http://dx.doi.org/10.1016/j.red.2012.07.003, Retrieved on 20/06/2015.
11. Ardic Oya Pinar, Damar H. Evren (2006), *Financial Sector Deepening and Economic Growth,* Online at http://mpra.ub.uni-muenchen.de/4077/ MPRA Paper No. 4077. Retrieved on 20/06/2015.
12. Adekunle Olusegun, Salami Ganiyu O., Adedipe Oluseyi (2013), "Impact of financial sector development on the Nigerian economic growth", *American Journal of Business and Management*, Vol 2(4): 347-356.
13. Aye Goodness C. (2013), "Causality between financial deepening, economic growth and poverty in Nigeria", *Business and Management Review*, Vol. 3 (3), pp 1-12.
14. Aye Goodness C (2015), "Causality between Financial Deepening and Economic Growth in Nigeria", *Journal of Economics, Business and Management*, Vol. 3 (8).
15. Oriavwote Victor E., Eshenake Samuel J., (2014), "An empirical assessment of financial sector development and economic growth in Nigeria", *International Review of Management and Business Research,* Vol. 3 (1).
16. *R* Core Team, *R*: A language and environment for statistical computing. *R* Foundation for statistical computing, (2018) Vienna, Austria. URL http://www.R-project.org/.