Full reference: Abdou, H. A., Mitra, S., Fry, J. & Elamer, A. A. (2019) Would two-stage scoring models alleviate bank exposure to bad debt?, *Expert Systems with Applications*. Accepted 15th March, 2019. https://doi.org/10.1016/j.eswa.2019.03.028.

Would two-stage scoring models alleviate bank exposure to bad debt?

Hussein A. Abdou*
Lancashire School of Business and Enterprise, University of Central Lancashire, Preston, PR1 2HE, UK
Email: habdou@uclan.ac.uk

Shatarupa Mitra
The University of Salford Business School, University of Salford, Greater Manchester, M5 4WT, UK
Email: rupamitra@ymail.com

John Fry
School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, Manchester, M15 6BH, UK,
Email: J.Fry@mmu.ac.uk

Ahmed Ahmed Elamer
School of Management, University of Bradford, Bradford, BD9 4JL, UK
Email: A.Elamer@bradford.ac.uk

Acknowledgments
The Authors would like to thank and acknowledge the Editor and the anonymous reviewers for their helpful suggestions and comments. We also would like to thank John English, Mohammad Zayed and Matt Burke for their useful comments on previous drafts of this paper.

***Corresponding Author**

Would two-stage scoring models alleviate bank exposure to bad debt?

Abstract
The main aim of this paper is to investigate how far applying suitably conceived and designed credit scoring models can properly account for the incidence of default and help improve the decision-making process. Four statistical modelling techniques, namely, discriminant analysis, logistic regression, multi-layer feed-forward neural network and probabilistic neural network are used in building credit scoring models for the Indian banking sector. Notably *actual* misclassification costs are analysed in preference to estimated misclassification costs. Our first-stage scoring models show that sophisticated credit scoring models, in particular probabilistic neural networks, can help to strengthen the decision-making processes by reducing default rates by over 14%. The second-stage of our analysis focuses upon the default cases and substantiates the significance of the timing of default. Moreover, our results reveal that State of residence, equated monthly instalment, net annual income, marital status and loan amount, are the most important predictive variables. The practical implications of this study are that our scoring models could help banks avoid high default rates, rising bad debts, shrinking cash flows and punitive cost-cutting measures.

Keywords: Credit; Indian Banks; Neural Networks; Actual Misclassification Costs, Timing of Default

1. Introduction

At a time when even the largest banks are not immune to distress, credit decision-making is crucially important. The Reserve Bank of India (RBI) and the Finance Ministry has thus far externally controlled and regulated the banking sector. Deregulation and the decoupling of state control pose new challenges, and intense competition is placing the survival of all but the fittest and the most efficient in doubt. Commercial banks are accordingly striving to adjust to a new economic and technological environment. Sound credit scoring models form an integral part of this adjustment process. This motivates our present purpose which is to propose suitably conceived and designed credit scoring models for personal loans with due allowance for the incidence of default.

The novel contribution of the present paper consists in integrating two stages of the decision process with reference to the Indian banking sector. Firstly, we build credit scoring models for our unique sample of personal loans, provided by one of the largest Indian banks. The sample includes a significant number of bad debts that is consonant with the current and evolving profile of personal indebtedness. Secondly, we explore in detail the characteristics of the defaulters in our sample. This feature is particularly important given the recent history of rising bad debt. In both stages, we identify the key predictor variables to be used in building models. Further, we evaluate our models by using *actual* misclassification costs.

FIGURE (1) ABOUT HERE

The sharp increase in household leverage ratios in recent years shown in Figure 1a (Leverage Ratios in India) portrays the increase in borrowers' vulnerability. Figure 1b (Growth of Personal Loans and Housing Loans) shows the muted growth of personal loans over recent years up to the end of 2010. However, the year ended March 2011 saw the increase of 17% portrayed in Table 1, against only 4.12% in the year ended March 2010. The rate slightly decreases in the next two years, 2012 and 2013, which is commensurate with the increase of non-performing assets reported on Indian banks' balance sheets (Financial Times, 2011). It should be emphasised that at the end of March 2014 retail credit has increased driven primarily by housing loans, personal loans and auto loans representing 47%, 36% and 14%, of gross credit respectively (RBI, 2014).

TABLE (1) ABOUT HERE

Indian market credit bureaux, for example Credit Information Bureau India Limited (CIBIL, 2016), collect credit data for the banking industry. CIBIL maintains a repository of the credit history of all commercial and consumer borrowers in the country and it provides information to any bank to facilitate their credit granting decisions. CIBIL's Consumer Credit Bureau deals with the credit history of individual customers while the Commercial Credit Bureau maintains the credit history of non-individual clients such as corporates. CIBIL provides credit information as distinct from opinions and does not classify any client's loan as being in default unless the lender has already classified it as such.

While many research papers have discussed credit scoring models for developed countries (Marshall, *et al.,* 2010; Akkoc, 2012; Brown & Mues, 2012; Tong, *et al.,* 2012; Majeske and Lauer, 2013; Ono, *et al.,* 2014; Leow and Crook, 2016; Bequé and Lessmann, 2017), relatively few have focused on building such models for developing and emerging markets (Abdou, *et al.,* 2008; Abdou, 2009a-b; Abdou and Pointon, 2009; Khashman, 2011; Louzada, *et al.,* 2012; Bekhet and Eletter, 2014; Abdou, *et al.* 2016; Fernandes and Artes, 2016). While these have addressed a wide range of cases none, to the authors' knowledge, have examined the Indian banking sector. Given the sensitivity of data access is significant. Particularly, in the light of past financial crises, banks become increasingly risk reverse due to security and clients data protection laws. Small samples are widely used in building scoring models in the literature, as this issue is well recognised (see for example Paliwal and Kumar, 2009; Abdou and Pointon, 2011; Lessmann *et al.,* 2015). For instance, consumer loan applications models are regularly

built using around1,000 observations or less (see for example Kim and Sohn, 2004; Lee and Chen, 2005; Sustersic *et al.*, 2009; Derelioğlu and Gürgen, 2011; Abdou, *et al.* 2016). In building scoring models, statistical techniques such as discriminant analysis and logistic regression are widely used (Tsai, *et al.*, 2009; Akkoc, 2012; Louzada, *et al.*, 2012; Wang, *et al.*, 2012; Abdou, *et al.* 2014; Bekhet and Eletter, 2014; Abdou, *et al.*, 2016). The logistic regression model does not necessarily require the assumptions of the discriminant analysis model and may prove to be more robust in practical applications.

