# Factors Leading to Non - Performing Assets (NPAs) : An Empirical Study

** P. K. Viswanathan*
*** M. Muthuraj*

## Abstract

The performance of the banking industry is one of the main indicators of economic growth. It plays a vital role in various socioeconomic activities. A strong banking sector is essential for a robust economy. The poor performance of the banking sector in terms of financial risk management may adversely impact the other sectors of the economy. In India, non - performing asset (NPA) is a key factor that enhances the credit risk substantially for any bank. The performance of the public sector banks in risk management in the recent past years has been declining in view of NPAs. The ability of the banks to identify defaulters before lending is paramount for minimizing the incidence of NPAs as well as developing effective mechanism to proactively deal with potential defaulters. Various financial indicators such as quick ratio, profit after tax (PAT) as percentage of net worth, total net worth, and cash profit as percentage of total income will enable the concerned authority to spot possible defaulters and take appropriate corrective measures. With this background, an attempt was made in this paper to study key factors leading to non - performing assets. This research study focused on how the key factors impact NPAs based on insights derived from three important classifications and predictive models namely random forest (RF), gradient boosting machine (GBM), and logistic regression. The findings of this study will pave way for policy makers in banks to assess the probability of borrowers repaying the loan and classify them as good credit or bad credit.

*Keywords:* RF, GBM, NPA, quick ratio, PAT, net worth

JEL Classification: G00, G210, M210, O160, P170

Paper Submission Date : April 18, 2018 ; Paper sent back for Revision : December 16, 2018 ; Paper Acceptance Date : December 22, 2018

The Indian economy has been growing in a steady manner during the last decade, and the banks play a crucial role in this context. The current situation shows that almost every bank in India suffers from a "bad-loan" problem. The loan amount that is borrowed by a customer and is not returned within the stipulated time horizon becomes a non-performing asset (NPA). In recent years, India is one of the highest in NPAs in South Asia with a figure in a vicinity of 2% of the total loan amount as NPAs.

The borrowers can be dichotomized into - bankrupt and willful defaulters. In recent times, there has been a significant growth in both these categories. This research paper attempts to provide key models for prediction of NPAs based on important variables. The paper also quantifies the relative importance of predictor variables that are responsible for the occurrence of NPAs.

* *Professor,* Great Lakes Institute of Management, Dr. Bala V. Balachandar Campus, East Coast Road, Manamai Village, Thirukazhukundram Taluk, Kancheepuram District - 603102, Tamil Nadu. E-mail : viswanathan.pk@greatlakes.edu.in
***Assistant Professor,* Great Lakes Institute of Management, Dr. Bala V. Balachandar Campus , East Coast Road, Manamai Village, ThirukazhukundramTaluk, Kancheepuram District - 603 102, Tamil Nadu.
E-mail : Muthuraj.m@greatlakes.edu.in

## Literature Review

The studies in the past have focused on how non-performing assets have been increasing over time in India due to important macro and microeconomic factors.

Arora and Singh (2015) pointed out that the banking institutions play a vital role in the economic development of a country. There are certain issues of concern that affect their performance and efficiency. The credit disbursed under this pattern increases the risk for the banks. The main idea of this study was to compare the public and private sector commercial banks on the basis of NPAs under the SHG bank linkage program. This paper also tried to explain the region wise differences in the amount of NPAs of the public sector banks.

Digal, Satpathy, and Behera (2015) showed that the rise in lending rate significantly reduced NPAs of private sector banks in India. It was pointed out that there is a tendency among borrowers to take lower quantum of loans if the interest rates are pretty high. Their research also expressed that high operational efficiency results in lower NPAs. According to the study, fiscal deficit, growth in GDP, and an increase in balance of trade help in bringing down the NPA levels of banks, while inflation causes increase in the NPA levels.

Polodoo, Seetanah, Sannassee, Seetah, and Padachi (2015) studied the Mauritian banks and deduced that the credit concentration ratios, bank size, exchange rate, and lagged loans significantly impacted NPA.

By focusing on the incidence of NPAs in the aviation, power, and telecom sectors, Khedekar (2012) recommended that banks should exercise caution in the selection of borrowers or projects by carefully scrutinizing the financial statements. Effective and regular follow up of the end use of the fund sanctioned is imperative for ascertaining any possible diversion of funds.

Ghosh's (2005) study on the relation of leverage ratio on NPAs confirmed that the financial condition of the corporate sector can provide useful leads on banks' asset quality. It also suggested that rapid expansion of lending by banks often worsened asset quality and adversely impacted their capital position.

