Methodology to Predict NPA in Indian Banking System

Gourav Vallabh¹, Digvijay Singh², Ruchir Prasoon², Abhimanyu Singh²

¹Finance Area, XLRI Xavier School of Management, Jamshedpur, India
²Business Management, XLRI Xavier School of Management, Jamshedpur, India
Email: gvallabh@xlri.ac.in

Received 28 July 2016; accepted 28 August 2016; published 31 August 2016

Abstract

Banking sector is the backbone of any economy and plays a pivotal role in the development of the nation. With the ever increasing need of involvement of banks in the economic growth process, the issue of non-Performing Assets (NPAs) has assumed mammoth proportions. Subsequently, as government in emerging countries pushes for financial inclusion, the risk associated with bank assets increases. With this background, this paper analyses problem of non-Performing Assets (NPAs) in Indian banking system. The authors have devised a unique way to forecast the NPAs in Indian banking system in 2020. The focus was to devise a model which would play a pivotal role in forecasting future NPAs in the Indian banking sector. This was achieved by looking into various methodologies and zeroing in on a model, which could be implemented to help understand how NPAs could be predicted.

Keywords
Non-Performing Assets, Multiple Linear Regression, Holt-Winter, Asset Restructuring Companies

1. Introduction

Banking sector is one of the most crucial aspects of an economy and it is high time the Indian banking sector starts to deliver the goods. Myriad schemes like Jan Dhan, Make in India, Direct benefit transfer require the Indian Banking sector to expand its spectrum of services and widen its capabilities to cater to larger masses. The necessity of it can be understood by the fact that India’s vision is to provide banking services to even its remotest geographical area and to accommodate people belonging to both ends of the spectrum. However, the road towards fulfilling the vision would not be easy as expansion of this scale attracts major roadblocks and challenges. The perennial problem of NPA will be the biggest issue to tackle as far as expansion is concerned.

Stringent RBI guidelines have added to the woes as it has resulted in greater stress for banks, since it is now asked banks to declare stressed assets. This has resulted in creation of provisions for bad loans, resulting in transfer of NPAs in the bad loans category. This has diminished profits and has created a negative sentiment for the banking sector.

The determination of NPA is very critical for a bank’s financial health and liquidity concerns. The health of the assets of a bank is measured by its NPA ratio. Given that the public banks in India are under a lot of financial stress due to accumulated non-performing asset, it is essential that they are in a position to plan for the future so that remedial steps can be taken.

It is crucial to understand what actually results in the rise of NPA before we delve into further details. Understanding the reasons for NPAs will provide a platform to better analyse the methodologies used to predict NPAs.

The causes for NPA can be attributed to many factors like bribery, corruption etc. One of the many causes is the lack of effort from the service provider during the pre-sanction process. Ignoring the process of conducting a due diligence process or conducting it in a haphazard manner leads to increasing cases of NPA. An excuse for corporations is blaming the economic scenario for defaulting; however forensic audits paint a different picture. In majority of the cases, borrowers have wilfully defaulted or transferred funds to other parties. The integrity of the promoter also comes into question in some high profile cases like that of Kingfisher. Obtaining tangible collateral security is also a big headache during credit enhancement as the loan amount outweighs the sales value of the attached security by a considerable margin, thereby diluting the provision of SARFAESI Act (Securitisation and Reconstruction of Financial Assets and Enforcement of Security Interest Act). The after disbursement monitoring process is also porous as in most of the cases there is no forensic audit after initial sanction of loan. There is a lack of due diligence for loan disbursement of the sector, where interest of promoter lies and after disbursement, monitoring is poor resulting in NPAs [1].

Corporate debt restricting scheme is another scheme which is misused by both corporations and banks. The objective of the scheme was to help genuine debt ridden companies by offering them flexible repayment terms and giving them time according to the economic scenario but the corporation often misused it to cajole the banks to lower their interest rates, prolong the repayment dates and so on.

Sluggish Legal system, inappropriate technology, poor quality management, improper monitoring and follow up processes, managerial deficiencies, and ineffective recoveries are other causes which results in rising NPAs [2].