Other classification techniques such as classification and regression tree, k-nearest neighbour and support vector machines are also in common use (Hsieh, 2005; Lee, et al., 2006; Bellotti and Crook, 2009; Huang, 2011; Brown and Mues, 2012; Majeske and Lauer, 2013; Abdou, *et al.*, 2016). Various neural networks, including artificial neural networks, multilayer perceptron neural networks and back-propagation neural networks, have also been used in building scoring models (Abdou, 2009a; Khashman, 2011; Akkoc, 2012; Wang, *et al.*, 2012; Bekhet and Eletter, 2014). Amongst these probabilistic neural networks provide results which are significantly more accurate in building personal loan scoring models (see, Bensic, *et al.*, 2005; Abdou, *et al.*, 2008; Abdou and Pointon, 2009; Mostafa, 2009; Wang, et al., 2009; Louzada and Ara, 2012).

Comparisons between traditional and advanced scoring techniques have been the subject of numerous studies (West, 2000; Abdou, et al., 2008; Abdou, 2009b; Tsai, *et al.*, 2009; Khashman, 2011; Akkoc, 2012; Brown and Mues, 2012; Majeske and Lauer, 2013; Abdou, *et al.*, 2016). A substantial number of these studies demonstrate the superiority of neural networks over conventional techniques (Malhotra and Malhotra, 2003; Lee and Chen, 2005; Abdou, et al., 2008; Wang, et al., 2009; Brown and Mues, 2012; Louzada and Ara, 2012; Abdou, *et al.* 2014; Bekhet and Eletter, 2014; Abdou, *et al.*, 2016). However, there is still a role for conventional techniques such as discriminant analysis and logistic regression in building scoring models for personal loans (see for example, Hand and Henley, 1997; Abdou, *et al.*, 2008; Bekhet and Eletter, 2014).

In this paper four statistical modelling techniques are applied to analyse bank personal loans using a data-set provided by an Indian bank. As motivated by the above literature these are discriminant analysis, logistic regression, multi-layer feed-forward neural networks and probabilistic neural networks. Three different criteria namely correct classification rate, error

rates and *actual* misclassification cost are used to compare the effectiveness and predictive capabilities of different models. Moreover, in this paper *actual* misclassification costs, provided by the bank's own credit officials, are used in preference to the more conventionally used estimated misclassification costs. This underscores the novelty of our contribution.

The layout of this paper is organised as follows: Section 2 reviews the current guidance note on credit risk management by RBI. Section 3 addresses research methodology and data sources. Section 4 discusses the empirical results. Section 5 concludes and discusses the opportunities for further research.

2. Current credit risk management practices in Indian banks

In the 21st Century banks are confronted with an increasingly complex combination of interdependent financial and non-financial risks. This includes credit, interest rate, liquidity issues, regulatory, reputational and operational risks. These risks need to be controlled and managed by banks' senior executives. Further, major decisions about whether or not to implement a centralised or decentralised structure to manage these risks are faced by banks all over the world. In India, banks have been guided by a centralised approach on their credit risk from the RBI "Guidance Note on Credit Risk Management" that was issued in 2002[1]. These guidelines recommend that banks need a credit risk framework that focuses on policy and strategy, organisational structure and systems, as discussed below.

Credit risk policy and strategy. Banks require a board-approved risk policy and strategy that clearly identifies how to manage the bank's lending portfolio. Strategic plans must establish the credit granting processes that will be utilised by the bank with due consideration for the target market and cost/benefit considerations. *Organisational structure*. Risk management committees and credit risk management departments are vital structural components in establishing successful risk systems that clearly identify accountability and ensure that responsibility flows from the Board of Directors down to lending officers.

Credit Risk Frameworks (CRFs) are used to avoid an overly simplistic approach to risk classification and a process that is used to formulate risk-ratings is as follows:

[1]This Guidance Note on Credit Risk Management is still current as of 2015 (RBI, 2015).

1. Identify all the principal business and financial risk elements.
2. Allocate weights to principal risk components.
3. Compare with weights given in similar sectors and check for consistency.
4. Establish the key parameters (sub-components of the principal risk elements).
5. Assign weights to each of the key parameters.
6. Rank the key parameters on the specified scale.
7. Arrive at the credit-risk rating on the CRF.
8. Compare with previous risk-ratings of similar exposures and check for consistency.
9. Conclude the credit-risk calibration on the CRF (RBI, 2015).

Credit risk modelling techniques encourage a more quantitative and less subjective approach to personal lending. These methods have enhanced the measurement of risk and performance in banks' lending portfolios. The modelling techniques suggested by the RBI Guidelines include econometric techniques, neural networks, optimisation models, rule-based or expert systems and hybrid systems. In this paper we explore the first two set of techniques (for details regarding the credit risk framework, see the Appendix). Credit risk models as described by RBI Guidance Notes encourage the statistical analysis of historical data including the Z-score model and Emerging Market Scoring (EMS) model (RBI, 2015).

3. Research methodology

The main aim of this paper is to investigate whether apposite credit scoring models can lead to more efficiently discriminating creditworthiness evaluation and ultimately towards lower default rates. At an early stage of this research we conducted structured interviews with key decision-makers in a number of private and foreign banks in India. This included state and regional sales managers, territory managers of personal loans, branch managers, credit approvals and credit default controllers. The importance of doing this was threefold. Firstly, these interviews enabled us to establish a list of explanatory variables, which are used as part of *actual* lending procedures. Secondly, the results of these interviews form a natural complement to the available academic literature. Thirdly, we were able to establish that there was no set method used in the evaluation of personal loan applications in India. In many cases a predominantly judgemental approach was employed.

In building our proposed scoring models we adopt a two-stage analysis and use four different statistical modelling techniques namely discriminant analysis, logistic regression, multi-layer

feed-forward neural networks and probabilistic neural networks. In the first stage, we build our scoring models and, using *actual* misclassification costs, test the predictive capabilities of the various scoring models. In the second stage we focus upon the *default* cases, using 'customer began to default' as a dependent variable, and the same set of explanatory variables as used in the first stage of the analysis. Furthermore, a Variable Impact Analysis is conducted as part of the two stage analysis to identify the key determinants of both successful and defaulted cases.