Singh (2016) focused on the impact of NPAs on the commercial banks in India. The study also threw light on the recovery of NPAs through various channels present in the market. The study further explained the appropriate measures that should be taken to reduce the future NPAs and the methods to manage the existing NPAs in the banks.

Mahboob (2013) showed the methods to compute and compare the financial performance of SBI and HDFC through profitability ratios. He used ratios that included adjusted cash margin, net profit margin, return on net worth (RONW), and adjusted return on net worth to arrive at the results.

Pašić and Omerbegović - Arapović (2016) discussed various repayment policies of loan repayment in Bosnia and Herzegovina so that non-performing assets of the banks could be minimized.

## Past Studies and their Relevance to the New Model

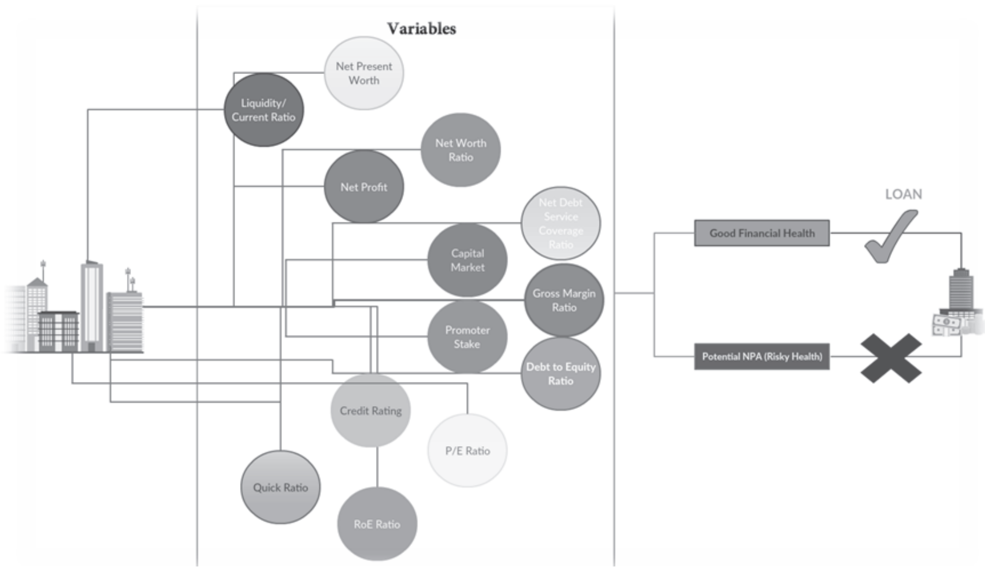
We find that the studies in the past focused mainly on determining the macroeconomic and microeconomic factors that resulted in non-performing assets for banks. In the past, factors like rise in lending rates, credit policy, credit concentration ratios, bank size, and lagged loans significantly impacted NPAs. Other factors that focused on reasons for generating NPAs were financial ratios such as leverage ratio, P/E, total debt to equity ratio, and gross margin ratio.

We thus are taking the earlier analysis from different researchers one step forward for predicting the increasing liability/financial health of the company in the future. We intend to do this with the help of various predetermined factors. We believe that the proposed model would help banks assess the financial health of a company and then decide whether to lend to a company or not. By lending the right amount to the right company, we believe that banks will subsequently decrease their accumulation of non-performing assets. Based on the past studies, we believe that this model will be a very useful tool for determining the parameters that are of paramount importance while disbursing loans to right customers.

## Conceptual Model

The various factors identified from literature review are used to build the conceptual model displayed in the Figure 1. The model attempts to show how banks can determine whether to disburse loan to a customer based on key variables (displayed in the middle box). If the customer meets certain parameters, then the bank disburses the loan. In this model, we also examine the key financial ratios to understand the financial health of the lender. If the customer's financial health is below a threshold score, then we can advise the banks to refrain from providing loans to such customers since it may result in potential non-performing assets.

**Figure 1. Model 1 : Measuring Customers' Eligibility for the Sanction of Loans Without Significant Threat of Causing NPAs for Banks**




## Data and Methodology

The models used for classification are logistic regression, gradient boost machine (GBM), and random forest. While logistic regression is the widely used method under-supervised learning, gradient boost method and random forest fall under the purview of recursive partition algorithms.

