

Examination of Movement of NPA Data over Time in Selected Public Sector Banks in India with Non Parametric Approach

Gour Bandyopadhyay

Abstract- The research paper is devoted to an analytical study to reveal the movement of the financial parameter, GNPA over time. Diagnostic tool like Quantile Comparison Plots of the Residuals have been used to check the presence of outliers and to detect departure from normality in the residual distribution for our data. Residuals vs Predictor Plot and Absolute Residuals vs Predictors have been employed to examine appropriateness of LOESS Fit (Non Parametric) model for the sample dataset. It is observed that there exist very few outliers in some of the selected banks and residuals follow approximately normal distribution for the sample dataset. It is also observed that LOESS Fit (Quadratic) model fits well with the given dataset and establishes curve linearity. Goodness of fit statistics represented by LOESS R^2 , Sig of F statistics establishes high precision of the model and excellent fit for dataset in respect of the parameter GNPA. Finally, the LOESS Fit model is extended to get the forecasted values for the respective data-set. Forecasted Values of GNPA for three years (2013, 2014 and 2015) of all the selected PSBs clearly demonstrates steep upward trend in future in respect of the financial parameter GNPA for all the selected PSBs, which is undoubtedly an alarming situation and deserves immediate attention on the part of the regulators to relook into the practices of credit appraisal and monitoring of credit in PSBs in India.

Index Terms— Examination of Trend of NPA, Forecasting Model, Gross Non Performing Assets, LOESS Fit Technique, Public Sector Banks in India, Non Parametric Approach, Time Series Data

1 INTRODUCTION

IN traditional banking business, lending to borrower (deficit units) are being financed by deposits from customers (surplus units) and in the process, commercial banks are exposed to the risk of default by the borrower in the payment of either principal or interest. This risk in banking parlance is termed as "credit risk" and accounts where payment of interest and/or repayment of principal are not forthcoming are treated as bad loans or distress assets or Non-Performing Assets (NPAs). Existence of NPA is an integral part of banking and therefore, all banks have NPAs in its advance portfolio, may be in varied proportions.

Asset quality expressed in terms of NPA of a bank is the true reflection of its sound and efficient management of credit portfolio. In an effort to examine trends in NPA and develop a forecasting model for the purpose of analysis and control of bad loans in selected PSBs, various approaches and techniques may be used by the researchers and experts. In this paper, LOESS Fit, a very popular non parametric technique has been fitted to GNPA data sets for six selected PSBs to examine trend in sample dataset and develop a forecasting model. The study contributes to the existing literature by modeling parameter GNPA of six selected PSBs, which has seldom been attempted earlier with such technique.

1.1 Statement of the Problem

NPAs have dampening effect on banking system since long, though they were not in the public domain till early 90s (Khasnobis, 2006) [1]. Norms on NPA was implemented for the first time by RBI in 1992-93, when Gross Non Performing Asset (GNPA) of all PSBs were Rs 39253 crores, representing 23.18% of Gross Advances (GAs), which rose to Rs 112489 crores, representing 3.17% of the GAs as on March 2012. Such a high figure of loan defaults is directly eating away vitality of the banks in India and forces us to relook into their credit appraisal and follow up system, (Sharma, 2005) [2]. The above estimates of NPAs in PSBs, is believed to significantly understate the actual gravity of the situation (Basu, 2005) [3]. Such a mammoth NPA figure in PSBs is also a constant worry for the regulators and the ministry, as banking plays a significant role in a developing country like India. Moreover, as slightest casual approach, on any front may put a bank into serious trouble, the task of managing asset quality in banks has become a serious concern for the practicing managers. Therefore, managing asset portfolio has become the top most priority in PSBs in India, which requires focused and planned effort including examining trend of NPA data over time, forecasting future value of distress asset and their effective management.

1.2 Defination of Terms

A) Non Performing Assets: Non Performing Assets (NPAs), are loan assets which cease to generate income to the bank. It includes borrowers' defaults or delays in payment of interest or principal repayment. In practice, a bank classifies an account as NPA only if the interest charged /instalment due

• Gour Bandyopadhyay is currently pursuing PhD degree program in Management in West Bengal University of Technology, India, PH-9119432083462. E-mail:gaur_banerjee@yahoo.co.in

during any quarter is not serviced fully within 90 days from the end of the quarter. These assets have well defined credit weakness that jeopardize the liquidation of debts and may be characterized by distinct possibilities that bank will sustain some losses. With a view to moving towards international best practices and to ensure greater transparency, it has been decided by RBI to adopt the "90 days' overdue" norm for identification of NPAs in India, from the year ended March 31, 2004.

B) Gross Non Performing Asset (GNPA) refers to the sum total of all loan assets that are classified as NPAs as per RBI guidelines as on balance sheet date and reflects the asset quality of a bank. It consists of all the nonstandard assets like as sub-standard, doubtful, and loss assets. In other words, GNPA is the amount outstanding in the borrower's accounts in the books of the banks other than the interest which has been recorded and not debited to the borrower accounts.

C) Public Sector Banks (PSBs) refers to those commercial banks where the Government (Central and/or State and RBI) holds majority shares of the company and have absolute control in the management of the same.

Public Sector Banks Group comprises of:

- ✓ State Bank of India and its Subsidiaries, popularly known as State Bank Group
- ✓ Other Nationalized Banks, popularly known as Nationalised Bank Group

1.3 Theoretical Framework

A model is a representation of an observed dataset of a real life phenomenon presented as per standard analytical practices and expressed by means of equation. It is well known that the linear model is most attractive because of its simplicity to fit, easily understandable results and availability of a wide range of techniques for testing the assumptions. However, it should also be noted that, in many cases because of an intrinsic nonlinearity in the data, linear model becomes inappropriate. Moreover, real life data situations are not easy to model because of the uncertainty and variation prevailing in the natural system and also due to presence of the interaction of unknown factors interplaying and governing the entire real life phenomenon. In respect of real life data, parametric non linear models and related methods may be too restrictive and as a result various non parametric techniques (that do not assume the data have any characteristic structure or parameters) for modeling non linear time series data have become popular amongst the researchers, to examine the underlying structure of the data. While analyzing a non linear time series data either parametric non linear or non parametric time series model is used. In the later method, according to Casas and Gao (2007) [4], the data may be allowed to speak for themselves in determining mathematical relationship between variables. This makes the non parametric techniques more feasible and flexible.

Wu (2000) [5] has stated that a general non parametric model can be written as

$$y_i = g(t_i) + e_i$$

where $i = 1, 2, 3 \dots, n$ and y_i is the response variable at time t_i ,

$g(\cdot)$ is a unknown smooth function and e_i is the random noise with mean zero.

For fitting smooth curves to time series non linear data, LOESS is considered as a powerful and simple technique. LOESS is the acronym for "local regression". In this paper, the data-sets relating to parameters GNPA in respect of six selected PSBs have been analysed with the help of LOESS procedure and forecasting future value for medium term have been attempted.

There are different types of nonparametric regression in use. One of the most commonly used method is LOESS or LOWESS procedure. The short form of locally weighted scatterplot smoother is LOWESS and local regression is LOESS. LOWESS and LOESS is essentially same thing. "LOESS has some highly desirable statistical properties, (it) is easy to compute, and ... (it) is easy to use" (Cleveland, 1993, p. 94) [6]. These models are used to fit local polynomial regressions and join them together.

Steps in Developing LOESS Smooth Fit

The procedure for developing a LOESS smooth fit as described by Jacoby (2000) [7], in his published paper 'LOESS: a non parametric graphical tool for depicting relationships between variables' and by Jacoby (n.d.) [8], in his Lecture note 'Lecture 14: Non-normal Error Distribution and Nonlinear Functional Forms' on 'Regression III : Advanced Methods' in Michigan State University is stated below:

Defining the window width

The first step is to define the window width (q) that encloses the closest neighbours to each data observation (the window half-width is labeled h). Here, for each data point the researcher selects the ' q ' nearest neighbours in terms of their X -value. This q was chosen to represent a particular percentage of the data. The researcher typically chooses the window width by trial and error. The researchers start with the first observation and move through the data, finding the ' q ' closest observations to each case. In each case, the observation is called the focal X . Initially, m equally-spaced locations across the range of X values have been defined calling these v_j , where $j = 1$ to m . At each of these v_j the LOESS curve will be evaluated. We choose $\hat{g}(v_j)$ as the LOESS fitted value at v_j .

After this the 2 parameters α (between 0 and 1) and λ (either 1 or 2) are supplied. The number of observations used in each local regression (called 'the window') is found by using the value of α . The window width is defined as, $q = \alpha n$ (where n is the number of empirical data points, and q is truncated to an integer value). The degree of polynomial that is fitted to the data is given by the value of λ .

Weighting the data

To give the greatest weight to the observations that are closest to the focal X observation, a weight function is chosen by the researchers. In most of the cases the tricube weight function is used. The distance from the point of evaluation (v_j) to the i th observation x_i is given by

$$\Delta_{[i]}(v_j) = |x_i - v_j|$$

The distances are sorted from smallest to largest. The same distance, expressed as a proportion of the distance from v_j to the farthest data point within the window is given by,

$$\Delta_{[i]}(v_j)^* = \frac{\Delta_{[i]}(v_j)}{\Delta_{[q]}(v_j)}$$

The neighbourhood weight is calculated for all observations, from $i=1$ to n . using the tricube weight function as follows:

$$w_i(v_j) = \begin{cases} (1 - |\Delta_{[i]}(v_j)^*|^3)^3 & \text{for } \Delta_{[i]}(v_j)^* < 1 \\ 0 & \text{otherwise} \end{cases}$$

Locally weighted least squares

The researchers then apply a polynomial regression to the focal X observation, using only the nearest neighbour observations. This polynomial regression uses weighted least squares (using the tricube weights). As a result the weighted residual sum of squares is minimized. Normally either local linear regression or local quadratic regression is used but the use of higher order polynomials is also allowed. The fitted value for the focal X value is then calculated from this regression. The fitted value is then plotted on the scatterplot. First of all, the coefficient b_k is to be found that minimizes the following:

$$\sum_{i=1}^n w_i(v_j) \left(y_i - \left[\sum_{k=0}^{\lambda} b_{kj} x_i^k \right] \right)^2$$

The value of λ can be set as 1 or 2, ($\lambda=1$ represent a linear equation, while $\lambda=2$ represent a quadratic equation) is fitted to the weighted data using above equation. The fitted value is obtained by taking the predicted value of Y at v_j , after the coefficients of the preceding equation are found.

$$\hat{g}(v_j) = \sum_{k=0}^{\lambda} b_{kj} v_j^k$$

The Nonparametric Curve

For each observation in the data, defining the window width, weighting the data and locally weighted least squares steps are carried out. For each value of X , there is a separate local regression. For each focal X , a fitted value from these regressions is obtained. In other words, all the points $(v_j, \hat{g}(v_j))$ are calculated using the above regression method. These points are plotted and connected as the solid lines. As a result, the local polynomial nonparametric regression curve is obtained. A smoother curve is obtained by increasing the window width. It is important to note that in this method, coefficient estimates are absent. Here, the relationship between X and Y is graphed.