3.1 Data collection and sampling procedures

In order to build our proposed credit scoring models, we use historical data comprising 2,093 personal loans supplied by one of the largest banks in India. Thus, given the data sensitivity, our sample size is in line with the previous literature (see for example, Lessmann *et al*., 2015; Paliwal and Kumar, 2009). The significance of our dataset is as follows. Firstly, based on literature reviews in Lessman *et al.* (2015) and Paliwal and Kumar (2009), our sample size appears to be in the top 20% of the published literature. Secondly, even when reported, larger sample sizes can be misleading. Often studies report results for multiple sub-samples. Though the average sub-sample size may be higher than our sample, it is common that several of the sub-samples may be significantly smaller than 2,000 observations (see e.g. Brown and Mues, 2012; Baesens *et al.* 2003; Lessman *et al.*, 2015). Thirdly, our application is interesting and important in its own right due to its focus upon developing countries. Of the ten papers identified in Lessman *et al.* (2015) as having larger sample sizes than our own, seven focus upon developed countries. In terms of applications to developing countries larger samples are either derived from externally funded research projects (Lee *et al.*, 2006; Huang *et al.*, 2007) or, whilst slightly larger, are of a similar order of magnitude (Yap *et al.*, 2011; 2,765 cases). Fourthly, it is important to recognise that our sample derives from a real-world credit scoring problem and data we ourselves collected. This stands in marked contrast to a small number of classical datasets that are regularly used in studies of credit scoring (see e.g. Table 3 in Lessman *et al.*, 2015). Furthermore, our unique blind data set used in this paper covers a lending range from Rupees ₹ crore 50,000 to Rupees ₹ crore 100,800,000 for its customers from 2009 to 2014, of which 1,233 are considered good loans and the remainder *860 are bad loans*. Having such a high percentage (41.09%) of bad loans, the dataset can be considered as '*pertinent'* (see for example, Abdou *et al.*, 2008).

The Indian bank provide 20 predictor variables which are mainly used in their decision making process. However, 6 predictors are excluded leaving 14 explanatory variables which are used

in building the scoring models, as shown in Table 2. Having a 'land line' is a mandatory decision criterion, without which the application is declined. Similarly, the provision of legal documentation is mandatory. Both 'state" and 'pin code' (equivalent to a postal code in the UK or a zip code in the USA) are considerably highly correlated (i.e. 97.70%) and therefore pin code is excluded[2]. We also excluded both the 'starting and the ending actual year' as we use 'term' as an explanatory variable[3]. The 'customer begin to default' variable is excluded when building the scoring models in the first stage. However, this variable is used as a dependent variable when running the sensitivity analysis investigating the incidence of the default cases[4], i.e. in the second stage, see Section 4.3.

TABLE (2) ABOUT HERE

In order to build our scoring models, Palisade Neural Tools, STATGRAPHICS Centurion XVI, IBM-SPSS Statistics 22 and R are used. We use a stratified 10-fold cross-validation technique to test the predictive capabilities of our scoring models. We randomise the data so that the percentage of bad customers in each group is the same, using R. The training set consists of 1883 cases (except for three folds, which consists of 1884 cases) and the hold-out set consists of 209 cases (except for three folds, which consists of 210 cases)[5].

3.2. Statistical scoring techniques

3.2.1. Discriminant Analysis

Discriminant analysis (DA) is a discrimination and classification technique, first popularised in bankruptcy prediction by Altman (1968). The following formula can be used for MDA:

$$Z = \alpha + \delta_1 X_1 + \delta_2 + \cdots + \delta_n X_n,$$

where,

[2]Our sample includes over 200 'pin codes' which make it almost impossible to be used as a categorical explanatory variable, and it does not add any value to be used as a numerical explanatory variable. However, retaining 'state', as an explanatory variable, can capture any loan quality differences between the states.
[3]Other Indian banks use a number of different variables as part of their credit evaluation which include, for example, length at current employment, spouse income and number of dependents.
[4]Interestingly, there is a belief stated by credit officials in the Indian banking sector that there is no need to include variables such as guarantees, field visits and feasibility studies in their credit evaluation processes.
[5]The correlation between the predictor variables are within an acceptable range i.e. <0.50.

Z represents the discriminant z-score, α is the intercept term, and δ_i is the respective coefficient in the linear combination of explanatory variables, X_i , for i = 1 to n (see, for example, Abdou, 2009a).

3.2.2. Logistic Regression

Logistic Regression (LR) is a widely used statistical modelling technique, in which the probability of a dichotomous outcome is related to a set of predictor variables in the form:

$$\log\left(\frac{p}{1-p}\right) = \alpha + \delta_1 X_1 + \delta_2 X_2 + \cdots + \delta_n X_n,$$

where,

p is the probability of default, α is the intercept term, and δ_i represents the respective coefficient in the linear combination of predictor variables, X_i, for i = 1 to n. The dependent variable is the logarithm of the odds ratio, $\{\log[\, p/(1-p)]\}$ (see, for example, Abdou *et al.* 2016).

3.2.3. Multi-Layer Feed-forward Network

It is convenient to use Multi-Layer Feed Forward Networks (MLFNs) to represent complex relationships between a set of variables. Figure 2 presents an example of a MLFN structure as follows:

FIGURE (2) ABOUT HERE

The following formula explains the MLFN function for two hidden layers:

$$Y = CF\left[\sum_{k=1}^{m} WO_k . CH_k^2 . \left\{\sum_{k=1}^{r} WH_{jk} . CH_j^1 . \left(\sum_{i=1}^{n} WI_{ij} . X_i\right)\right\}\right]$$

where,

Y = the output of the network; CF = conversion function for the output layer; WO_k = connection weighted summation to the output layer from the second hidden layer; CH_k^2 =

conversion function for the second hidden layer for node k; WH_{jk} = conversion weighted summation from the first hidden layer to the second hidden layer; CH_j^1 = conversion function for the first hidden layer for node j; WI_{ij} = conversion weighted summation from the input layer to the first hidden layer; X_i = inputs variables for node i; m = number of nodes in the second hidden layer; r = number of nodes in the first hidden layer; and n = number of input nodes (see, Abdou, 2009a, p. 102).

3.2.4. Probabilistic Neural Network

A Probabilistic Neural Network (PNN) is primarily a classifier, mapping inputs to a number of classifications, which might be imposed into a more general function. Figure 3 presents an example of a PNN structure, as follows:

FIGURE (3) ABOUT HERE

The Bayesian probability density function, for the respective output from PNN pattern node, can be represented as follows (see, Abdou, 2009a):

$$P(\underline{X}/C_i) = \frac{1}{(2\pi)^{m/2}\sigma^m n_i} \sum_{j=1}^{n} \exp\left[\frac{-(\underline{X} - \underline{X}_{ij})^T(\underline{X} - \underline{X}_{ij})}{2\sigma^2}\right]$$

where,

$\underline{X}$ = vector of observed inputs; n_i = number of training patterns for class C_i; $\underline{X}_{ij}$ = j[th] training vector for class C_i; m = vector-dimension; σ = standard deviation parameter for smoothing purposes; C_i = category class; T = transposition function for vector; and P = probability.

The conditional probability can be written as:

$$P(C_i / \underline{X}) = \frac{P(\underline{X}/C_i)P(C_i)}{P(\underline{X})}$$

for each class, using the basic Bayes' formula (see, Abdou, 2009a, p. 100).