With this paper, the authors want to propose a methodology for forecasting NPAs, which can thereby be used to forecast NPA of an Indian bank and gauge the crisis which Indian banking system is facing. We intend to forecast the percentage of NPA which banks will have in 2020. The forecasted value can be used by RBI to estimate whether the system will survive or collapse.

In order to develop a holistic model for forecasting NPAs, data from all Scheduled Commercial Banks (SCB) were taken into the fray, and various forecasting models were tried. Iterations were performed so as to come up with a model which was comprehensive and provided satisfactory results. Subsequently, the model was tried on SBI so as to forecast their NPAs by 2020. Ideally, we can choose banks of different sizes and levels of exposure to risks, in order to analyse the NPA. However, we have chosen SBI as a reference point as it is a credible public sector bank. SBI is mammoth in size and its data are not affected by minor fluctuations.

2. Literature Review

Researchers in the past have tried countless number of times to forecast a bank’s performance based on certain macro-economic factors. Roger et al. (2011) [3] focuses on forecasting recession based on a bank’s commercial and industrial loan commitments. The paper emphasises on the fact that a bank’s commitment of loans depends upon future interest rates hence can be used as an indicator for nation’s growth Tobback et al. (2014) [4] has forecasted loss from bad loans on a country’s performance using linear regression models [4]. The magnitude of NPA is comparatively higher in public sectors banks than private sector banks (Singh, 2013) [5]. A high level of NPAs also suggests high probability of a large number of credit defaults that affect the profitability and liquidity of banks. A study on trends of NPAs in India from various dimensions explains how mere recognition of the problem and self-monitoring has been able to reduce it to a great extent (Meenakshi et al., 2010) [6]. On the similar lines, it was found that public sector banks in India, which function to some extent with welfare motives, have as good a record in reducing NPAs as their counterparts in the private sector (ibid). The literature suggests
that the lending policies of various Banks were not proper due to having improper financing. Banks should provide detailed information to the customer about their lending policy (Jain and Sheikh, 2012) \[7\]. Moreover, various Private Banks are not granting Loans outside India, so they could do so to expand their business. They pointed out that instead of focusing on urban areas only, the Banks should set up branches in the rural regions also which could improve their profitability (Beck Roland et al. 2013) \[8\]. study the macroeconomic determinants of nonperforming loans (NPLs) across 75 countries during the past decade. According to their dynamic panel estimates, the following variables are found to be significantly affecting NPL ratios: real GDP growth, share prices, the exchange rate, and the lending interest rate. In the case of exchange rates, the direction of the effect depends on the extent of foreign exchange lending to un-hedged borrowers which is particularly high in countries with pegged or managed exchange rates. In the case of share prices, the impact is found to be larger in countries which have a large stock market relative to GDP.

None of the models have tried to forecast NPA values itself using the macro-economic indicators and we have made an attempt to do the same using both step wise regression as well as logarithmic regression.

3. Research Methodology

The main objective of this research is to propose a methodology for forecasting NPAs which can thereby be used to forecast NPA of an Indian bank and gauge the crisis which Indian banking system is facing.

There are many forecasting methods which can be used to predict the variable of interest. They are broadly divided into two categories:

1) Regression models: Simple Linear Regression (SLR), Multiple Linear Regression,
2) Time series based methods: Moving Average, Simple Exponential Smoothing, Holt, and Holt-winter method.

Regression models are used for prediction purposes. Based on the trend of independent variables, they predict the trend for the dependent variable. Since we will be predicting the NPA’s in 2020 based on external factors, regression models are the best fit in this particular situation. SLR could not be used because it gives the value of dependent variable (Gross NPA’s) in terms of only one independent variable. NPA is often dependent on various factors so this method will never give satisfactory results.

We finally conducted the analysis by MLR method.

3.1. Variables Used in Forecasting Technique

To use MLR we had to determine the factors on which NPA’s in India depend on. Qualitative factors cannot be modelled in the forecasting equations and hence they were neglected for the purpose of this study.

After much deliberation and thought process we came up with the following 4 quantitative variables which were used to predict NPA’s in 2020 of banking sector.

1) Repo rate
   Repo rate is the rate at which the central bank of a country (RBI in case of India) lends money to commercial banks in the event of any shortfall of funds. Repo Rate is generally related to the Bank Prime lending rate as well as reverse repo and MLR. It is an indicator of the prevailing interest rate in the country. Interest rate and inflation has a cumulative effect on the economy and ability of the borrower to pay back. Hence, repo rate is a crucial factor impacting NPAs.