Robustness Weights

In the LOESS fitting procedure the robustness step is optional. Belsley (1980, as cited in Jacoby, 2000) [7] observes that the unusual observations can adversely and strongly influence the LOESS like other least square methods.

Such unusual observations (outliers) are down weighted by

the robustness step of the LOESS procedure and as a result, the more concentrated areas of data points are more likely to be followed by a smooth curve. The robustness weights in the LOESS fitting procedure is routinely included to produce a graphical summary of the bi-variate data. The residuals may not be normally distributed with a mean of zero and constant variance, which is considered to be the only potential disadvantage of robust LOESS estimation. Thus it can be said, a no robust LOESS fit can produce a misleading representation of the predominant structure within the data (Jacoby, 2000) [7].

Diagnostics Procedures

Best fit in nonparametric regression is found with proper variable selection as well as the choice of smoothing parameter. Like parametric regression, residuals are the most important diagnostic component in case of nonparametric regression also. Several plots are used as diagnostic tool as described below.

Quantile Comparison Plots of the Residuals

In non parametric regression, residuals are defined as the difference between the observed values of the Y variable, and the corresponding fitted values for the respective occurrences of the X variable values. This plot shows the residual plot from the original LOESS curve that are fitted to the data. The points in this figure are obtained by plotting the LOESS residual values (on the vertical axis) against normal norm quantiles plot (on the horizontal axis) to check the presence of outliers in the data variables. The plot is also used to detect departures from normality in the residual distribution.

Residuals vs Predictor Plot

The LOESS residuals are defined as:

$$E_i = y_i - \hat{g}(x_i)$$

where y_i is the observed value and $\hat{g}(x_i)$ is the fitted value. It is to be noted that, since the m evaluation points have been used to find the LOESS curve (the v_j s) are imaginary values which are usually different from the n observed values of the independent variable hence interpolation between the two closest occurrences of the equally-spaced evaluation points are used to find the fitted values for the empirical observations, $\hat{g}(x_i)$.

The LOESS residuals are plotted against either the corresponding fitted values or the values of the original independent variable (X) and a LOESS curve is fitted to the points within the residual plot. A flat line located at the zero value on the vertical axis in the residual plot is obtained as a result of this new application of the LOESS smoother.

Jacoby (2000) [7] states that:

“The reasoning is as follows. The LOESS residuals measure the variability in Y that remains after the dispersion of the fitted values (and hence, the smooth curve) is taken into account. Any systematic functional dependencies between X and Y should be picked up by the original smooth curve fitted to the bivariate data. To the extent that the LOESS fitting process does so successfully, there should be no discernible patterns of any kind among the residuals; this, in

turn, would produce a horizontal line when a smooth curve is fitted to the residual plot (Cleveland, 1993)". (p. 590)

If the plot does not reveal any kind of systematic relationship (whether linear, cubic, quadratic, etc.) between the residuals and the predicted values, we may say LOESS model is appropriate for our data.

Jacoby (2000) [7] argues that:

Residual plots can provide the analyst with useful guidance for controlling the LOESS fitting process (Cleveland, 1993, 1994). (p.591)

"This is particularly important for selecting the proper value of the smoothing parameter α , as the appropriate λ value and the need for the robustness weights can often be determined through visual inspection of the original scatter plot" (Jacoby, 2000, p. 591) [7].

Absolute Residuals vs Predictors

The LOESS absolute residuals are plotted against the corresponding predicted values of the original independent variable (X). The points in this figure are obtained by plotting the LOESS absolute residual values (on the vertical axis) against predictors (on the horizontal axis). More effective is to plot residuals (or their absolute or squared values) against the predictor variable or, equivalently, against the fitted values. This plot is primarily used to detect dependence of the residuals on predictors. The lack of fit would result in a graph showing the residuals departing from zeros in a systematic fashion likely a "megaphone" shape.

Goodness of Fit Statistics

R^2_{LOESS}

The LOESS smooth curve discloses the structure within the data. By examining the 'goodness of fit' the researchers can understand how well the smooth curve characterizes the data values. In case of nonparametric methods like LOESS the 'goodness of fit' is partially appropriate. The ratio of the sum of squares in the LOESS fitted values to the total sum of squares in the dependent variable gives us a summary fit statistics similar to R^2 .

$$R^2_{loess} = \frac{\sum_{i=1}^n (\hat{g}(x_i) - \bar{\hat{g}(x_i)})^2}{\sum_{i=1}^n (y_i - \bar{Y})^2}$$

Where, $\hat{g}(x_i)$ is the LOESS fitted value for observation i , $\bar{\hat{g}(x_i)}$ is the mean of the LOESS fitted values, y_i is the dependent variable value for observation i , \bar{Y} is the sample mean for the dependent variable.

There are three major limitations in case of interpretation of R^2_{LOESS} value.

The R^2_{LOESS} cannot be interpreted as variance explained in true sense.

When robustness weights are used in the fitting process, misleading or even meaningless results can be produced by R^2_{LOESS} statistic.

A large R^2_{LOESS} value should not always be expected.

The LOESS smoother does not fit a particular, narrowly defined model to data (Weisberg, 1996) [9]. Therefore the 'goodness of fit' concept becomes tricky. However, the statistics can be used to give limited interpretation. It conveys the size of the fitted value variance, expressed as a ratio of the total variance in Y and provide an effective summary of the degree to which the LOESS fitted values track the empirical data points in the scatterplot (Jacoby, 2000) [7].

Decision Rule: A relatively high R^2_{LOESS} value would lead to the conclusion that the smooth curve summarizes nearly all of the total dispersion in the dependent variable.

LOESS F statistics

The researchers use the approximate degrees of freedom to carry out F tests. The degrees of freedom is also used to construct the confidence envelopes around the fitted curve.

The approximate degrees of freedom is given by,

$$df_{MOD} = \text{trace}(S).$$

There are several other equivalent ways to define the degrees of freedom also. The residual degrees of freedom is given by,

$$df_{RES} = n - df_{MOD}.$$

The estimated error variance is given by,

$$s^2 = \sum e_i^2 / df_{RES}$$

It is to be noted that in case of nonparametric regression the degrees of freedom are not necessarily whole numbers. The researchers generally plot the 95% confidence band or envelope for the regression function by joining the confidence intervals for each of the X-values together.

The researcher use F tests to compare the residual sums of squares for alternative nested models. Since the degrees of freedom are approximate hence these tests are approximate only. It is worthwhile to note that the test of nonlinearity by constructing it with a linear regression model is permitted by the nonparametric regression. Here, we have nested models because a linear model is a special case of a general nonlinear relationship.

The F test takes the usual form:

$$F_0 = \frac{\frac{RSS_0 - RSS_1}{\text{trace}(S) - 2}}{\frac{RSS_1}{n - \text{trace}(S)}}$$

Where, RSS_0 is the residual sum of squares from the linear model, RSS_1 is the residual sum of squares from the nonparametric model and $\text{trace}(S)$ is the df for the nonparametric model.

Decision rule: If p-value of F test is less than 0.05 then we conclude that there is no linear relationship between response and predictor variables.

1.4 Objective of the Study

In view of the relative importance of NPAs in banking sector in India in general with PSBs in particular, it is perceived that a comprehensive study in this area should be made. The present study is a humble endeavour to examine various aspects of NPAs in selected PSBs. The specific objectives embodied under the research are as follows:

- i) to examine the overall trends of NPAs and to explore the dynamicity of NPA as the variable under study over time in six selected PSBs;
- ii) to develop a forecasting model with the help of univariate dataset of GNPA (dependent variable) and Time (independent variable)

1.5 Methodology

The study undertakes an empirical approach to analyse the movement of NPAs in six selected PSBs India over the last two decades, based on secondary data related to the strategic banking variable, ie, NPAs. The secondary data have been collected from RBI publications, Prowess Database of Centre for Monitoring Indian Economy and the Annual Reports of the selected PSBs.

The study includes examination of trends of NPA, developing forecasting model for medium term with NPA as dependent variable. For the purpose of our analysis, a time series data-set on parameter GNPA for 17 years i.e March 1996 to March 2012 for six selected PSBs have been captured and analysed by using R software in order to draw relevant inference.

To examine dynamicity of NPA over time as stated in the first objective, nonparametric, nonlinear regression models have been invoked by employing LOESS Fit technique and the best fitted model to represent the trend has been obtained in case of each data-set. The forecasted values have been generated based on the said best fitted model as stated in objective number two.

1.6 Hypothesis

Null Hypothesis (H_0): There is no linear relationship between Non Performing Assets (response variable) and time (predictor variable).

Alternate Hypothesis (H_1): There is a linear relationship between Non Performing Assets (response variable) and time (predictor variable).

1.7 Review of Literature

In view of the seriousness of the problem, numerous research studies have been conducted on different issues concerning credit risks including modeling of NPA, a brief account of which are given below.