4. Empirical results and analysis

We present descriptive statistics for our predictor variables followed by our two-stage results. Stage one, focuses on presenting the results of the four statistical models (shown in Section 3.2.) using the 10-fold cross validation. Then we compare different statistical techniques results predictive capabilities using average classification rates, errors rates and *actual misclassification costs*. In addition, we present a ranking of the relative importance of the predictor variables. Stage two performs an additional sensitivity analysis of the *default* cases.

4.1. Descriptive statistics

Table 3 provides descriptive statistics for the categorical variables used in building our scoring models. It can be concluded that 'state' is the most important predictive variable as it has the highest information value of 0.021. It is clearly evident that State C has the worst Weight of Evidence (WOE) value of -40.42 compared to 12.52 for State A. This may imply a preference of lending to clients from State A. Similarly, and counter-intuitively, females (WOE = -23.65) are less creditworthy compared to their male counterparts (WOE = 4.95). Our descriptive statistics show that other predictor variables are less important, with lower information values, when compared to State and Gender. As to the continuous predictors, five variables are also used in building our scoring models as follows: Age ranges from 23 to 56 years old; Term ranges from 2 to 4 years; EMI ranges from Rupees ₹ crore 1,468.5 to Rupees ₹ crore 2,960,496; Loan Amount ranges from Rupees ₹ crore 50,000 to Rupees ₹ crore 100,800,000; and Net Income ranges from Rupees ₹ crore 570,000 to Rupees ₹ crore 1,310,000.

TABLE (3) ABOUT HERE

The following sub-sections present classification results, including *Actual* Misclassification Costs (AMC), for our scoring models presented in Section 3.2. We use actual ratios of 6.5:1.6 and 15:1.7 for 2006 and 2011, respectively, to calculate the AMC associated with Type II and Type I errors. These actual ratios were provided by the Indian bank's own credit officials. This offers a refinement of the traditional approximate way of incorporating expected misclassification costs in the literature (see for example, Abdou, 2009b). Our unique AMC can be calculated using

$$AMC = \{ACR_1 \times P_{(B/G)} \times \pi_1\} + \{ ACR_2 \times P_{(G/B)} \times \pi_0\},$$

where,

ACR_1 denotes the corresponding actual cost ratio associated with a Type I error; $P_{(B/G)}$ denotes the associated probability of a Type I error; π_1 denotes the prior probability of good cases; ACR_2 denotes the corresponding actual cost ratio associated with a Type II error; $P_{(G/B)}$ denotes the associated probability of a Type II error; π_0 denotes the prior probability of bad cases.

These actual misclassification cost ratios that were provided, pre credit crunch, demonstrated a more favourable outlook in India with a 2006 ratio of 1.6:6.5 compared to previous studies (see for example, Abdou *et al.* 2008; Abdou, 2009b) who used a ratio of 1:5. However, the later figures used reflect a clear deterioration in the Indian lending climate with a ratio of 1.7:15 being used from 2011. This deterioration is confirmed by observations that the RBI raised interest rates to tame inflation and, due to worsening credit conditions, asked lenders to double their provisions for bad loans (see Financial Times, 2011; 2015).

Furthermore, as an additional robustness test, for the two neural network models, namely PNN and MLFN, we run the 10-folds cross validation again, this time allowing the 10-folds to be chosen at random.

4.2. Statistical scoring techniques: Stage 1

4.2.1. Discriminant analysis

Table 4 summarises the classification results for the 10 DA scoring models hold-out sub-samples using a default cut-off score of 0.50. The Average Correct Classification Rates (ACCR) range from 63.33% to 78.10% with a mean ACCR of 67.61%. Type I errors range from 8.06% to 30.08%; Type II errors range from 41.86% to 60.47%; and Total Error (TE) rates range from 21.90% to 36.67%. The average mean for Type I, Type II and TE are 21.82%, 47.56% and 32.39%, respectively. Notably, the *actual* misclassification costs for years 2006 and 2011 range from 1.19 to 1.72, and from 2.65 to 3.85, with an average mean of 1.48 and 3.15, respectively (see Table 4). Clearly, this suggests that AMC has significantly increased over time. This should motivate decision-makers to apply scoring models to reduce default rates.

4.2.2. Logistic regression

Results of the 10 LR scoring models hold-out sub-samples using a default cut-off score of 0.50, are shown in Table 5. The ACCR range from 50.24% to 77.62% with an average mean of

66.03%. Type I error rates range from 11.29% to 50.24% with an average mean of 27.49%. Type II error rates range from 38.37% to 51.16% with an average mean of 43.26%. The TE rates range from 22.38% to 49.76% with an average mean of 33.97%. As per actual misclassification costs, they range from 1.13 to 1.69 and from 2.47 to 3.47 for years 2006 and 2011, respectively. The average mean for the AMC for years 2006 and 2011 are 1.41 and 2.94 (see Table 5). Again, our results show notable increases in AMC over time. These results are in line with DA scoring models results shown in Section 4.2.1.

TABLE (4) ABOUT HERE

TABLE (5) ABOUT HERE

4.2.3. Multi-layer Feed-forward Networks

Tables 6 and 7 give the classification results for the 10 MLFN scoring models hold-out sub-samples and the additional 10 MLFN scoring models based on random runs, respectively. As per the former, the ACCR ranges from 63.16% to 76.67% with an overall mean of 67.13%. Type I, Type II and TE rates range from 10.57% to 44.72%, from 19.77% to 54.65%, and from 23.33% to 36.84%, respectively. The overall mean for these error rates are 27.74%, 40.23%, and 32.87%, respectively. For MLFN the AMC ranges from 0.95 to 1.68, and from 1.67 to 3.60 for years 2006 and 2011, respectively. The overall means for these AMC are 1.34 and 2.76, respectively (see Table 6). As per the latter, our 10 MLFN scoring models based on random runs show slightly better results under each of the previous criteria. As shown in Table 7, the overall means are 70.57%, 23.15%, 39.13%, and 29.43% for ACCR, Type I, Type II and TE rates, respectively. More importantly, the AMC results also improved showing that the overall means are 1.22 and 2.55 for years 2006 and 2011, respectively. These results emphasise that MLFN can offer better results compared to conventional statistical techniques shown in Sections 4.2.1.-4.2.2.