2) Gross domestic product
   A country’s GDP is the total market value of all final goods and services produced in a country in a given year which is equal to the total consumer, investment and government spending. GDP is a growth indicator of an economy. As GDP grows, loans and advances also grow and hence it directly impacts NPAs. Moreover, when the economy is in shambles, corporate will not be able to pay the debts which will thereby lead to an increase in NPA’s.

3) Loans and advances
   Loans and advances are considered the most important factor while forecasting NPAs. As the size of loans and advances increases, the proportion of NPA’s increase due to increase in risk in that case.

4) Inflation rate
   The index is a measure of the average price which consumers spend on a market-based “basket” of goods and services. Inflation based upon the consumer price index (CPI) is the main inflation indicator in most of the
countries. Inflation rate in India is based upon the Indian Consumer Price Index. As inflation rises it becomes cheap for borrowers to borrow money, because of inflation purchasing power of consumer’s fall resulting in a drop in profits for the companies. Combination of both these factors results in rise in NPA’s [9].

3.2. Standard Error Terms

To choose the best model for forecasting, we have standard error terms to be calculated. A comparison of these can help us choose the best fit model.

1) Mean Absolute Deviation (MAD)

It measures the size of the error in units. It is calculated as the average of the unsigned errors.

\[
\text{MAD} = \frac{1}{n} \sum |y_a - y_f|
\]

2) Mean Absolute Percentage error (MAPE)

The MAPE measures the size of error in percentage terms.

\[
\text{MAPE} = \frac{1}{n} \left( \sum |y_a - y_f| \right) / y_a \times 100
\]

\[y_a = \text{actual value of the NPA},\]
\[y_f = \text{forecasted value of the NPA},\]
\[n = \text{Number of the observation}.\]

MAPE and bias are the two statistics which are used by organizations worldwide for choosing the best forecasting model. They use MAPE numbers and subjective knowledge to choose the final forecasting model. We did the same and MAPE was our deciding criteria for the forecasting model.

3.3. Research Process

Research was conducted in two steps.

1) Annual Data (4 factors) for last 11 years (2004 - 2011) was taken for Indian Banks. MLR models were developed using combination of various factors. The best fit model was used to predict the NPAs in the Indian banking sector in year 2020.

2) The same methodology is then used to predict the NPAs for SBI. In case of SBI we took quarterly data. This helped us to check the impact of seasonality via Holt-Winter method. Finally the best fit model was used to predict NPA for SBI in 2020.

3.4. Data Collected

Data for All Scheduled Commercial Banks

<table>
<thead>
<tr>
<th>Year</th>
<th>Gross Advances</th>
<th>CPI</th>
<th>GDP</th>
<th>REPO rate</th>
<th>Gross NPAs</th>
<th>NPA%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>63,548</td>
<td>5.88</td>
<td>2074</td>
<td>7.75</td>
<td>2987</td>
<td>4.70%</td>
</tr>
<tr>
<td>2014</td>
<td>51,025</td>
<td>6.37</td>
<td>2049</td>
<td>8</td>
<td>1759</td>
<td>3.45%</td>
</tr>
<tr>
<td>2013</td>
<td>58,022</td>
<td>10.92</td>
<td>1862</td>
<td>7.75</td>
<td>2395</td>
<td>4.13%</td>
</tr>
<tr>
<td>2012</td>
<td>44,308.34</td>
<td>9.30</td>
<td>1832</td>
<td>8.25</td>
<td>1308.03</td>
<td>2.95%</td>
</tr>
<tr>
<td>2011</td>
<td>38,121.14</td>
<td>8.87</td>
<td>1836</td>
<td>6.25</td>
<td>890.2</td>
<td>2.34%</td>
</tr>
<tr>
<td>2009</td>
<td>28,618.7</td>
<td>10.83</td>
<td>1365</td>
<td>4.75</td>
<td>627.69</td>
<td>2.19%</td>
</tr>
<tr>
<td>2008</td>
<td>23,879.8</td>
<td>8.32</td>
<td>1224</td>
<td>6.50%</td>
<td>522.27</td>
<td>2.19%</td>
</tr>
<tr>
<td>2007</td>
<td>18,458</td>
<td>6.39</td>
<td>1239</td>
<td>7.75%</td>
<td>288.28</td>
<td>1.56%</td>
</tr>
<tr>
<td>2006</td>
<td>14,794</td>
<td>5.79</td>
<td>949</td>
<td>7.25%</td>
<td>491.4</td>
<td>3.32%</td>
</tr>
<tr>
<td>2005</td>
<td>11,153</td>
<td>4.25</td>
<td>834</td>
<td>6.25%</td>
<td>558.7</td>
<td>5.01%</td>
</tr>
<tr>
<td>2004</td>
<td>8674</td>
<td>3.77</td>
<td>722</td>
<td>6.00%</td>
<td>632.56</td>
<td>7.29%</td>
</tr>
</tbody>
</table>