LOESS methodology was first proposed by Cleveland (1979, as cited in Jacoby, 2000) [7] for modeling and smoothing bivariate data and further developed by Cleveland and Devlin (1988, as cited in Jacoby, 2000) [7]. According to McLeod (n.d.)

[10], "the technique provides a general and flexible non linear family of models to fit bivariate data" (p. 1). LOESS does not require any priori specification of the relationship between the dependent and independent variables and is often used as a scatter plot smoother. The technique also contains inferential procedures for confidence intervals and other statistical tests and can also be applied to multivariate data. For all of these reasons, LOESS is a useful tool for data exploration and analysis in the social sciences (Jacoby, 2000) [7]. Dias (n.d.) [11] states "depicting the 'local' relationship between a response variable and a predictor variable over parts of their ranges, which may differ from a 'global' relationship determined using the whole data set toward smoothing a scatter plot is the essence of the LOESS procedure" (p. 1).

Bhide (2002) [12] observes that the most critical issue in Indian banking is that of its bad loans which has been understated over years due to lax regulatory provisions towards asset classification and provisioning in banking in India.

Ranjan and Dhal (2003) [13] examine the factors responsible for increasing levels of NPAs and observe that NPLs are influenced by three major sets of economic and financial factors, i.e., terms of credit, bank size induced risk preferences and macroeconomic shocks. A panel regression model for forecasting GNPA ratio has been developed. The empirical results from panel regression models suggest that terms of credit variables have significant effect on the banks' NPL in the presence of bank size induced by risk preferences and macroeconomic shocks. The study also observes that factors like maturity of credit, better credit culture, and favorable macroeconomic and business conditions lead to lowering of NPAs. Business cycle may have differential implications adducing to differential response of borrowers and lenders.

Gopalakrishnan (2006) [14] develops a statistical model for forecasting future value of gross advances, standard advances and NPAs with the help of regression analysis. Time series data for nine years have been taken and regressed with a few strategic variables under two scenarios to enable prediction of the above parameters for 11 years as a matter of reference for the bankers in formulating their strategies.

Sethi and Bhatia (2007) [15], introduce the readers to the banking business, with exclusive and detailed discussion on banking sector reforms. Issues, dimensions and management of NPAs in commercial banks have also been deliberated with reasonable precision.

Rajeev and Mahesh (2010) [16] in their exploratory paper examine the trends of NPAs in India from various dimensions. The study of trend in this paper has been examined on both GNPA, NNPA and sector wise distribution of NPA both on absolute figure of NPA as well as on ratio data with a time horizon of 2002-2009. The study observes that decline in NNPA is sharper than GNPA because of the increasing level of provisions. The study also observes that mere recognition of the problem and self-monitoring has been able to reduce it to a great extent. It also shows that PSBs in India, which function to some extent with welfare motives, have performed rather well in reducing NPAs as their counterparts in the private sector.

Pradhan (2012) [17], in his study of trends of NPA in PSBs

during post reform period, examines trends in GNPA, GNPA Ratio, PSB group-wise as well as individual PSB wise, and observes that rate of decline in GNPA in absolute term has been extremely low for PSBs group-wise during the last decade. However asset quality in PSBs in ratio terms like GNPA to Gross Advances has reduced remarkably during post reform period. With regard to individual bank's performance in management of NPA, it is found that Andhra Bank and Indian Bank are the leaders during the period under consideration.

Rajamohan (2012) [18] makes an attempt to examine the status, growth and sector wise movement of NPA in PSBs, group wise with dataset covering a period from 2001-02 to 2010-11. The study observes that NPA figure in absolute term has registered consistent fall for six to seven years for both Nationalised Bank (NB) Group and State Bank (SB) Group after which NPA rose very sharply for the remaining three to four years. With regard to NPA in priority sector, the study also observes that it remained moderately stable till 2009 beyond which NPA figure rose quite sharply. The study also examines whether there is significant difference in the growth rate of NPA between NB Group and SB Group with the help of Mann Whitney U Test and observes that there is significant difference between the growth rate of NPAs in these banks.

Rawlin, Sharan and Lakshmipathy (2012) [19] examine the relationship between gross and net NPA% of a bank with its aggregate advances. A strong correlation is observed between gross and net NPA% and the advances made. Based on the above observations attempt is made to predict gross and net NPA% from advances by fitting to Linear and non linear models. A non linear curve estimation model linking both Gross and Net NPA to advances provided the best curve fit and the least deviation from actual values. Thus by simply looking at advances an overall picture of the banks NPA level can be ascertained.

Veerakumar (2012) [20] in an attempt to study NPA in Priority Sector Advances (PSA) in a commercial bank examines time series data from 2000-01 to 2009-10 of both GNPA and NNPA (on absolute figure as well as ratio data) and presents trend analysis with the help of curve estimation regression technique. It is observed that a polynomial regression model, particularly a cubic curve is the best fit with very high R^2 value, signifying a dependable forecasting model. The researcher has also used multiple linear regression analysis to examine the influence of different variables on GNPA of SCBs.

Bandyopadhyay (2013) [21] examines movement of NPA over time. Diagnostic tool like Residuals vs Predictor Plot, Quantile Comparison Plots of the Residuals and QQ Plot have been employed to examine appropriateness of Penalized Spline (Semi Parametric Curve Fit) model, to check the presence of outliers, and to detect departures from normality in the residual distribution for our data respectively. It is observed that Penalized Spline model fits well with the given dataset. It is also observed that there exist very few outliers in some of the selected banks and residuals follow approximately normal distribution for the sample dataset. Penalized Spline model, establishes curve linearity in the given data. Goodness of fit statistics represented by R^2 , Sig of F statistics establishes high precision of the model and excellent fit for dataset in re-

spect of the NPA. For all banks, the Penalized Spline model is extended to get the forecasted values for the respective dataset. Forecasted Values of NPA for three years (2013, 2014 and 2015) of all the selected PSBs clearly demonstrates future upward trend in respect of NPA, which puts question mark on the wisdom and integrity of the top management in PSBs in India in handling credit portfolio. Such a situation undoubtedly deserves immediate and serious attention on the part of the regulators to relook into the practices of credit appraisal and monitoring of credit in PSBs in India.

1.8 Scope of the Study

Since NPA in PSB constitute approximately 79% of NPA as on March 2012 of total banking industry. Our study covers dataset from 1995-96 (the earliest bankwise NPA data available) to 2011-12 (the latest year till the dataset is available).

For the purpose of the study, we have taken one large sized PSB {State Bank of India (SBI)}, two medium sized PSBs {Punjab National Bank (PNB) and Central Bank of India (CBI)} and three small sized PSBs {State Bank of Travancore (SBT), UCO Bank (UCO) and Syndicate Bank (SB)}.

1.9 Limitation of the Study

The study makes an attempt to examine empirically trends of NPAs for six selected PSBs. However studies revealed macro economic factors like global recession, high rate of inflation have significant influence on NPAs of a bank which is beyond the scope of our study.

Time series GNPA data is available for seventeen years for six selected PSBs. Therefore the dataset can be said to be very small under any standard.

1.10 Significance of the Study

Despite the importance of monitoring non-performing loans, forecasting on NPA have only received moderate attention in literature. This study contributes to the existing literature by modeling NPA of six selected PSBs in India using non parametric LOESS Fit technique.

By examining trends in NPA time series data and forecasting NPA for medium term in six selected PSBs, as attempted in the study, a bank will be in a position to initiate corrective actions as appropriate towards improving level of distress asset in the bank resulting in a great relief for the economy and society.

2 BODY

2.1 Background of the Study

A number of factors make the NPAs in PSBs in India an interesting subject for study.

- ✓ First, during the 1990s, India underwent liberalization of the banking sector with the objective of enhancing efficiency, productivity, and profitability (India. RBI, 1991)[22].

- ✓ Second, the banking sector underwent an important transformation, driven by the need for creating a market-driven, productive, and competitive economy in order to support higher investment levels and accentuate growth (India. GOI, 1998) [23].
- ✓ Third, studies on NPAs in banking industry in emerging economies like India has great relevance, as it dampens the bottom-lines and thereby poses a serious threat to the very existence of the most important sector in propelling the desired growth and development of the economy.

In view of the seriousness of the problem, numerous research studies have been conducted on different issues concerning credit risks in banks including NPAs. However, empirical works on NPA problems in PSBs are inadequate. The exhaustive review of literature on NPA demonstrates that majority of the research work has been undertaken on aggregate PSBs data.

However empirical work on individual bank-wise trend study and modeling for the purpose of forecasting future value on NPA in PSB has seldom been attempted. It is against this backdrop that the present study is undertaken to fill up this gap and make a modest contribution in the field of management of NPA in banks in India. Accordingly, examination of movement of NPA over time and forecasting the same for medium term with the help univariate time series data by employing non parametric (LOESS Fit) for six selected PSBs has been attempted in the paper.

2.2 Developing LOESS Smooth Fit

Defining the window width

In GNPA dataset, there are 17 v_j values (that is, $n=17$), uniformly spaced in the closed interval from 1 to 17 (i.e., year 1996-2012); hence, $v_1=1$, $v_2=2$, and so on, up to $v_{17}=17$. The locations of the various v_j s are shown as tick marks on the horizontal scale of all the figures. Note that the v_j s would usually range only across the observed values of the X variable. In this case, α is set as 0.6 (i.e., 60% of the data), so $\alpha n=10$. Thus, the window will always enclose the 10 empirical data points that fall closest to the current v_j on the horizontal axis. Figures below show the window for $v_j=8$ th year (Year 2003) (an arbitrarily-selected value, located by the asterisk along the horizontal axis in the figure). Note that the physical width of this window will change at different values of v_j , as the distance to the 10th closest data point changes. Also, evaluation points close to either the maximum or minimum X values will be asymmetric; to pick some obvious examples, v_1 will have all 10 points in its window arrayed to the right, while v_{17} will have all 10 points arrayed to its left. In any event, the window will always contain the 10 closest empirical data points, regardless of their direction and/or distance from v_j .