4.2.4. Probabilistic Neural Networks

Table 8 summarises classification results for the 10 PNN scoring models hold-out sub-samples. The ACCR ranges from 59.81% to 81.90% with an average mean of 68.71%. Error rates results show that they range from 5.65% to 31.71% for Type I error with an average rate of 23.11%; they range from 36.05% to 52.33% for Type II errors with an overall mean rate of 43.02%; and they range from 18.10% to 40.19% for the TE rates with an overall mean of 31.29%. AMC results show that they range from 1.01 to 1.70 and from 2.27 to 3.55 for years 2006 and 2011,

with average means of 1.37 and 2.88, respectively (see Table 8). Results shown in Table 9 are for the 10 PNN scoring models based on random runs. Clearly, these results are the best amongst our scoring models with exception of the AMC 2011 results. The overall means are 73.20%, 18.49%, 38.73%, and 26.85% for ACCR, Type I, Type II and TE rates, respectively. Furthermore, the AMC results show that the overall means are 1.21 and 2.59 for years 2006 and 2011, respectively. These results demonstrate that our neural network models, namely PNN and MLFN, can lead to further material reductions in default losses.

TABLE (6) ABOUT HERE

TABLE (7) ABOUT HERE

TABLE (8) ABOUT HERE

TABLE (9) ABOUT HERE

4.2. Comparison of different statistical scoring models

Comparing different models where the same 10-folds are used, neural network models, namely PNN and MLFN, outperform conventional models, namely DA and LR, used in this paper. That is, PNN models show the highest ACCR of 68.71% and the lowest TE of 31.29%; whilst MLFN show the lowest AMC of 1.34 and 2.76 for 2006 and 2011, respectively. Furthermore, when the 10-folds are randomly chosen both PNNran and MLFNran results show improvement under different criteria and both models are still outperform other techniques. On the one hand, PNNran has the highest ACCR of 73.20%, the lowest TE of 26.85% and the lowest AMC of 1.21 for 2006, whilst MLFNran has the lowest AMC of 2.55 for 2011. Our results suggest that the default rate of 41.09% could be reduced to 26.85% using PNNran scoring models (see Table 9).

We then use a *General linear* model, which is a one-way Analysis of Variance (ANOVA), to investigate whether there are significant differences between different models for the scoring criteria outlined above[6]. The general linear model with categorical variables is formed by setting

$$X_i = \mu + \alpha_i + \varepsilon_i,$$

where,

[6]The focus here is upon the hold-out sub-samples.

μ is the overall mean, α_i is the *i*th treatment effect (under the identifiability constraint $\sum_i \alpha_i = 0$), and the ε_i are iid $N(0,\sigma^2)$ (see for example, Bingham and Fry, 2010). Table 10 shows our results and there is an evidence of statistically significant differences between the scoring models for each criterion. The graphical illustration (see Figure 4) confirms the findings shown in Table 10.

TABLE (10) ABOUT HERE

FIGURE (4) ABOUT HERE

4.2.1. Importance of different predictor variables used in building the scoring models

Table 11 shows the Average Variable Impact (AVI) for each of the 14 predictor variables under each of the scoring models applied in this paper across 10-folds. Clearly, alternative models may treat various predictor variables differently when it comes to their impact on loan quality. By averaging the variable impact weight over 60 scoring models, for each predictor variable under each of the statistical techniques, we identified net income (NINC), marital status (MRST) and loan amount (LAMT) as of key importance in distinguishing clients' creditworthiness. In contrast, vehicle ownership (OVEH), loan duration (TERM) and client's job (JOB) are the least important determinants of clients' creditworthiness.

TABLE (11) ABOUT HERE

4.3. Sensitivity analysis of default credits: Stage 2

The main aim of this stage is to shed light upon the default cases given that they constitute a relatively large proportion of the entire sample (over 41%, 860 out of a total of 2,093 cases). We use a stratified 5-fold cross-validation technique to explain the timing of the incidence of default. We use the same four statistical modelling techniques shown in Section 3.2. We rerun additional 5-fold cross validation with folds randomly chosen by the software for both MLFN and PNN. However, it should be emphasised that the main focus of this section is to identify the key determinants of the incidence of default. Interestingly, in our sample, default occurs only in the first and second years, and none in later years. We randomise the data so that the percentage of bad customers who start to default in their first year and those who start to default

in their second year are the same, using R. The training set consists of 688 cases and the hold-out set consists of 172 cases.

4.3.1. Descriptive statistics for default customers

In building our scoring models, we use the same 14 explanatory variables, as shown in Table 2. However, the dependent variable used in this section is 'customer begin to default' replacing 'loan quality' in the original modelling. As to the five continuous predictors, Age ranges from 23 to 56 years old; EMI ranges from Rupees ₹ crore 1,469 to Rupees ₹ crore 469,920; Loan Amount ranges from Rupees ₹ crore 5,000 to Rupees ₹ crore 16,000,000; Net Income ranges from Rupees ₹ crore 570,000 to Rupees ₹ crore 1,250,000; and Term ranges from 2 to 4 years. Nine categorical variables are used in building our models. *Inter alia* the sample consists of 693 males and 167 females; 353 single, 498 married and 9 others; 442 graduates and 418 post-graduates; 251 work in the public sector and 609 work in the private sector. Our sample show that 288 start to default during the first year of the loan facility, and 572 start to default during the second year.

4.3.2. Importance of different variables for the default cases

It is crucial for decision-makers to become fully aware of the key determinants of the incidence of default, which in turn may reflect on their final decision. Table 12 shows the AVI for each of the 14 predictor variables under each of the models across 5-folds. By averaging the variable impact weight over 30 models, for each predictor variable under each of the statistical techniques, we identified the following three key determinants of the incidence of default, in order of importance: State of residence (STATE); equated monthly instalment (EMI) and actual loan amount (LAMT). This stands in marked contrast to vehicle ownership (OVEH), previous employment (PEMP) and educational level (EDU) which are the least important predictor variables.

Considering both the first and the second stages impact analyses of predictor variables, we strongly recommend the Indian banking sector to take into account the following set of predictor variables when making lending decisions: *STATE, EMI, NINC, MRST and LAMT*. This can have a demonstrable impact on the loan quality and subsequently on the overall lending decision making process.

TABLE (12) ABOUT HERE

We run additional statistical tests to distinguish between early and late defaulters in relation to our key variables namely, *STATE, EMI, NINC, MRST and LAMT*. There are no significant differences between different MRST sub-categories namely single, married and others. Likewise, there are no significant differences between different levels of income. In contrast, early defaulters are associated with higher levels of EMI and LAMT. Furthermore, none of the residents in State C has defaulted in the first year; however, much larger numbers defaulted in the second year. Finally, the largest number of both early and late defaulters are located in State B.

In summary, and as part of our policy implications, recent news report that high default rates, rising bad debts and shrinking cash flows has led to enforced redundancies and the closure of a significant number of branches throughout India (Quartz India, 2015; Financial Times, 2015). Thus, evidence clearly demonstrates that it would have been less costly for the bank had it adopted our credit scoring models rather than implementing their own strategic decisions to downsize. These lessons are not limited to the Indian bank that provided our loan data-set as confirmed by recent news that four major foreign banks have reduced their exposure to the Indian market (Quartz India, 2015; Financial Times, 2015).