Source: Gross Advances and Gross NPAs data: Annual Publication, RBI (“04-” 15); GDP & CPI data: tradingeconomics.com.
3.5. Data for SBI

The data is collected for all the above quantitative factors for SBI from 2010 Q2. In total there are 23 data points.

<table>
<thead>
<tr>
<th>Year</th>
<th>Quarter</th>
<th>CPI</th>
<th>CPI%</th>
<th>GDP (in USD Billion)</th>
<th>Repo rate</th>
<th>Gross NPA (Rs crore)</th>
<th>Loans and advances (Rs crore)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>Q4</td>
<td>238.63</td>
<td>6.32%</td>
<td>2066</td>
<td>7.25</td>
<td>72,791.73</td>
<td>1,300,026</td>
</tr>
<tr>
<td>2015</td>
<td>Q3</td>
<td>237.94</td>
<td>5.14%</td>
<td>2061.63</td>
<td>7.5</td>
<td>56,834.28</td>
<td>1,232,545</td>
</tr>
<tr>
<td>2015</td>
<td>Q2</td>
<td>236.52</td>
<td>6.10%</td>
<td>2057.26</td>
<td>7.75</td>
<td>56,420.77</td>
<td>1,209,648</td>
</tr>
<tr>
<td>2015</td>
<td>Q1</td>
<td>236.11</td>
<td>6.28%</td>
<td>2052.89</td>
<td>7.75</td>
<td>56,725.34</td>
<td>1,198,903</td>
</tr>
<tr>
<td>2014</td>
<td>Q4</td>
<td>238.34</td>
<td>2.07%</td>
<td>2048.52</td>
<td>8</td>
<td>61,991.45</td>
<td>1,209,829</td>
</tr>
<tr>
<td>2014</td>
<td>Q3</td>
<td>238.03</td>
<td>1.66%</td>
<td>2001.84</td>
<td>8</td>
<td>60,712.38</td>
<td>1,148,901</td>
</tr>
<tr>
<td>2014</td>
<td>Q2</td>
<td>234.81</td>
<td>0.76%</td>
<td>1955.16</td>
<td>8</td>
<td>60,434.24</td>
<td>1,103,090</td>
</tr>
<tr>
<td>2014</td>
<td>Q1</td>
<td>236.29</td>
<td>1.51%</td>
<td>1908.48</td>
<td>8</td>
<td>61,605.35</td>
<td>1,060,689</td>
</tr>
<tr>
<td>2013</td>
<td>Q4</td>
<td>233.5</td>
<td>1.76%</td>
<td>1861.8</td>
<td>7.5</td>
<td>67,799.33</td>
<td>1,045,617</td>
</tr>
<tr>
<td>2013</td>
<td>Q3</td>
<td>234.14</td>
<td>1.18%</td>
<td>1854.295</td>
<td>7.25</td>
<td>64,206.3</td>
<td>978,115</td>
</tr>
<tr>
<td>2013</td>
<td>Q2</td>
<td>233.04</td>
<td>1.50%</td>
<td>1846.79</td>
<td>7.5</td>
<td>60,891.46</td>
<td>926,919</td>
</tr>
<tr>
<td>2013</td>
<td>Q1</td>
<td>232.77</td>
<td>1.47%</td>
<td>1839.285</td>
<td>7.75</td>
<td>51,189.39</td>
<td>916,841</td>
</tr>
<tr>
<td>2012</td>
<td>Q4</td>
<td>229.47</td>
<td>1.66%</td>
<td>1831.78</td>
<td>8</td>
<td>53,457.79</td>
<td>978,115.31</td>
</tr>
<tr>
<td>2012</td>
<td>Q3</td>
<td>231.4</td>
<td>1.99%</td>
<td>1832.7875</td>
<td>8.5</td>
<td>49,202.46</td>
<td>926,918.78</td>
</tr>
<tr>
<td>2012</td>
<td>Q2</td>
<td>229.6</td>
<td>1.73%</td>
<td>1833.795</td>
<td>8.5</td>
<td>47,156.38</td>
<td>897,248.835</td>
</tr>
<tr>
<td>2012</td>
<td>Q1</td>
<td>229.39</td>
<td>2.65%</td>
<td>1834.8025</td>
<td>8.25</td>
<td>39,676.46</td>
<td>867,578.89</td>
</tr>
<tr>
<td>2011</td>
<td>Q4</td>
<td>225.72</td>
<td>3.56%</td>
<td>1835.81</td>
<td>8</td>
<td>40,098.43</td>
<td>846,265.96</td>
</tr>
<tr>
<td>2011</td>
<td>Q3</td>
<td>226.88</td>
<td>3.87%</td>
<td>1803.97</td>
<td>7.25</td>
<td>33,946.31</td>
<td>790,601.23</td>
</tr>
<tr>
<td>2011</td>
<td>Q2</td>
<td>225.69</td>
<td>2.68%</td>
<td>1772.13</td>
<td>6.5</td>
<td>27,768.28</td>
<td>773,660.34</td>
</tr>
<tr>
<td>2011</td>
<td>Q1</td>
<td>223.46</td>
<td>2.68%</td>
<td>1740.29</td>
<td>6.25</td>
<td>25,326.29</td>
<td>756,719.45</td>
</tr>
<tr>
<td>2010</td>
<td>Q4</td>
<td>217.96</td>
<td>1.05%</td>
<td>1708.45</td>
<td>6</td>
<td>23,437.75</td>
<td>718,734.465</td>
</tr>
<tr>
<td>2010</td>
<td>Q3</td>
<td>218.43</td>
<td>1.14%</td>
<td>1622.69</td>
<td>5.75</td>
<td>23,204.59</td>
<td>680,749.48</td>
</tr>
<tr>
<td>2010</td>
<td>Q2</td>
<td>219.79</td>
<td>1.78%</td>
<td>1536.93</td>
<td>5</td>
<td>20,825.22</td>
<td>642,764.495</td>
</tr>
</tbody>
</table>