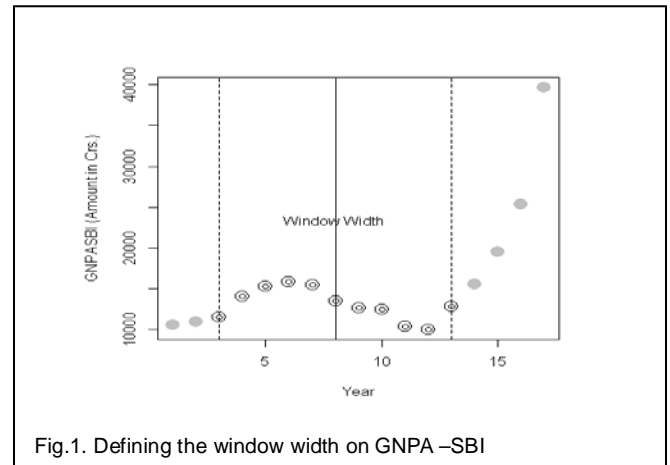


Fig.1. Defining the window width on GNPA –SBI

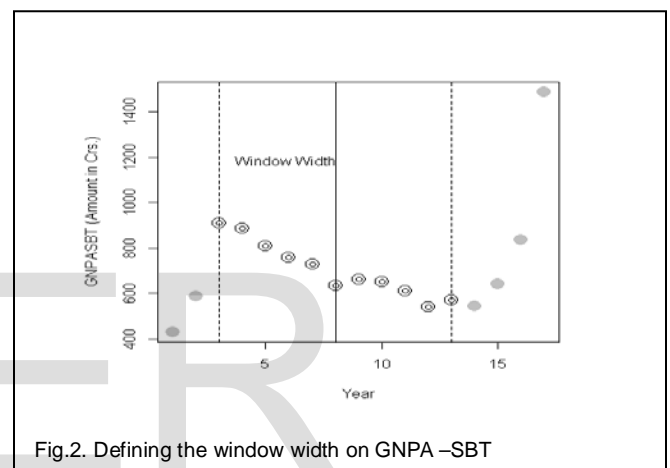


Fig.2. Defining the window width on GNPA –SBT

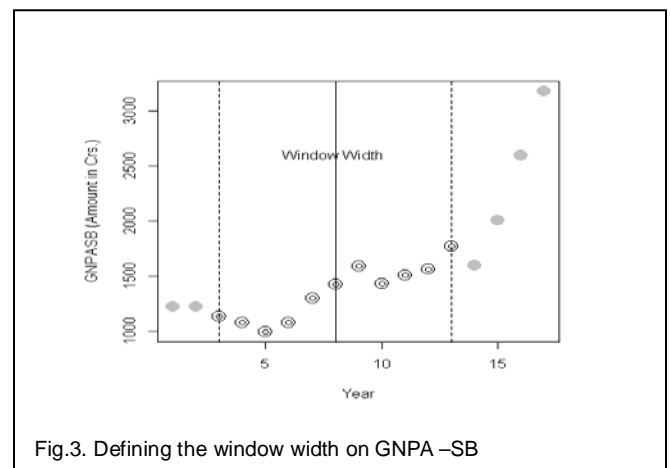


Fig.3. Defining the window width on GNPA –SB

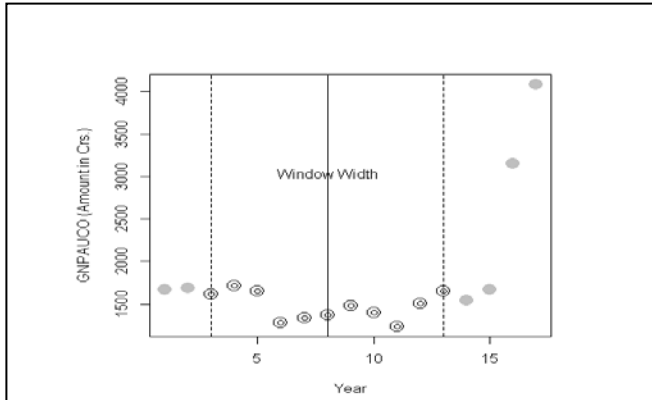


Fig.4. Defining the window width on GNPA –UCO

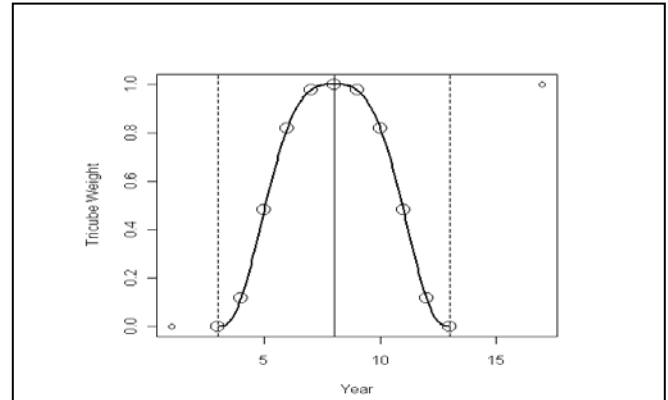


Fig.7. Tricube weight on GNPA

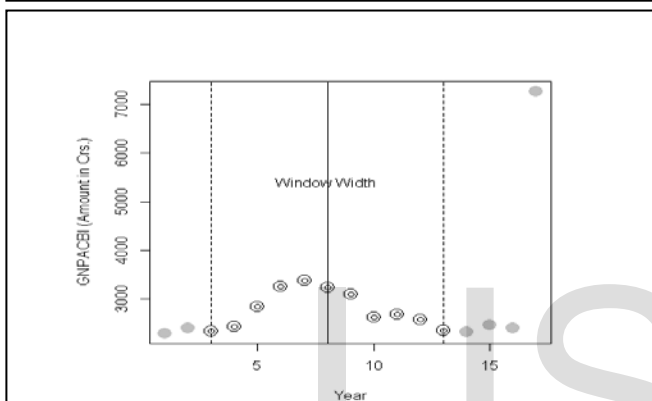


Fig.5. Defining the window width on GNPA –CBI

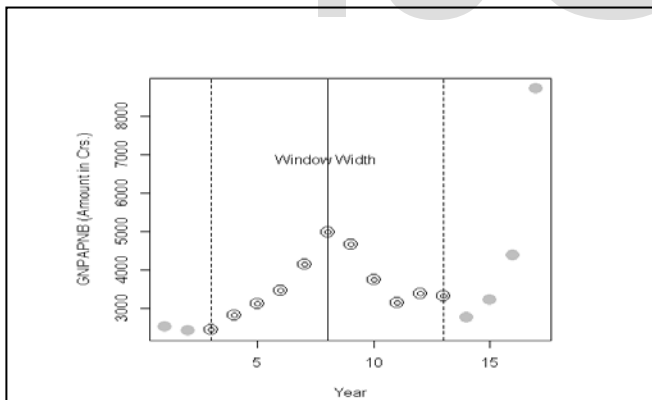


Fig.6. Defining the window width on GNPA –PNB

Summary Statistics of the Non parametric (LOESS) Fit (Parameter GNPA)

Now the GNPA dataset as dependent variable with time (year as 1996, 1997, 1998...2012 taken as 1, 2, 3, ...17 respectively) as independent variable are processed through R software and summary statistics generated by R are given below.

TABLE 1
 SUMMARY STATISTICS FOR LOESS FIT MODELS

No. of Observation	17
Equivalent Number of Parameters	5.36
Trace of smoother matrix	5.91
Control settings	Normalize: TRUE, Span: 0.6, Degree: 2, Family: Gaussian, Surface: interpolate, Cell = 0.2

LOESS Fit (Quadratic) Model

The graphs and plots of the non parametric models (LOESS fit) for six selected PSBs are given below:

The GNPA of the various banks plotted (on the vertical axis) against the Time (Years) from 1996-2012 (on the horizontal axis) are shown below.

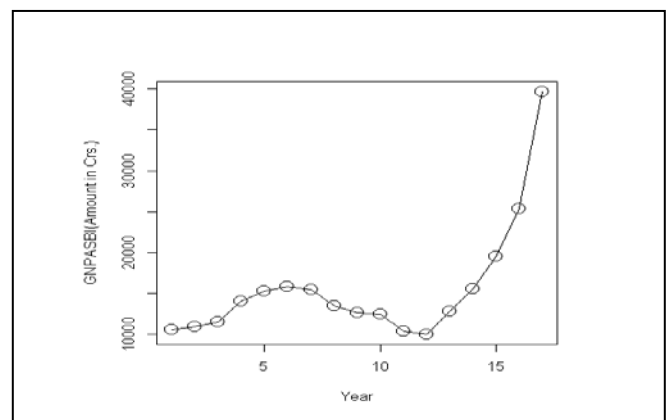


Fig.8. LOESS fit Quadratic model on GNPA –SBI

Weighting the data

For each v_i , neighborhood weights are calculated. Neighborhood weights are calculated for all observations, from $i=1$ to n for all GNPA values. The shape of the tricube weight function, as well as the specific weights assigned to the 11 observations, is shown in figure below. It is noted that the observations with X values close to 8th (year 2003) have large weights, near 1.0. The weights fall off fairly quickly for observations with X values substantially different (in either direction) from the current evaluation point of $v_i=8$ th (year 2003).