5. Conclusions and areas for further research

The main aim of our paper is to use a two-stage analysis to investigate whether scoring models can efficiently distinguish the Indian banking clients' creditworthiness, and reduce default rates. Working alongside the bank, our fresh contribution includes the incorporation of *actual* misclassification costs when evaluating our models. Our statistically rigorous analysis also stands in marked contrast to the predominantly subjective approach the bank were using to make lending decisions. In building our models we use four statistical modelling techniques namely discriminant analysis, logistic regression, multi-layer feed-forward neural network and probabilistic neural network. This is combined with a bespoke data-set with a default rate of over 41%.

As to our first stage, our 10-folds analysis shows that both PNN and MLFN, outperform conventional statistical models. PNN models perform better compared to other models in terms of conventional classification criteria such as ACCR and TE. However, MLFN models outperform others (including PNN) once *actual* misclassification costs are incorporated

achieveing the lowest AMC of 1.34 and 2.76 for 2006 and 2011, respectively. Moreover, when the randomly seclected 10-folds are incorporated, PNNran models outperform all other techniques (including MLFNran) achieveing the highest ACCR, the lowest TE, and the lowest AMC of 1.21 for 2006. However, there is still a role for MLFNran achieveing a marginally lower AMC of 2.55 for 2011. We have evidence of statistically significant differences between the scoring models for each criterion using a g*eneral linear* model. Out of 60 scoring models, we identified NINC, MRST and LAMT as key determinants of creditworthiness in the Indian banking sector. As to our second stage, we use 5-folds cross validation to build our models using the same set of statistical modelling techniques to explain the timing of the incidence of default. Out of our 30 models, we further identified STATE, EMI and LAMT as key determinants of the timing of default.

Moreover, when combining both stages outcomes, we identified *STATE, EMI, NINC, MRST and LAMT* as the most important predictor variables for the Indian banking sector. Further analysis shows that early defaulters are associated with higher levels of EMI and LAMT. STATE level effects are also prevalent in the incidence of default. This suggests that, in practice, greater care needs to be exercised when granting loans to clients from different states. In summary, by applying our proposed scoring models to the Indian banking sector, and alongside successful implementation, we argue that the challenges facing the Indian market could be significantly reduced. In particular, our best scoring models can significantly reduce our sample default rate by 14.24% (i.e. 41.09%, the original default rate – 26.85%, default rate using PNNran). *Inter alia* problems such as increasing interest rates in an attempt to restructure default debt, inflation and the increased cost of banks' debt could be mitigated. Other consequences of the high default rates have been the redundancy and branch-closure policies that some Indian banks followed in an attempt to cut costs. We submit that some of these cost-cutting measures could thus ultimately have been avoided.

In terms of the theory of expert and intelligent systems our proposed two-stage approach forms a natural complement to previous neural network (Gaganis et al., 2007; Öğüt et al., 2009) and hybrid (Li *et al.*, 2016) modelling of credit risk. We also show that methods such as neural networks can lead to better assessments of credit risk than classical statistical methods (Abdou, *et al.* 2016; Abellán and Castellano, 2017). Beyond reproducing aspects of real decision-making our results show that neural network models can lead to improved financial decision-making in industrial applications. In particular, neural network models may be particularly

useful when the distribution of instances in the dataset is unbalanced (Zhao et al., 2015) or information is scarce (Falavigna, 2012).

There are a number of opportunities for further work. This includes the application of additional techniques and their possible combination into integrated models with larger sample sizes. In particular, gene expression programming, fuzzy algorithms, proportional hazard models and SVM etc. Limitations of our study include potential concerns over the accuracy of industry-standard costings and the need for high computational efficiency in industrial-sized financial applications (see for example Zhao *et al.*, 2015). Results may also be sensitive to the economic conditions associated with the timing of the business cycle (see for example Derelioğlu and Gürgen, 2011). However, recent financial turbulence in India suggests extending our study to other products including credit cards, business loans and mortgages would also be extremely timely.

Appendix: grading system for calibration of credit risk

In this section, we discuss the rating scales and weighted scoring systems as typically applied in the lending departments of Indian banks.

Rating scales:

(i) Numerical values from 1 to 9 are utilised in rating scales with 1 to 5 representing levels of acceptable credit risk as shown in Table A1 below, and 6 to 9 representing unacceptable credit risk (RBI, 2015).

Table A1: Risk classification scheme

Risk Class	Description
1	Customer with no risk of default
2	Customer with negligible risk of default [Default Rate less than 2%]
3	Customer with little risk of default [Default Rate between 2% to 5%]
4	Customer with some risk of default [Default Rate between 5% to 10%]
5	Customer with significant risk of default [Default Rate in excess of 10%]

Source: Gosalia, (2010, p. 38), modified

(ii) Alphabetical and symbol rating scales such as AAA, AA+, A-, BBB are recognisable alternatives and widely used by various credit rating agencies, for example, Moody's, Fitch and Standard & Poors.

Weighted scoring systems: weighted systems apply a score or grade for risk profiling with suitably applied percentages assigned to each of the risk-ratings to produce a weighted average risk-rating. The example as shown in Table A2 below would be considered as a potentially low-risk rating:

Table A2: CRF weighted scoring system

Risk-rating area	Score	Weighting
If gross revenues between Rs. 800 to Rs. 1000 crore	2	20%
If operating margin is 20% or more	2	20%
If ROCE (Return On Capital Employed) is 25% or more	1	10%
If debt-equity ratio is between 0.60 to 0.80	2	20%
If interest cover is 3.5 or more	1	20%
If DSCR (Debt Service Coverage Ratio) is 1.80 or more	1	10%

Source: RBI (2015, p. 17).

Clearly the problem is how the Credit Risk Framework (CRF) assigns those weightings. In this paper and as a starting point we are assigning weightings for personal loans based on advanced statistical techniques such as neural networks to avoid any subjective bias in assigning these weightings.

References

Abdou, H. A. (2009a). *Credit scoring models for Egyptian banks: neural nets and genetic programming versus conventional techniques.* Ph.D. Thesis, The University of Plymouth, UK.

Abdou, H. A. (2009b). Genetic programming for credit scoring: The case of the Egyptian public sector banks. *Expert Systems with Applications, 36 (9),* 11402–11417.

Abdou, H. A., Alam, S. & Mulkeen, J. (2014). Would credit scoring work for Islamic finance? A neural network approach, *International Journal of Islamic and Middle Eastern Finance and Management. 7 (1),* 112–125.

Abdou, H. A. & Pointon, J. (2009). Credit scoring and decision making in Egyptian public sector banks, International Journal of Managerial Finance. 5 (4), 391-406.