Source: Annual Report SBI Corporate Website (“10-” 15).

4. Results

4.1. MLR Analysis (Linear, Stepwise and Logarithmic) and Results for Overall Banking Sector

In Multiple Linear Regression we have used a combination of independent variables (CPI, GDP, Repo Rate, Loans and Advances) for finding out their relationship with the NPAs. It can be used to predict the NPA of a given bank on the basis of the independent variables [10] [11]. The MLR model is:

\[
Y_i = \beta_0 + \beta_1X_{i1} + \beta_2X_{i2} + \beta_3X_{i3} + \cdots + \beta_kX_{ik} + \epsilon_i
\]

where,
- \(Y_i\) = Gross NPA for the quarter “i” (dependent/response variable),
- \(X_{ik}\) = independent/explanatory variable taken for regression such as GDP,
- \(\beta_0\) = Y intercept,
- \(\beta_k\) = slope of \(Y\) with respect to \(X_{ik}\), holding other variables \(X_{i1}, X_{i2}, X_{i3}, \cdots\) constant,
- \(\epsilon_i\) = random error in \(Y\) for observation “i”.

MLR models were developed using combination of all variables. The results are shown below:
Independent Variables | R² | Adjusted R² | MSE | MAPE | MAD
--- | --- | --- | --- | --- | ---
CPI, GDP and Gross advances and REPO rate | 0.997 | 0.994 | 2430 | 5.50% | 42.42

**MLR Equation:**

\[ Y_1 = 1534.992 - 9453.88 \times CPI - 1.88389 \times GDP + 0.089585 \times L&A + 23.49054 \times L&A. \]

There can be a case of multicollinearity in this kind of situation. For example, we can fairly guess here that with GDP growth, loans and advances will grow. Hence rather than directly doing MLR analysis we do a stepwise regression which gives us the model with only the significant independent variable and also it keeps in check any issue which arises due to multicollinearity. As expected with stepwise regression we got only three significant variables CPI%, GDP and Loans & Advances. The table below shows results from stepwise regression.