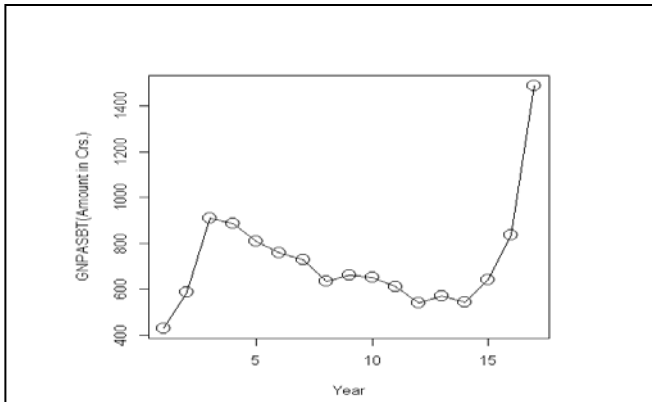


Fig.9. LOESS fit Quadratic model on GNPA –SBT

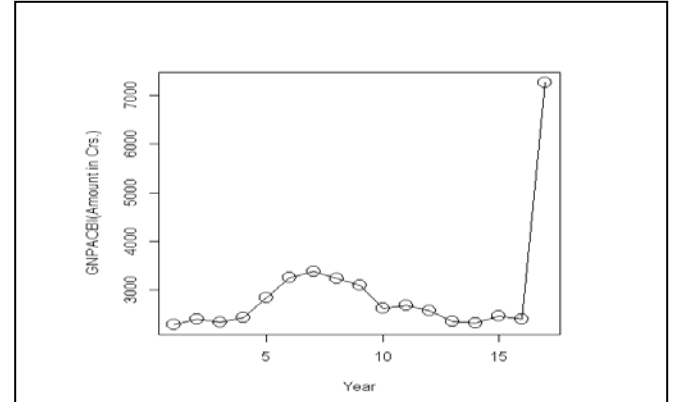


Fig.12. LOESS fit Quadratic model on GNPA –CBI

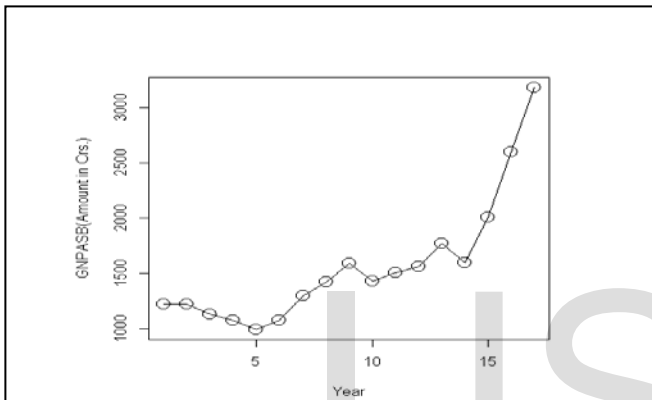


Fig.10. LOESS fit Quadratic model on GNPA –SB

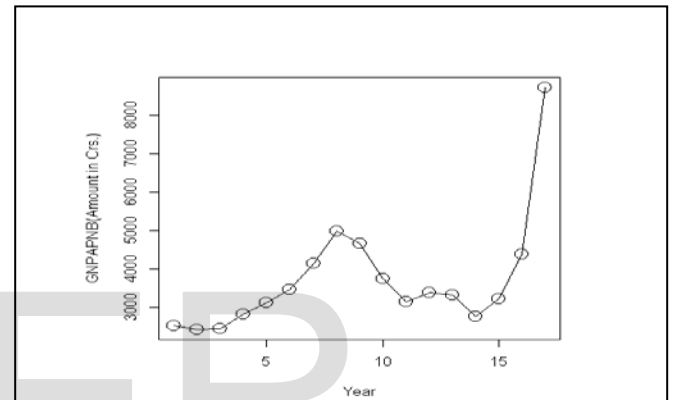


Fig.13. LOESS fit Quadratic model on GNPA –PNB

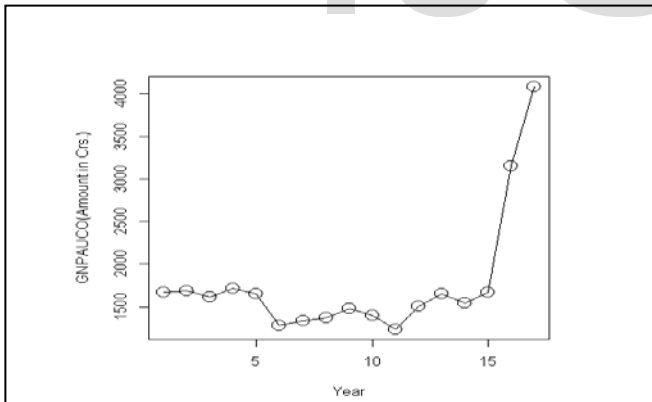


Fig.11. LOESS fit Quadratic model on GNPA –UCO

Robustness Weights

The effect of robust fitting option is given below. The solid black lines show the robust LOESS (LOESS.smooth) curves. The red lines show the LOESS curves obtained when the robustness weights are omitted from the fitting procedure. The deep blue lines show the robust lowess curves. The light blue lines show the lowess curves obtained when the robustness weights are omitted from the fitting procedure. All the robust LOESS and LOESS curves are fit with α set at 60% and $\lambda=2$.

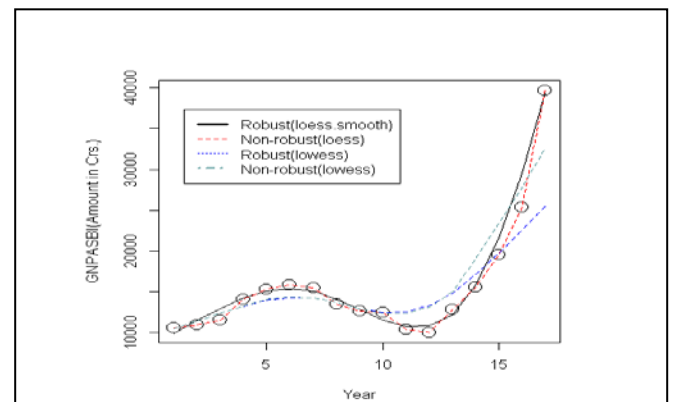


Fig.14. Robust LOESS fit on GNPA –SBI

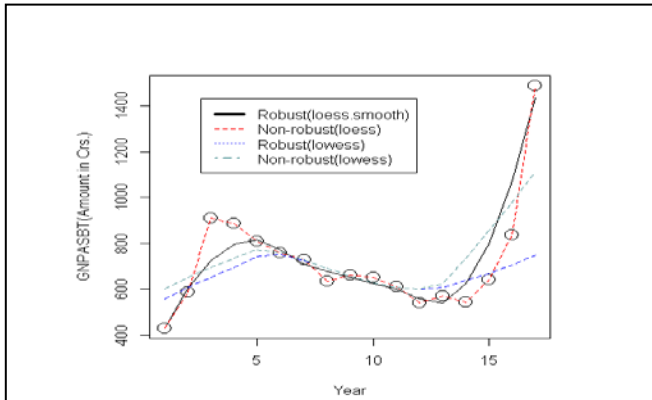


Fig.15. Robust LOESS fit on GNPA –SBT

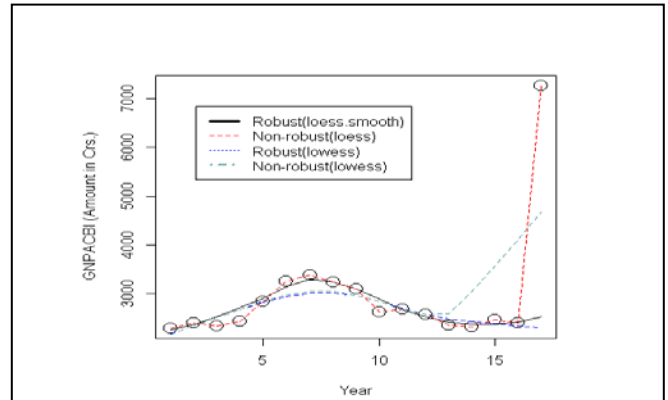


Fig.18. Robust LOESS fit on GNPA –CBI

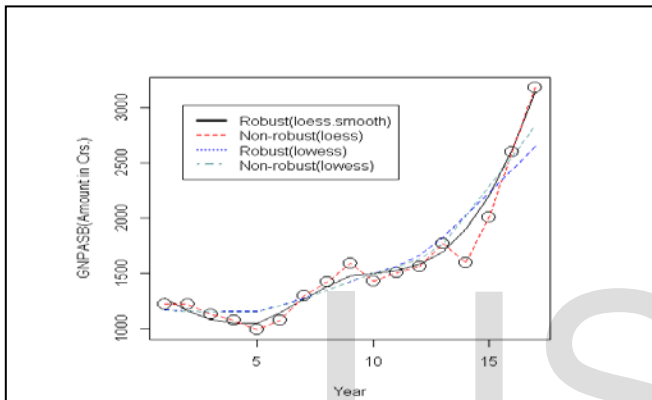


Fig.16. Robust LOESS fit on GNPA –SB

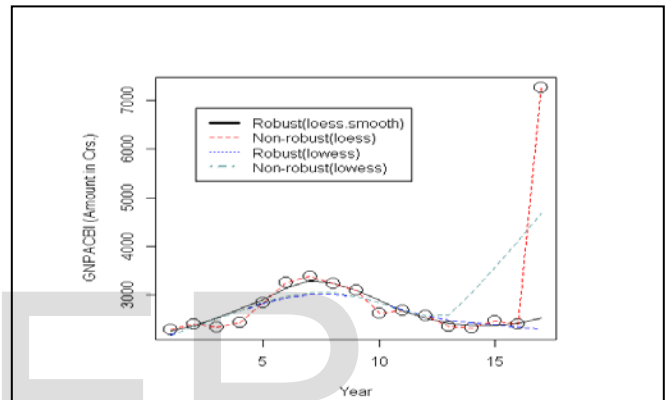


Fig.19. Robust LOESS fit on GNPA –PNB

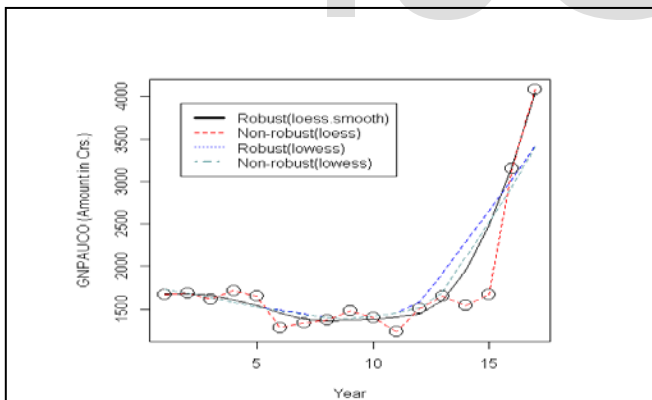


Fig.17. Robust LOESS fit on GNPA –UCO

Observations on the above non parametric models (LOESS Fit) given in Fig number 14 to 19 for six selected PSBs are given below:

1. The general diagonal orientations of the points suggest that GNPA values are highly correlated to Year for all the banks.
2. The curves also follow the central tendency of the Y variable's values across the range of the X variable which reveals the curvilinear nature of the relationship between GNPA's and Time (Year).
3. These curves were obtained without any prior specification about the functional form of the relationship. The sigmoid (i.e. S-like) shapes of all the curves are produced by the LOESS procedure. The slopes of the fitted curves are negative at some Years on the horizontal axis. In contrast, the slopes become very shallow near the right and left sides of the curve.
4. These curves clearly show that a linear model would provide a misleading depiction of the relationship between GNPA values and Time (Year).