Logistic regression is a variation of ordinary regression in which the dependent variable is binary in the sense it takes only two values 0 or 1. The dependent variable is categorical that usually represents the occurrence or non-occurrence of an event and the independent variables can be continuous, categorical, or both. Logistic regression has been widely used in the financial service industry for credit scoring models. On theoretical grounds, logistic regression is a more appropriate statistical tool than linear regression. Ordinary least square regression (OLS) cannot guarantee estimated probability to be between 0 and 1. On the contrary, logistic regression will ensure the estimated probability to fall in the range of 0 to 1 because it is based on a sigmoid function. In logistic regression, the individual parameters can be tested for statistical significance.

Recursive partitioning algorithm (RPA) is associated with the term decision tree for classification and prediction. Recursive partitioning creates a decision tree that aims to correctly classify members of the population based on several dichotomous dependent variables. It is a learning system that creates classification matrix displayed in the decision tree format. If we have data divided into classes, for example, high-risk loan and low risk loan, recursive partitioning algorithm can be used for obtaining the solution with good accuracy.

Gradient boosting is based on the idea of whether a weak learner can be modified to become stronger. GBM is built around a forward stage-wise additive model by implementing gradient descent in function space. Similar to gradient descent in parameter space, at the *K*th iteration, the direction of the steepest descent is provided by the negative gradient of the loss function. Taking a step along this direction ensures reduction in loss. At each iteration, a regression tree is fitted to predict the negative gradient.

Random forest is an ensemble learning method for classification, regression, and other tasks. It operates by constructing a multitude of decision trees using training data. The output is based on the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random forest is well known for giving high prediction accuracy, though it can over fit sometimes.

Prowess for interactive querying (ProwessIQ) is an Internet based application for querying database on performances of listed and unlisted companies. It provides charting tools and formatted reports on performance of companies. These reports provide financial information, including analytical ration and benchmark comparisons. It is a user friendly quality product which provides reliable data.

We obtained the financial data such as financial market/equity market, promoter stake, net debt service coverage ratio, credit rating, price to earnings ratio/ P/E growth ratio, quick ratio, return on equity ratio, total debt to equity ratio, long term debt to equity ratio, short term debt to equity ratio, gross margin ratio (gross profit/revenue) for different customers to study the trend. The data set contains a little over 3500 (3545) data points spanning over a time horizon of 10 years (2007-2016). These input variables are financial ratios of organizations that include defaulters (bad credit) and non-defaulters (good credit) in an adequate manner. Hence, the number of records which is over 3500 is not only robust, but is also considered more than adequate for building predictive models in the context of NPA.

We started our analysis based on the net worth of a company. The default classification *(D)* is based on the net worth value. *D* will have a value of 1 if the predicted net worth in the next year is less than zero, else it will have a value of 0. The data set has been partitioned with 70% as training data and the remaining 30% as test data.

As mentioned earlier, the data set has been partitioned into the development set (training data) and the test set (test data). The development set will contain 70% of the data, and the test set will contain the remaining 30% of the data. After constructing the model from the development set, we have validated the efficacy of the model using the test data. We have identified the number of defaulters that the model predicts and the actual number of defaulters and then based on this, we calculate the percentage accuracy of our model.

## Analysis and Discussion

**(1) Random Forest :** The following are the variables used in this model : (a) total asset, (b) total income, (c) profit after tax, (d) PBDITA, (e) PBT, (f) cash profit, (g) PBT as of income, (h) PAT as of income, (i) sales, (j) quick ratio times, (k) current ratio, (l) debt to equity ratio times, (m) PAT as of net worth, (n) total net worth, (o) cash profit as of income.

On performing the random forest analysis, the following output results. This throws light on the relative importance of the variables used in understanding NPAs given in the Table 1.