Independent Variables | R² | Adjusted R² | MAPE | MAD
--- | --- | --- | --- | ---
CPI, GDP, Loan & Advances | 0.995 | 0.994 | 6.25% | 45.00

**MLR Equation:**

\[ Y_1 = 1410.268 - 9135.33 \times CPI - 1.7918 \times GDP + 0.091402 \times L&A. \]

For MLR we will choose the equation given by stepwise regression for the reasons mentioned above. While performing Stepwise MLR all other assumptions (homoscedasticity etc.) were also checked. We also explored the possibility of a nonlinear relationship between NPA and the variables identified from stepwise regression. A logarithmic regression model was run between Ln (NPA) as dependent variable and Ln (CPI), Ln (GDP) & Ln (Loan & Advances) to check if there exists a polynomial relationship between NPA and the mentioned independent variables. The table below shows results from the logarithmic regression model.

Independent Variables | R² | Adjusted R² | MAPE | MAD
--- | --- | --- | --- | ---
CPI, GDP, Loan & Advances | 0.90 | 0.86 | 36% | 2.39

**MLR Equation:**

\[ \ln (Y_1) = 2.23982 - 1.13435 \times \ln (CPI) - 3.61072 \times \ln (GDP) + 3.214642 \times \ln (L&A). \]

It can be clearly seen from above results that MAPE of logarithmic regression is very high and hence it cannot be used. We would use stepwise regression equation as our final model [11].

4.2. Analysis and Recommendations for Overall Banking Sector

To compare the results in each methodology we have calculated the MAPE for each one of them.

The MAPE of MLR is 5.50% when all four variables are taken into consideration. R² of MLR is 0.995 which is considered a very good indicator. However, this is because of the fact that data set is only of 11 years. On RBI website data for NPA till last 11 years is mentioned and this resulted in high R². But the purpose was only to devise a methodology and hence 11 data sets also fulfill the purpose.

As discussed in section 5.1 to improve the results of MLR, a stepwise regression model was run for both linear and polynomial relationship. MAPE for linear stepwise model is 6.25% whereas for polynomial relationship it is 36%. MAPE of 6.25% is very good and linear stepwise regression takes care of multicollinearity issue. Hence we would use linear stepwise regression model for NPA prediction as it is the best fit.

4.3. Final Forecasting for Overall Banking Sector

The final forecasting model to be used will be linear stepwise regression model as discussed above. The forecasting equation is:

\[ Y_1 = 1410.268 - 9135.33 \times CPI - 1.7918 \times GDP + 0.091402 \times L&A. \]
We will calculate NPA in 2020 based on the above equation. In 2020 assumptions regarding the independent variables are:
CPI (X1) in 2020 = 6% (assuming stable inflation),
Loans and Advances in 2020 = INR 158,127 billion (Taking loans and advances growth rate of 20% year on year as we expand the banking services),
GDP in 2020 = 4189 billion USD (Taking growth rate of 7% year on year).
With this the Forecasted NPA = INR 7808 billion which is 4.94% of the Average Loans and Advances of all scheduled commercial banks.

4.4. MLR Analysis (Linear, Stepwise and Logarithmic) and Results for SBI

As discussed above the same model was applied on data from SBI to check the accuracy of the model and predict NPA of SBI in 2020.
The Results are shown below:

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>R²</th>
<th>Adjusted R²</th>
<th>MSE</th>
<th>MAPE</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI, GDP, Repo, Loan &amp; Advances</td>
<td>0.86442521</td>
<td>0.834297478</td>
<td>42,201,679.083</td>
<td>10.4765%</td>
<td>4589.064658</td>
</tr>
</tbody>
</table>

**MLR Equation:**

\[ Y_1 = -61850 + 2547 \cdot \text{CPI} - 13.21306 \cdot \text{GDP} + 2795 \cdot \text{REPO RATE} + 0.09570 \cdot \text{L&A}. \]

For multi-collinearity we performed linear stepwise regression. As expected with stepwise regression we got only two significant variables CPI% and Loans & Advances. The important point to notice here is that for doing stepwise regression and accurate prediction we first remove the seasonality component from raw data and then run stepwise regression. We then predict the values from the model and finally seasonize the values to get final output. In the entire process we have assumed multiplicative seasonality. The table below shows results from stepwise regression.