2.3 Diagnostics – Residual vs Predictor Plot

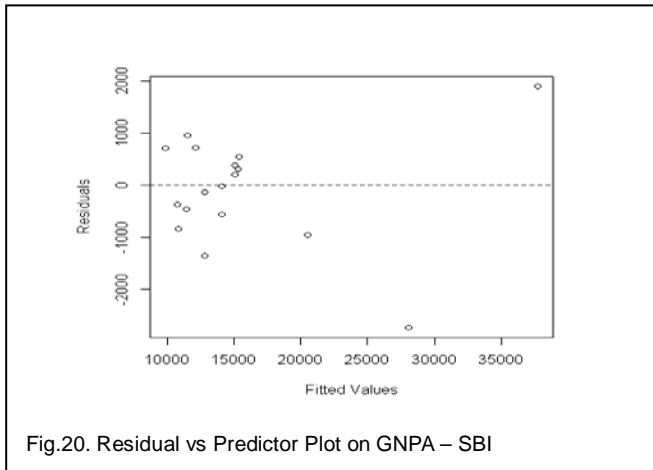


Fig.20. Residual vs Predictor Plot on GNPA – SBI

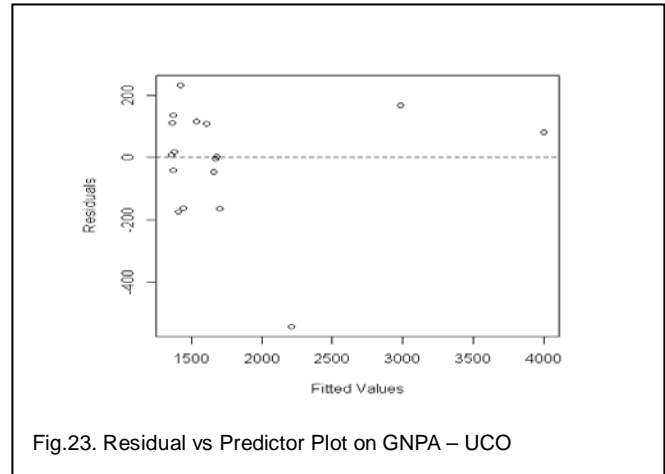


Fig.23. Residual vs Predictor Plot on GNPA – UCO

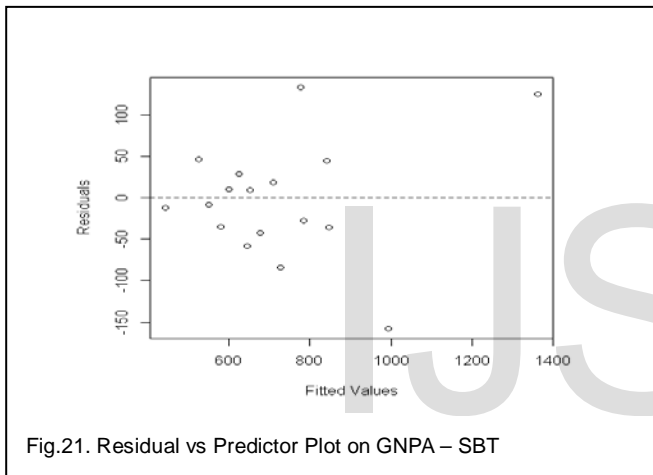


Fig.21. Residual vs Predictor Plot on GNPA – SBT

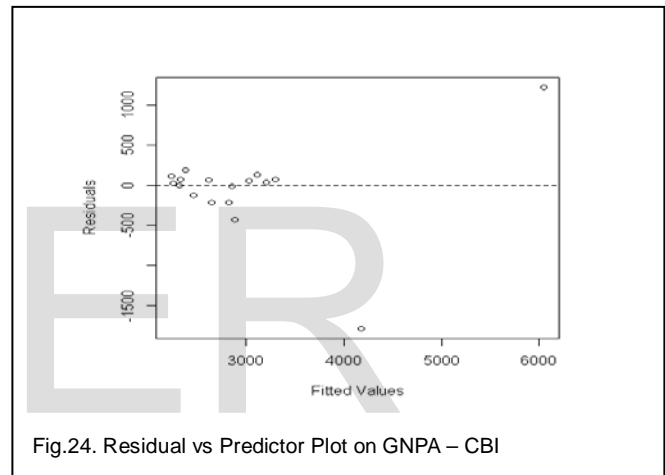


Fig.24. Residual vs Predictor Plot on GNPA – CBI

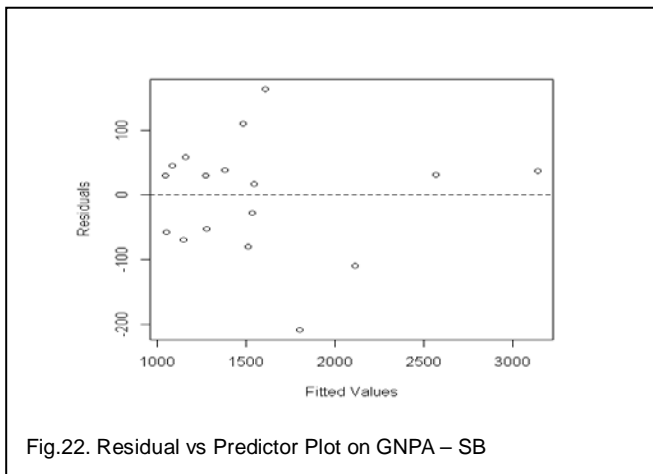


Fig.22. Residual vs Predictor Plot on GNPA – SB

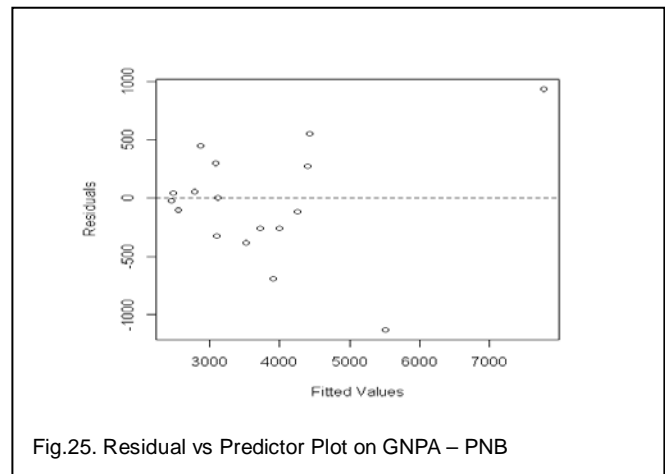


Fig.25. Residual vs Predictor Plot on GNPA – PNB

The dotted horizontal line is a visual baseline corresponding to residual value of zero. The above 'Residual vs Predictor' (Fig. 20 to 25) shows that the residuals appear randomly scattered around zero indicating that the penalized spline model describes the data well. It can also be stated that the model provide adequate representation of the GNPA bivariate dataset. Looking at the plots above we can say that for

our GNPA dataset, the plots exhibit the absence of relationship between residual and predicted value and therefore it can be concluded that penalized spline model exhibiting simple curvilinear relationship is appropriate for our GNPA dataset.