**Table 1. Results of Random Forest Analysis**

| Variables | Valuable Importance (Type 0 Data) | Valuable Importance (Type 1 Data) | Mean Decrease Gini | % Relative Importance |
|---|---|---|---|---|
| Total Asset | 3.9595957 | 0.1528561 | 11.818460 | 5.42 |
| Total Income | 4.2949809 | 0.6073147 | 5.910829 | 2.71 |
| Profit After Tax | 4.0226628 | 1.4616979 | 10.420069 | 4.78 |
| PBDITA | 4.8186020 | 0.6380102 | 10.094290 | 4.63 |
| PBT | 4.0717218 | 4.6792176 | 15.104788 | 6.93 |
| Cash Profit | 4.2151050 | 3.8224476 | 12.166796 | 5.58 |
| PBT as of Income | 4.8216746 | 3.4649768 | **17.720258** | **8.12** |
| PAT as of Income | 3.2733160 | 0.7989783 | 9.529840 | 4.37 |
| Sales | 4.0893178 | -1.2701403 | 5.969670 | 2.74 |
| Quick Ratio Times | 2.1255994 | 2.4373036 | 9.692694 | 4.44 |
| Current Ratio | 0.2800468 | 2.8335087 | 12.410237 | 5.69 |
| Debt to Equity Ratio Times | 4.3265571 | 7.2092004 | **38.853204** | **17.81** |
| PAT as of Net Worth | 4.4683011 | 3.2765370 | **23.292577** | **10.68** |
| Total Net Worth | 4.9908778 | 5.9561318 | **23.357281** | **10.71** |
| Cash Profit as of Income | 3.5733527 | 1.3167958 | 11.770651 | 5.40 |

The results obtained from random forest analysis point out debt to equity, PAT as of net worth, total net worth, and PBT as of income are the most important ratios that separate the group 0 from group 1 (Table 1). Mean decrease in Gini index that is part of the random forest machine learning technique indicates in some way the degree of purity in the overall assessment of the model. Unlike a traditional model such as logistic regression, in random forest, one cannot precisely say that one unit increase in the independent variable impacts how much increase/decrease in the dependent variable. One way to quantify the relative importance in the random forest method is to normalize mean decrease in Gini in terms of percentages so that relative importance can be measured. Debt to equity, PAT as of net worth, total net worth, and PBT as of income are relatively more important in separating group 0 from group 1 (see figures given in bold under mean decrease as well as percent relative importance in Table 1). We further validate the model by taking four variables randomly at a time, for which mean decrease Gini is highest and the results are given in the Table 2.

**Table 2. Results of Mean Decrease**

| Variables | Valuable Importance (Type 0 data) | Valuable Importance (Type 1 data) | Mean Decrease | |
|---|---|---|---|---|
| | | | Accuracy | Gini |
| Quick Ratio | 1.179363 | 5.367426 | 4.131685 | 37.33661 |
| PAT as of Net Worth | 9.247413 | 15.422414 | 14.458266 | 94.31687 |
| Total Net worth | 8.556821 | 9.052154 | 13.632386 | 59.08499 |
| Cash Profit as of Income | 4.412182 | 9.571320 | 10.149944 | 32.39823 |

It is inferred from the Table 2 that the four variables (quick ratio, PAT as of net worth, total net worth, and cash profit as of income) have been obtained by performing many iterations for getting the highest mean decrease Gini values.

**Table 3. Confusion Matrix**

| | Valuable Importance (Type 0 data) | Valuable Importance (Type 1 data) | Class Error |
|---|---|---|---|
| Valuable Importance (Type 0 data) | 1704 | 31 | 0.01786744 |
| Valuable Importance (Type 1 data) | 60 | 61 | 0.49586777 |

**Table 4 . Results of GBM Method**

| | Variables | % Relative Influence |
|---|---|---|
| PATNW | PAT as of Net Worth | 53.021132 |
| DE | Debt to Equity Ratio Times | 22.589526 |
| TNW | Total Net worth | 11.046061 |
| PBDITA | PBDITA | 7.843254 |
| CR | Current Ratio | 5.500027 |
| TA | Total Asset | 0.000000 |
| TI | Total Income | 0.000000 |
| PAT | Profit After Tax | 0.000000 |
| PBT | PBT | 0.000000 |
| CPFT | Cash Profit | 0.000000 |
| PBTI | PBT as of Income | 0.000000 |
| PATI | PAT as of Income | 0.000000 |
| SALES | Sales | 0.000000 |
| QR | Quick Ratio Times | 0.000000 |
| CPFTI | Cash Profit as of Income | 0.000000 |

The overall accuracy of the model is 95.1%. It can be observed from the Table 3 that the model is able to predict almost 99% cases correctly for non-defaulting cases, while for defaulting cases, the prediction is almost 50%, which can be further improved by looking at the cut off values critically.

**(2) GBM Method:** As we can see from the Table 4, beyond CR, nothing impacts. So, we have considered the following variables. The visual (Figure 2) succinctly and profoundly captures the essence of the key variables that influence NPA.

**(3) Logistic Regression :** We performed logistic regression using SPSS on the development data and obtained the following results.