Abdou, H. A. & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: a review of the literature. Intelligent Systems in Accounting, Finance and Management, 18 (2–3), 59–88.

Abdou, H. A., J. Pointon, & A. El-Masry. (2008). Neural Nets versus Conventional Techniques in Credit Scoring in Egyptian Banks. *Expert Systems with Applications, 35,* 1275-1292.

Abdou, H. A., Tsafack, M., Ntim, C. & Baker, R. (2016). Predicting creditworthness in retail banking with limited scoring data. *Knowledge-Based Systems, 103,* 89-103.

Abellán, J. & Castellano, J. G. (2017). A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Systems with Applications, 73,* 1-10.

Akkoc, S. (2012) An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research, 222,* 168-178.

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance, 23, 589-609.*

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J. & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society, 54,* 627-635.

Bekhet, H. A. & Eletter, S. F. K. (2014). Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. *Review of Development Finance, 4,* 20-28.

Bellotti, T. & Crook. J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications, 36,* 3302-3308.

Bensic, M., Sarlija, N. & Zekic-Susac, M. (2005). Modelling small-business credit scoring by using logistic regression, neural networks and decision trees. *Intelligent Systems in Accounting, Finance and Management, 13,* 133-150.

Bequé, A. & Lessmann, S. (2017). Extreme learning machines for credit scoring: An empirical evaluation. *Expert Systems with Applications, 86,* 42-53.

Bingham, N. H. & Fry, J. M. (2010). *Regression: Linear models in statistics.* London Dordtrecht Heidelberg New York: Springer.

Brown, I. & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications, 39,* 3446-3453.

CIBIL. Credit Information Bureau India Limited. (2016) http://www.cibil.com/faqs.htm Accessed April 15 2016.

Derelioğlu, G. & Gürgen, F. (2011). Knowledge discovery using neural approach for SME's credit risk analysis. *Expert Systems with Applications, 38,* 9313-9318.

Falavigna, G. (2012). Financial ratings with scarce information. *Expert Systems with Applications, 39,* 1784-1792.

Fernandes, G. B. & Artes, R. (2016) Spatial dependence in credit risk and its improvement in credit scoring. *European Journal of Operational Research, 249,* 517-524.

Financial Times. India banks fear rising bad loans. (2011). https://www.ft.com/content/5fd4a346-ccab-11e0-b923-00144feabdc0 Accessed July 16 2015.

Financial Times. Bad loans spiral in India's 'broken' banking system. (2015) https://www.ft.com/content/23890fb4-6d77-11e5-aca9-d87542bf8673 Accessed 16 May 2016.

Gaganis, C., Pasiouras, F. & Doumpos, M. (2007). Probabilistic neural networks for the identification of qualified audit opinions. *Expert Systems with Applications, 32,* 114-124.

Gosalia, K. H. (2010) *Credit delivery mechanism in Indian banks and suggestion for improvements.* Mumbai, Maharashtra, India.

Hand, D., & Henley, W. (1997). Statistical Classification Methods in Consumer Credit Scoring: A Review. *Journal of the Royal Statistical Society A, 160,* 523−541.

Hsieh, N-C (2005). Hybrid mining approach in the design of credit scoring models. *Expert Systems with Applications*, *28*, 655-665.

Huang, S-C. (2011). Using Gaussian process based kernel classifiers for credit rating forecasting. *Expert Systems with Applications*, *38*, 8607-8611.

Huang, C-L., Chen, M-C., & Wang, C-J. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications 33*, 847-856.

Khashman, A. (2011). Credit risk evaluation using neural networks: Emotional versus conventional models. *Applied Soft Computing*, *11*, 5477–5484.

Kim, Y.S. & Sohn, S.Y. (2004). Managing loan customers using misclassification patterns of credit scoring model. *Expert Systems with Applications 26(4)*, 567–573.

Lee, T.-S., & Chen, I.-F. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, *28*, 743–752.

Lee, T-S., Chiu, C-C., Chou, Y-C. & Lu, C-J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis*, *50*, 1113-1130.

Leow, M. & Crook, J. (2016). A new Mixture model for the estimation of credit card Exposure at Default. *European Journal of Operational Research*, *249*, 487-497.

Lessmann, S., Baesens, B., Seow, H. V. & Thomas, L. C (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124-136.

Li, K., Niskanen, J., Kolehmainen, M. & Niskanen, M. (2016). Financial innovation: credit default hybrid model for SME lending. *Expert Systems with Applications, 61,* 343-355.

Louzada, F. & Ara, A. (2012). Bagging *k*-dependence probabilistic networks: An alternative powerful fraud detection tool. *Expert Systems with Applications, 39*, 11583-11592.

Louzada, F., Ferreira-Silva, P. H. & Diniz, C. A.R. (2012). On the impact of disproportional samples in credit scoring models: An application to a Brazilian bank data. *Expert Systems with Applications*, *39*, 8071-8078.

Majeske, K. D. & Lauer, T. W. (2013). The bank loan approval decision from multiple perspectives. *Expert Systems with Applications*, *40*, 1591-1598.

Malhotra, R., & Malhotra, D. K. (2003). Evaluating consumer loans using neural networks. *Omega*, *31*, 83-96.

Marshall, A., Tang, L. & Milne, A. (2010). Variable reduction, sample selection bias and bank retail credit scoring. *Journal of Empirical Finance, 17*, 501-512.

Mostafa, M. M. (2009). Modeling the efficiency of top Arab banks: A DEA–neural network approach. *Expert Systems with Applications, 36*, 309-320.

Öğüt, H., AktaşR., Alp, A. & Doğanay, M. M. (2009). Prediction of financial information manipulation by using support vector machine and probabilistic neural network. *Expert Systems with Applications, 36,* 5419-5423.

Ono, A., Hasumi, R. & Hirata, H. (2014). Differentiated use of small business credit scoring by relationship lenders and transactional lenders: Evidence from firm–bank matched data in Japan. *Journal of Banking and Finance*, 42, 371-380.

Paliwal, M. & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications,* 36, 2-17.

Quartz India. Foreign banks are hurting in India after failing to recover globally. (2015). https://qz.com/358346/foreign-banks-are-hurting-in-india-after-failing-to-recover-globally/ Accessed 16 July 2015.

RBI. Guidance Note on Credit Risk Management. (2015) http://rbidocs.rbi.org.in/rdocs/notification/PDFs/32084.pdf Accessed on 16 October 2015.

RBI Annual Report (2014). https://rbidocs.rbi.org.in/rdocs/AnnualReport/PDFs/00A157C1B5ECBE6984F6EA8137C57AAEF493C.PDF Accessed October 2016.