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>R²</th>
<th>Adjusted R²</th>
<th>MAPE</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI, Loan &amp; Advances</td>
<td>0.852</td>
<td>0.837</td>
<td>7.78%</td>
<td>4920.4</td>
</tr>
</tbody>
</table>

**MLR Equation:**

\[ Y_1 = -69720.095 + 0.096 \cdot \text{L&A} + 2994.97 \cdot \text{CPI}. \]

A **logarithmic regression** model was run between Ln (NPA) as dependent variable and Ln (CPI) and Ln (Advances) to check if there exists a polynomial relationship between NPA and the mentioned independent variables. The table below shows results from the logarithmic regression model.

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>R²</th>
<th>Adjusted R²</th>
<th>MAPE</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI, Loan &amp; Advances</td>
<td>0.852</td>
<td>0.837</td>
<td>9.08%</td>
<td>4023.01</td>
</tr>
</tbody>
</table>

It can be clearly seen from above results that stepwise regression and logarithmic regression give comparable results and based on MAPE figures which are marginally better in case of linear stepwise regression we use the same for prediction of NPA.

4.5. Analysis and Recommendations for SBI

To compare the results in each methodology we have calculated the MAPE for each one of them. Contrary to our expectations the MAPE is high in models.
The MAPE of MLR is 10.47% when all four variables are taken into consideration. R² of MLR is 0.8644 which is considered a very good indicator of the fact that the variables chosen determine 86% variability in NPA.
MAPE for linear stepwise model is 7.78% whereas for polynomial relationship it is 9.08%. MAPE of 7.78% is very good and hence we would use linear stepwise regression model for NPA prediction as it is the best fit.
4.6. Final Forecasting for SBI

The final forecasting model to be used will be linear stepwise regression model as discussed above. The forecasting equation is:

\[ Y_1 = -69720.095 + 0.096 \times \text{L&A} + 2994.97 \times \text{CPI}. \]

We will calculate NPA in 2020 based on the above equation.

In 2020, assumptions regarding the independent variables are:

- CPI (X1) in 2020 = 6% (assuming stable inflation),
- Loans and Advances in 2020 = 2,093,704.87 crore (Taking loans and advances growth of 10% year on year as we expand the banking services).

With this the Forecasted NPA = 147,921 crore which is 7.07% of the Average Loans and Advances of SBI.

4.7. Holt-Winter Method for SBI to Account for Seasonality

Holt Method of forecasting has been employed to estimate the NPAs of SBI based on the past 5 years’ quarterly data. The actual data is first de-seasonalized by considering the seasonality in the value of NPAs for a bank. Then as per Holt method, the level and trend adjusted forecasted data is calculated by considering a nominal value of “alpha” and “beta” and the standard sum of errors (SSE) is calculated. This SSE is then minimized using the “solver” function of Excel which takes into account the reference variables and calculates the most optimal value of alpha and beta for the present data. This helps us in forecasting the value of next quarters’ NPA with least amount of error [12].

Forecast Equation: \[ y_{t+1} = \ell_t + hb_t, \]
Level Equation: \[ \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}), \]
Trend Equation: \[ b_t = \beta (\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}, \]
where \( \ell_t \) denotes an estimate of the level of the series at time \( t \), \( b_t \) denotes an estimate of the trend (slope) of the series at time \( t \), \( \alpha \) is the smoothing parameter for the level, \( 0 \leq \alpha \leq 1 \) and \( 0 \leq \alpha \leq 1 \) and \( \beta \) is the smoothing parameter for the trend, \( 0 \leq \beta \leq 1 \).