2.4 Diagnostics – Quantile Comparison plot of residuals

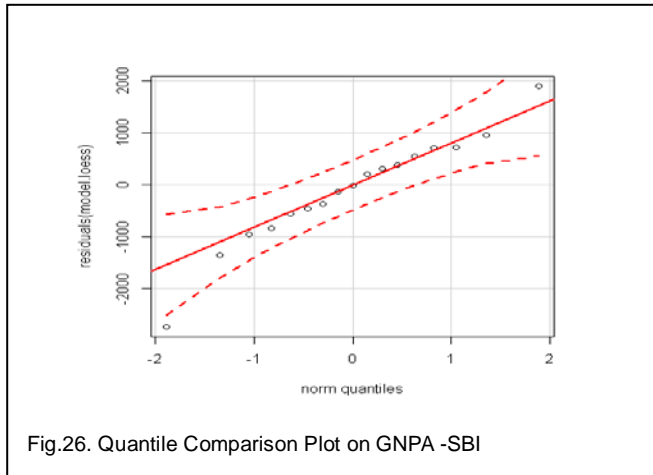


Fig.26. Quantile Comparison Plot on GNPA -SBI

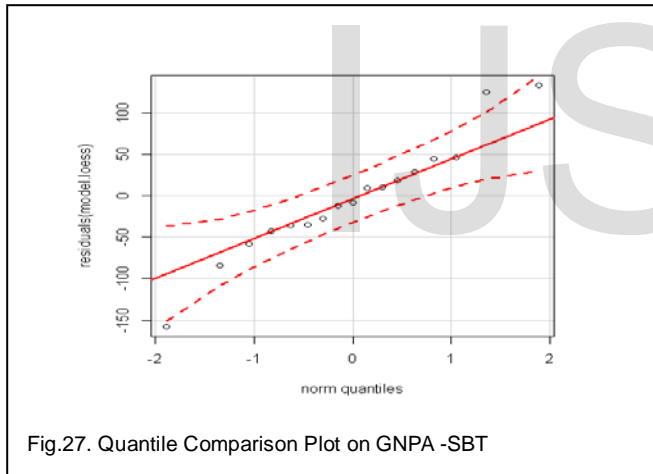


Fig.27. Quantile Comparison Plot on GNPA -SBT

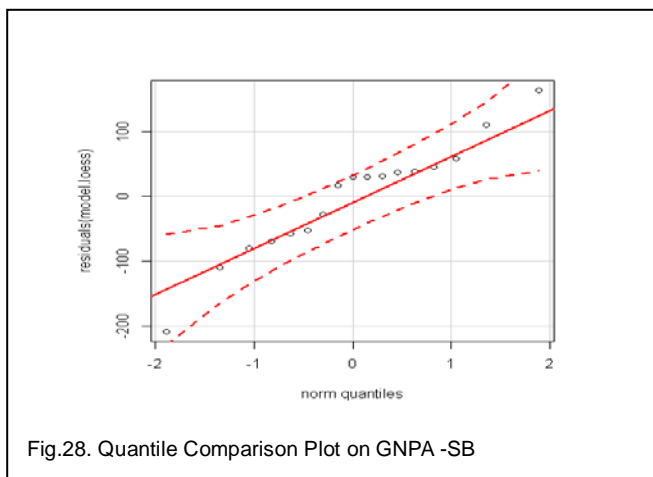


Fig.28. Quantile Comparison Plot on GNPA -SB

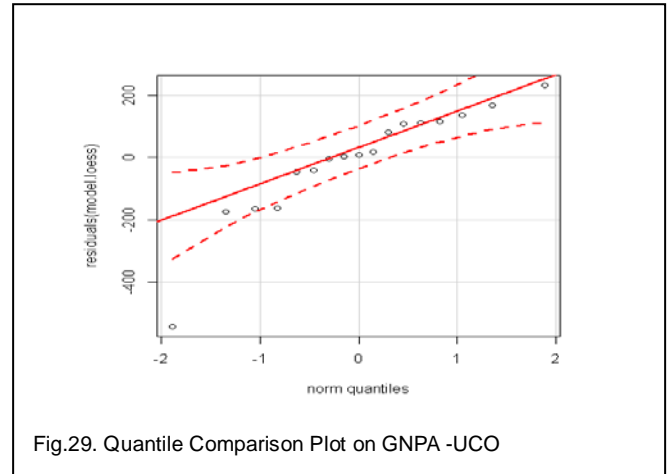


Fig.29. Quantile Comparison Plot on GNPA -UCO

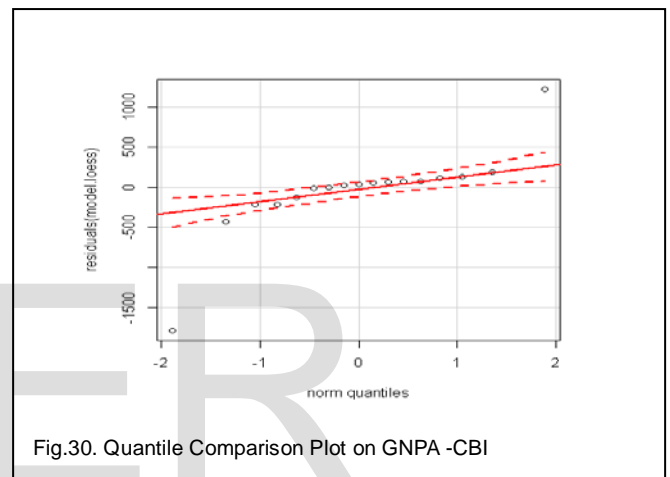


Fig.30. Quantile Comparison Plot on GNPA -CBI

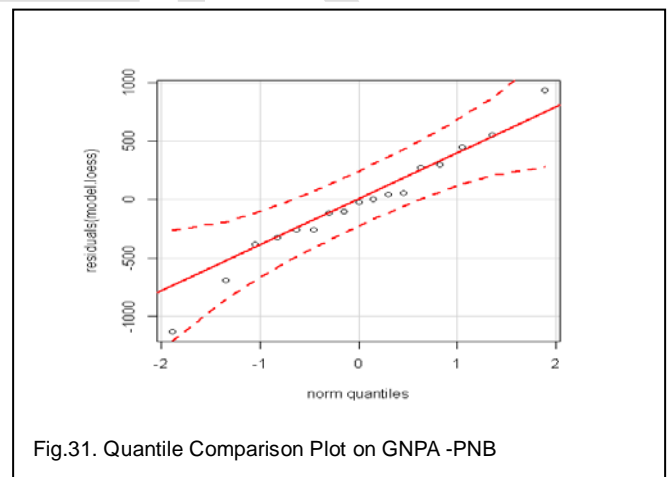
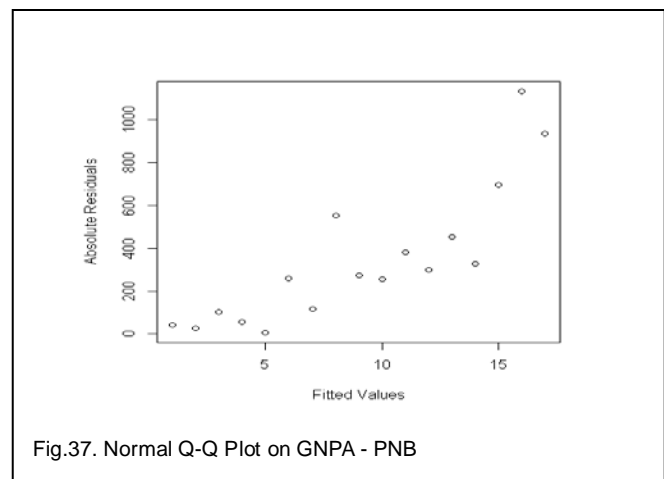
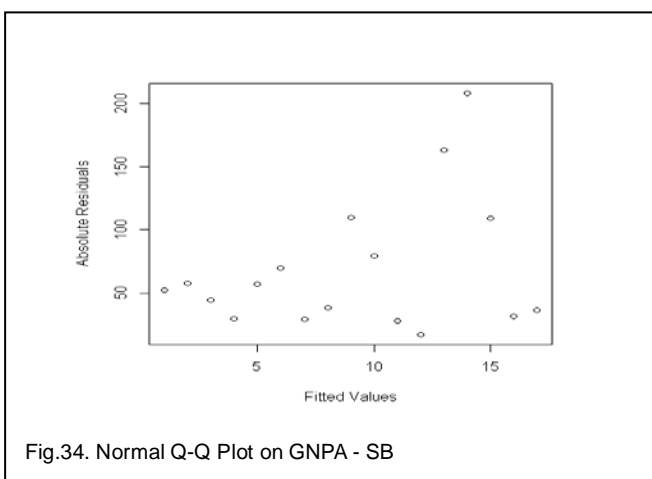
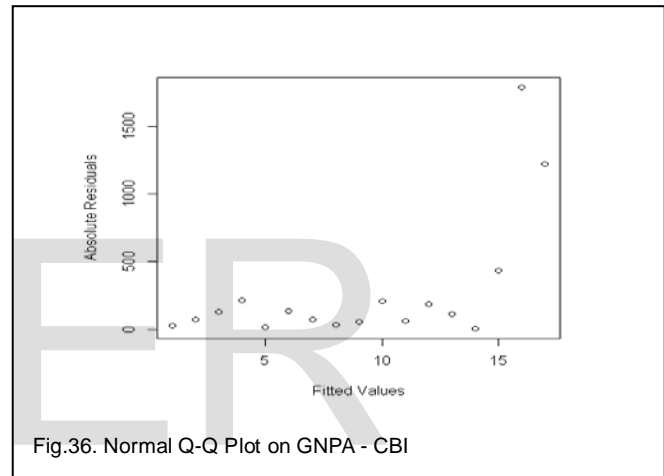
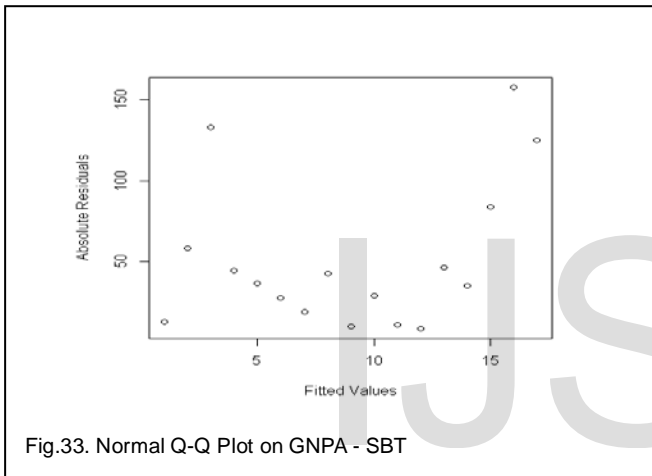
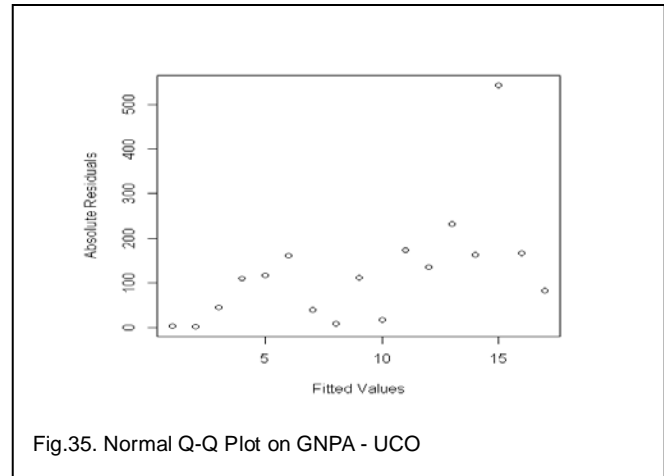
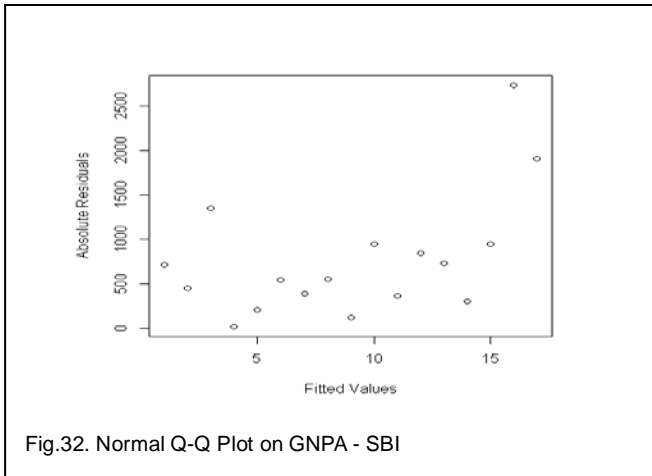


Fig.31. Quantile Comparison Plot on GNPA -PNB

From the above figures it is evident that in case of SBT and CBI outliers are present. Rests of the banks (SBI, SBT, SB and UCO) are free from outliers. The curves also confirm that GNPA dataset are approximately normal.