From the Table 6, we can infer that the overall accuracy of the classification is 94.2%. We have got an accuracy of 65.3% for predicting defaulted case given default. Non default has an accuracy of 96.2%. The odds ratio given by logistic regression is dominated by total net worth in terms of default prediction. All the predictor variables, that is, quick ratio times, PAT as of net worth, total net worth, and cash profit as of total income are statistically overwhelmingly significant at the 5% level (see Table 7).

It is seen from the Table 5 that the Nagelkerke $R$ square implies that 43.7% of the uncertainty produced by the intercept only model (intercept only model means only the intercept exists and the rest of the variables are not there in the model) is explained by the full model.

**Figure 2. Relative Influence of Variables**

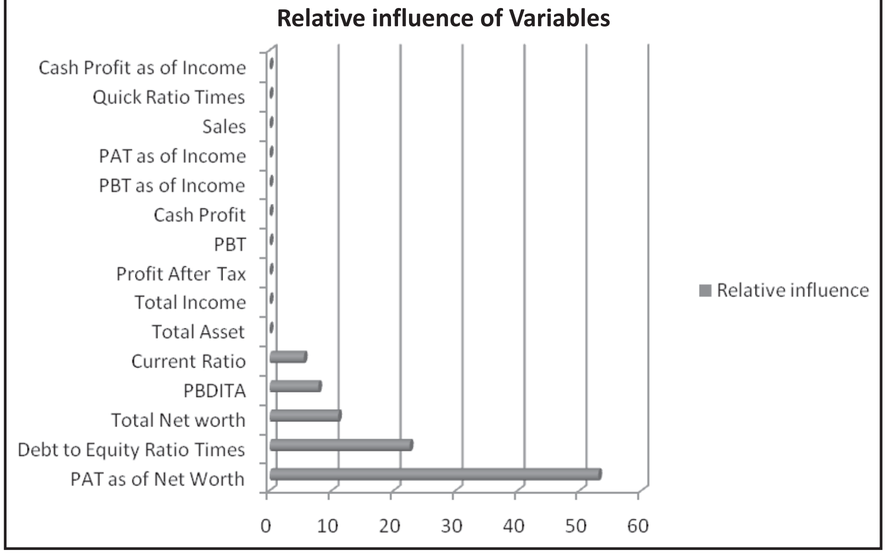



**Table 5. Model Summary**

| Step | -2 Log Likelihood | Cox & Snell *R* Square | Nagelkerke *R* Square |
|---|---|---|---|
| 1 | 762.941[a] | .169 | .437 |

Note. [a] Estimation terminated at iteration number 7 because parameter estimates changed by less than 0.001.

**Table 6. Classification Table[a]**

| Observed | | | Predicted | | |
|---|---|---|---|---|---|
| | | | *D* | | Percentage Correct |
| | | | 0 | 1 | |
| Step 1 | *D* | 0 | 2255 | 88 | 96.2 |
| | | 1 | 58 | 109 | 65.3 |
| | Overall Percentage | | | 94.2 | |

Note. [a] The cut value is .150.

**Table 7. Classification Table**

| | *B* | S.E | Wald | *df* | Sig. | Exp (B) |
|---|---|---|---|---|---|---|
| Step 1 [a] quick ratio times | -.317 | .158 | 4.038 | 1 | .044 | .728 |
| PATas net worth | -.030 | .003 | 86.297 | 1 | .000 | .971 |
| TOLTNW | .210 | .019 | 116.713 | 1 | .000 | 1.234 |
| Cash Profit as of Total Income | -.013 | .004 | 10.853 | 1 | .001 | .987 |
| Constant | -3.373 | .185 | 330.970 | 1 | .0001 | .034 |

Note. [a] Variable (s) entered on step 1: Quick ratio times, PAT as net worth, TOLTNW, Cash profit as of total income.

In summary, we can say that four variables namely quick ratio times, PAT as of net worth, TOLTNW, and cash profit as of total income are important predictors of group classifications.

## Major Findings

**(1)** Debt to equity, PAT as of net worth, total net worth, and PBT as of income are relatively more important in separating group 0 from group 1.

**(2)** The four variables (quick ratio, PAT as of net worth, total net worth, and cash profit as of income) are found to be important predictor variables in a steady-state level after several iterations resulting in the highest mean decrease Gini values.

**(3)** Further, the results are robust when the model is validated with the test data.

**(4)** The results of random forest analysis have almost perfect linear separation as mean decrease Gini value is highest for the key ratios mentioned above.