Holt method we will not use for forecasting as it does not take care of seasonality. We now introduce a third equation to take care of seasonality (sometimes called periodicity). The resulting set of equations is called the “Holt-Winters” (HW) method after the names of the inventors. The exponential smoothing formulae applied to a series with a trend and constant seasonal component using the Holt-Winters additive technique are:

\[ a_t = \alpha (Y_t - S_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1}), \]
\[ b_t = \beta (a_t - a_{t-1}) + (1-\beta)b_{t-1}, \]
\[ S_t = \gamma (Y_t - a_t) + (1-\gamma)S_{t-p}, \]

where, \( \alpha, \beta, \gamma \) are smoothing parameters at, \( B_t \) is the change in the trend at time \( t \), \( S_t \) is the seasonal smooth at time \( t \), \( p \) is the number of seasons per year.

Holt Winter was used to forecast SBI NPA’s and results are given below and see if there is any effect of seasonality in the data. The error indices are on the higher side which is indicative to the fact that NPA’s depend mainly on external macro-economic factors.

**Results of Holt-winter:**

<table>
<thead>
<tr>
<th>Error Indices</th>
<th>MAD</th>
<th>MSE</th>
<th>MAPE</th>
<th>SMAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>4726.82</td>
<td>33,642,435.45</td>
<td>14.2%</td>
<td>19.6%</td>
</tr>
</tbody>
</table>

5. Remedies

The model proposed by us predicts 7.07% of the Average Loans and Advances of SBI will fall under the NPA bracket. The sheer value of the figure, INR 147,921 crore, is enough to point towards a major situation in the making. It is imperative to implement remedial measures so as to deflect the major repercussions of the same.
Some of them have been discussed in detail below:

**1) Strengthening the pre-sanction and after disbursement monitoring process by using technology**

Banks need to monitor their internal and external processes and constantly look for avenues to fortify the same. This is where technology can stake is claim. Data analytics is the new buzzword and has the potential to provide keen insights in the NPA problem. Our PSUs can avail the services of data analytics to create state of the art processes for cash flow generation capacity of a particular company, past records of the borrowers, track a particular sector’s growth projections, track the loan payment schedule etc. Creation of a Central Data repository for all corporate borrowers of all banks would go a long way in establishing coordination and provide a way to track defaulters, thus restricting their entry into the system. Detailed analysis of the fixed assets claimed by the borrowers needs to be done. Proper processes should be in place to check the fixed assets as claimed by the borrowers. Innovative techniques like surprise visits to the factories will keep the borrowers on their toes. Mid-term audits will also ensure a status check of the borrowers, thus allowing banks to take preventive measures [13].

**2) Sector wise planning**

Certain sectors need to be monitored more closely than other. RBI financial stability report states that sectors like infra, aviation, iron and steel, mining, textiles make up 24.8% of total advances of commercial banks whereas they hold a staggering 51.1% in total stressed advances. The concerned sectors should be dealt with extreme caution and extensive due diligence should be carried out before sanctioning anything. Moving forward, banks can think on the lines of creating a credit appraisal team which would cater to specific industries and monitor new applications and their status quo, subsequently raising a red flag if any anomalies are found [14] [15].

**3) Standardisation of 3rd party Vendors**

3rd party agencies like financial analysts, engineers, verification agencies play a pivotal role in concurring the veracity of the claims of borrowers. It is easy to obtain the required documents, provided the 3rd parties are also on the same page. It is important for the government to intervene and create a mechanism which would be conducive to both the borrower and the lender. Monitoring of 3rd party vendors, standardisation via government authorisation can pave the way for reducing forgery and initiating a step towards creating a transparent system [13].

**6. Conclusion**

The linear stepwise regression model was used to forecast NPAs for Indian Banking Sector and it was also used to predict the same for SBI, as a special case. The model predicted that if the current path was treaded, the issue of NPAs would snowball into a major problem with serious repercussions. As far as the Indian Banking Sector is concerned, it was seen that bad loans would amount to INR 7808 billion which is 4.94% of the Average Loans and Advances of all scheduled commercial banks. The forecasting for SBI showed a grim picture where the bad debt would rise to INR 1479.21 billion, which would be 7.07% of the Average Loans and Advances of SBI. The results point that the need of the hour is NPA control so as to avoid the serious consequences that the future holds for us. We also covered various causes for NPAs and their possible solution which could prevent a loan from converting it into an NPA.

**References**