2.5 Diagnostics – Absolute vs Predictor plot



Above figures (Fig No 32 to 37) exhibit that none of the graphs, excepting for the models of SB, looks like megaphone shape. Hence the graphs show the justification of our Loess models for the given GNPA dataset.

2.6 Goodness of fit statistics

A summary of R^2_{LOESS} , Residual standard error, LOESS F-statistic and P-value for all the selected PSBs along with PSB in aggregate are given in the following table:

TABLE 2

SUMMARY STATISTICS OF PRECISION CRITERIA FOR SELECTED PSBS ON PARAMETER GNPA

Bank	R^2_{LOESS}	Residual Standard Error	LOESS F-statistic	P-value
SBI	0.97929	1296	84.26958	3.37E-08
SBT	0.91502	86.8	27.03662	1.21E-05
SB	0.97668	108.2	36.23729	2.77E-06
UCO	0.93869	223.8	31.79255	5.37E-06
CBI	0.7803	694.1	7.936263	0.002934
PNB	0.90009	595.1	16.55592	0.000126

From the above table it is apparent that R^2 value is very high (above 0.9) for all banks excepting CBI while it is moderately high for CBI. Hence we may say that the smooth curve summarizes nearly all of the total dispersion in the dependent variable.

The p-value of F test is less than 0.05 and therefore we may conclude that there is no linear relationship between response and predictor variables, which establishes the null hypothesis that there is no linear relationship between Non Performing Assets (response variable) and time (predictor variable).

2.7 Aspect on Forecasting

The main purpose of constructing a time series model is to forecast. Forecasting is a quantitative estimate about the likelihood of future events which is developed on the basis of current and past information. This information is embodied in the form of a model. By extrapolating models beyond the period over which they were estimated, one can make forecast about the future events. The non parametric models (smoothed by LOESS Fit) obtained based on our dataset from 1996 to 2012, have been extrapolated beyond the sample period to get the forecasted values for the years, 2013, 2014 and 2015 with respect to the parameters GNPA is given in the Table 3.

Forecasted Values of GNPA for three years (2013, 2014 and 2015) of all the selected PSBs exhibit an alarming phenomenon of continuous upward rise, which puts question mark on the overall management of credit portfolio of PSBs in India. Such a situation undoubtedly deserves immediate and serious attention on the part of the regulators to relook into the practices of credit appraisal and monitoring of credit in PSBs in India.

TABLE 3

FORECAST VALUE FOR PARAMETER GNPA

Bank	df	Residual scale	Forecasted value (Rs in Crores)		
			2013	2014	2015
SBI	11.74	1460.24	47769.9	60512.5	75269.8
SBT		107.457	1688.72	2163.68	2730.29
SB		121.977	3704.51	4436.93	5283.92
UCO		250.99	4960.88	6260.83	7780.24
CBI		765.946	7613.93	10026.4	12954.5
PNB		696.484	9864.09	12887.2	16519.7

3. Conclusion

The study on the aspects of movement of NPA offers a perpetual challenge to the bankers and researchers in general as forecasting future NPA is essential to monitor, manage and mitigate the deadly virus in banking system. In this paper, LOESS, a very popular non parametric technique, (which do not possess nice parametric interpretation, but offer more precise representation, when evaluated against different measure of closeness) has been fitted to GNPA data sets for six selected PSBs and it has been observed in respect of GNPA dataset, these non parametric models are very much relevant and useful as it enhances the precision level of the predicted values of the models in respect of parameter GNPA in six selected PSBs in India

The study also establishes that NPA in PSB in India can be adequately captured by non parametric regression (LOESS technique). In future the researchers can use this study as a reference to examine the trend of NPA for other PSBs and PrSBs and develop appropriate forecasting models and further compare these results among different group of banks for examining the relative strengths and weaknesses of different banks. Forecasted values of GNPA for three years (2013, 2014 and 2015) of all the selected PSBs clearly demonstrates future upward trend in respect of the financial parameter GNPA for all the selected PSBs which is a matter of great concern in a developing country like India.

REFERENCES

[1] Khasnobis, S. (2006) 'NPAs: Emerging Challenges'. The Indian Banker, 1(2): 17-20.
 [2] Sharma, M. (2005) 'Problem of NPAs and its Impact on Strategic Banking Variables'. Finance India, XI(3): 953-967.
 [3] Basu, P. (2005) 'How Distressed is India's Banking System'. In P. Basu (Ed.), India's Financial Sector, Recent Reforms, Future Challenges, (pp. 89 - 110). New Delhi, India: Macmillan India Ltd.
 [4] Casas, I. and Gao, J. (2007) 'Nonparametric Methods in Continuous Time Model Specification'. Econometric Reviews, 26(1), 91-106.

Available from:

<http://www.jitigao.com/uploads/3/1/8/5/3185830/paper07c.pdf>
[Accessed on 25th April, 2013].

[5] Wu, L. (2010) *Mixed Effects Models for Complex Data*. New York, USA: CRC Press.

[6] Cleveland, W.S. (1993) *Visualizing Data*. Hobert Press, Summit, NJ, USA.

[7] Jacoby, W.G. (2000) 'Loess: A Nonparametric, Graphical Tool for Depicting Relationships between Variables'. *Electoral Studies*, 19, 577-613. Available from:

<http://www.med.upenn.edu/beat/docs/Jacoby2000.pdf>
[Accessed on 22nd April, 2013].

[8] Jacoby, B. (n.d.) *Lecture 14: Non-normal Error Distribution and Nonlinear Functional Forms – Regression III: Advanced Methods*. Lecture Notes [pdf]. Michigan, USA: Michigan State University. Available from:

http://www.waikato.ac.nz/__data/assets/pdf_file/0017/51632/PA_Referencing_6th_ed.pdf [Accessed on 22nd April, 2013].

[9] Weisberg, S. (1996) 'The R-Code: A Graphical Paradigm for Regression Analysis'. In R. Stine & J. Fox (Eds.), *Statistical Computing Environments for Social Research*. Thousand Oaks, CA, USA: Sage Publications.

[10] McLeod, A.I. (n.d.) *Statistics 259b -Robust Loess: S lowess*. [pdf]. Ontario, Canada: University of Western Ontario. Available from: <http://www.stats.uwo.ca/faculty/aim/2004/04-259/notes/RL.pdf> [Accessed on 20th April, 2013].

[11] Dias, R. (n.d.) *Nonparametric Regression: Lowess/Loess*. [pdf]. Available from:
<http://www.ime.unicamp.br/~dias/loess.pdf> [Accessed on 25th April, 2013].

[12] Bhide M. G. Prasad A and Ghosh S (2002) 'Banking Sector Reform: A Critical Overview'. *Economic and Political Weekly*, (February).

[13] Ranjan, R., and Dhal, S.C. (2003) 'Non Performing Loans and Terms of Credit of Public Sector Banks in India: An Empirical Assessment'. *Reserve Bank of India Occasional Papers*, 24(3), 81-120. Available from:

<http://rbidocs.rbi.org.in/rdocs/Publications/PDFs/60613.pdf>
[Accessed on 5th September, 2011].

[14] Gopalakrishnan, T.V. (2006) *Management of Non Performing Asset: A Study with Reference to Public Sector Banks*. New Delhi, India: Northern Book Centre.

[15] Sethi, J., and Bhatia, N. (2007) *Elements of Banking and Insurance*. New Delhi, India: PHI Learning Pvt. Ltd.

[16] Rajeev, M., and Mahesh, H.P. (2010) 'Banking Sector Reforms and NPA: A Study of Indian Commercial Banks (252)'. *The Institute for Social and Economic Change, Bangalore, Working Paper*. Available from:

<http://www.isec.ac.in/WP%20252%20-%20Meenakshi%20Rajeev%20and%20H%20P%20Mahesh.pdf>
[Accessed on 12th October, 2012].

[17] Pradhan, T.K. (2012) 'A Study of Trends of Non Performing assets in Public Sector Banks during Post Reform Period'. *Research Journal of Economics & Business Studies*, 1(11), 30-37.

[18] Rajamohan, S. (2012) 'Analysis of NPAs of the Public Sector Banks in India'. *International Journal of Business and Management Tomorrow*, 2(9), 1-6.

[19] Rawlin, R., Sharan, M.S., and Lakhshmiopathy, B.P. (2012) 'Modeling the NPA of a Midsized Indian Nationalised Bank as a Function of Advances'. *European Journal of Business and Management*, 4(5), 78-88.

[20] Veerakumar, K. (2012) 'Non Performing Assets in Priority Sector: A Threat to Indian Scheduled Commercial Banks'. *International Research Journal of Finance and Economics*, 93, 6-23.

[21] Bandyopadhyay, G. (2013) 'Modeling NPA Time Series Data in Selected Public Sector Banks in India with Semi Parametric Approach'. *International Journal of Scientific & Engineering Research*, 4(12), 1876 - 1889.

[22] India. Reserve Bank of India (1991) *Report of the Committee on the Financial System*. Mumbai, India: Reserve Bank of India.

[23] India. Government of India (1998) *Report of the Committee on Banking Sector Reforms*. New Delhi, India: Government of India.