**(5)** The findings succinctly and profoundly capture the essence of the key variables that influence NPA.

**(6)** Logistic regression on the training data points out that the overall accuracy of the classification is 94.2%.

**(7)** The four variables namely quick ratio, PAT as of net worth, TOLTNW, and cash profit as of total income are important predictors of group classifications in the logistic model.

## Implications

From the point of view of policy makers in banks, an alert system can be developed as follows :

**(1)** If the predicted health falls below the cutoff score based on these financial ratios, then it would be wise for the banks to refrain from providing loans to such institutions.

**(2)** The policy makers can avoid the probability of default incidence that can trigger a possible creation of non-performing assets beforehand.

**(3)** Prevention is better than cure implies here, which means that policy inputs will prevent NPAs to occur in the first place rather than treating it later when it may be too late.

**(4)** The policy makers in banks can spot the potential defaulters using the models discussed in this study and can refuse to sanction loan because there is a classifier to separate good credit from non-bad credit.

## Limitations of the Study and the Way Forward

The following are the limitations of the study :

**(1)** The study does not examine how a macroeconomic factor like interest rate affects NPAs.

**(2)** The scope of the study is limited only to the selected indicators, for example, in our data, we have taken all the companies which gave a zero or negative revenue in a year.

**(3)** The study is based on secondary data culled out from RBI and other reports. These are based on historical data which does not capture the essence of inflation.

As an extension to this study, a neural network model that is considered a universal approximation and is non-parametric in nature can be applied for improving predictive accuracy with support of vector machines (SVM).


## Acknowledgment

We wish to acknowledge the efforts of three graduate students of Great Lakes Institute of Management - Pallav Abhishek, Sushrit Anand, and Dinesh Singh, who helped us in field work and data collection and without whose contribution, this paper would not have been possible.


## References


Arora, M., & Singh, S. (2015). An evaluation of the non-performing assets of public and private sector banks under the SHG bank linkage programme. *Indian Journal of Finance, 9* (6), 41-50. doi:10.17010/ijf/2015/v9i6/71161

Digal, S. K., Satpathy, A., & Behera, S. (2015). Macroeconomic factors affecting the NPAs in the Indian banking system: An empirical assessment. *The IUP Journal of Bank Management, 14* (1), 57 - 74.

Ghosh, S. (2005). Does leverage influence banks' non-performing loans: Evidence from India. *Applied Economics, 12* (15), 913 - 918.

Khedekar, P. S. (2012). Performance of non-performing assets in India. *Aweshkar Research Journal, 13* (1), p. 136.

Mahboob, A. (2013). A comparative study on profitability of SBI and HDFC. *International Journal of Management Science and Technology, 3* (6), 1 - 5.

Pašić, S., & Omerbegović - Arapović, A. (2016). The influence of macroeconomic trends on the repayment of loans by households : Evidence from the Federation of Bosnia and Herzegovina and policy recommendations. *South East Journal of Economics & Business, 11* (1), 76 - 87.

Polodoo, V., Seetanah, B., Sannassee, R.V., Seetah, K., & Padachi, K. (2015). An econometric analysis regarding the path of non-performing loans - A panel data analysis of Mauritian banks and implications for the banking industry. *The Journal of Developing Areas, 49* (1), 53 - 64.

Singh, V. R. (2016). A study of non-performing assets of commercial banks and it's recovery in India. *Annual Research Journal of Symbiosis Centre for Management Studies, 4,* 110 - 125.

## About the Authors

**Dr. P. K. Viswanathan is currently a Professor in the area of operations and analytics at Great Lakes Institute of Management, Chennai. Apart from executing corporate consultancy assignments, he has designed and conducted training programs for many leading organizations in India. After completing his Masters from Madras University, he went on to do an MBA from FMS, Delhi and MS from Manitoba, Canada. He holds a Ph.D. from Madras University.**

**Dr. M. Muthuraj is working as an Assistant Professor at Great Lakes Institute of Management, Chennai since June 2009. His research areas include agricultural economics, environmental economics, public finance, and rural developmental economics. He has a work experience of about five years in micro economics, macro economics, and business economics. He was previously associated with the Loyola College as Lecturer and University of Madras as Lecturer cum Research Officer in the Department of Economics and Agro Economic Research Centre (AERC) from July 2006 - June 2007. During his tenure at the AERC, he co-authored the final reports of two research projects entitled "Factors Affecting Fertilizer Consumption in Tamil Nadu" and "Estimation of Seed, Feed and Wastage Ratio for Major Food Grains in Tamil Nadu".**