Sustersic, M., Mramor, D. & Zupan, J. (2009). Consumer credit scoring models with limited data. *Expert Systems with Applications 36(3),* 4736–4744.

Tong, E. N. C., Mues, C. & Thomas, L. C. (2012). Mixture cure models in credit scoring: If and when borrowers default. *European Journal of Operational Research*, *218,* 132-139.

Tsai, M-C., Lin, S-P., Cheng, C-C. & Lin, Y-P. (2009). The consumer loan default predicting model – An application of DEA-DA and neural network. *Expert Systems with Applications*, *36,* 11682-11690.

Wang, Y., Li, L., Ni, J. & Huang, S. (2009). Feature selection using tabu search with long-term memories and probabilistic neural networks. *Pattern Recognition Letters, 30,* 661-670.

Wang, G., Ma, J., Huang, L., & Xu, K. (2012). Two credit scoring models based on dual strategy ensemble trees. *Knowledge-Based Systems, 26,* 61-68.

West, D. (2000) Neural network credit scoring models. *Computers and Operations Research, 27*, 1131–1152.

Yap, B. W., Ong, S. H. and Husain, N. H. M. (2011). Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Systems with Applications 38*, 13274-13283.

Zhao, Z., Xu, S., Kang, B. H., Kabir, M. M. J. & Liu, Y. (2015). Investigation and improvement of multi-layer perceptron neural networks for credit scoring. *Expert Systems with Applications, 42,* 3506-3516.

TABLES

Table 1: Growth of personal loans, RBI

Year ended March	Personal loans outstanding Rupees ₹ crore	Variation Absolute Rupees ₹ crore	 Per cent
2007	452758	-	-
2008	507488	54730	12.09
2009	562479	54991	10.84
2010	585633	23154	4.12
2011	685372	99739	17.03
2012*	789990	104618	15.26
2013*	900890	110900	14.04

Source: RBI Annual Reports (2009/10, 2010/11, 2012/13), adapted.
*Numbers for these years are converted from billion to crore.

Table 2: List of predictor variables used in building the scoring models

Variables	**Code**	**Unit**	**Comments**
x_1 Gender	GEN	Categorical	0=Male, 1=Female
x_2 Marital Status	MRST	Categorical	0=Single, 1=Married, 2=Others e.g. divorced
x_3 EMI	EMI	Numerical	Refers to the actual Equated Monthly Instalment
x_4 Loan Amount	LAMT	Numerical	Actual loan amount in Rupees ₹ crore
x_5 Term	TERM	Numerical	Loan duration is between 2 and 4 years
x_6 State	STATE	Categorical	0=State A, 1=State B, 2=State C
x_7 Loan Purpose	LPRP	Categorical	0=Customer durable, 1=Home renovation, 2= Luxury purchase, 3= Travel and tourism, 4=Unplanned expenses.
x_8 Job	JOB	Categorical	0=Public sector job, 1=Private sector job
x_9 Previous Employment	PEMP	Categorical	0=No and 1=Yes
x_{10} Age	AGE	Numerical	Actual age of the client, and range between 23-56
x_{11} Education	EDU	Categorical	0=Graduate, 1=Post graduate
x_{12} Net Income	NINC	Numerical	Actual net income in Rupees ₹ crore
x_{13} Vehicle	OVEH	Categorical	0=Does not own a vehicle, 1=Own vehicle(s)
x_{14} Other Loan	OTLO	Categorical	Have taken loan from other bank or not. 0=Yes, 1=No, 2=Unknown
y Loan Quality	LQUA	Categorical	0=Bad, 1=Good

Table 3: Descriptive statistics for categorical variables

Characteristic	Code	No. of cases	Total %	Good cases	Good cases %	Bad cases	Bad cases %	Bad Rate	WOE
Gender									
Male	0	1737	82.99%	1044	84.67%	693	80.58%	39.90%	4.951
Female	1	356	17.01%	189	15.33%	167	19.42%	46.91%	-23.652
Information value:0.012*									
Marital Status									
Single	0	842	40.23%	489	39.66%	353	41.05%	41.92%	-3.438
Married	1	1227	58.62%	729	59.12%	498	57.91%	40.59%	2.08
Others e.g. Divorced	2	24	1.15%	15	1.22%	9	1.05%	37.50%	15.055
Information value:0.001									
State									
State A	0	819	39.13%	507	41.12%	312	36.28%	38.10%	12.523
State B	1	1092	52.17%	637	51.66%	455	52.91%	41.67%	-2.38
State C	2	182	8.70%	89	7.22%	93	10.81%	51.10%	-40.424
Information value:0.021									
Loan Purpose									
Consumer Durables	0	357	17.06%	202	16.38%	155	18.02%	43.42%	-9.543
Home Renovation	1	539	25.75%	320	25.95%	219	25.47%	40.63%	1.898
Luxury Purchase	2	513	24.51%	312	25.30%	201	23.37%	39.18%	7.943
Travel & Tourism	3	523	24.99%	302	24.49%	221	25.70%	42.26%	-4.801
Unplanned Expense	4	161	7.69%	97	7.87%	64	7.44%	39.75%	5.555
Information value:0.004									
Job									
Public	0	640	30.58%	389	31.55%	251	29.19%	39.22%	7.785
Private	1	1453	69.42%	844	68.45%	609	70.81%	41.91%	-3.394
Information value:0.003									
Previous Employment									
No	0	248	11.85%	137	11.11%	111	12.91%	44.76%	-14.982
Yes	1	1845	88.15%	1096	88.89%	749	87.09%	40.60%	2.041
Information value:0.003									
Education									
Graduate	0	1060	50.65%	618	50.12%	442	51.40%	41.70%	-2.509
Post Graduate	1	1033	49.35%	615	49.88%	418	48.60%	40.46%	2.587
Information value:0.001									
Vehicle									
Does Not Own	0	688	32.87%	407	33.01%	281	32.67%	40.84%	1.019
Own	1	1405	67.13%	826	66.99%	579	67.33%	41.21%	-0.498
Information value:0.000									
Other Loan									
Yes	0	1051	50.22%	617	50.04%	434	50.47%	41.29%	-0.845
No	1	573	27.38%	347	28.14%	226	26.28%	39.44%	6.852
Unknown	2	469	22.41%	269	21.82%	200	23.26%	42.64%	-6.388
Information value:0.002									

**Information Value, or total strength of the characteristics*, relates directly to the WOE, which may be used to identify the strength of the association between different variables. The higher the information values the greater the contribution of attributes to the final scores (for more details see Abdou *et al.*, 2016).

Table 4: Cross-validation results for the 10 Discriminant Analysis (DA) scoring models (hold-out sub-samples)

DA	Classification results			Error results			*Actual* Misclassification Costs	
	GG	BB	ACCR%	Type I	Type II	TE	AMC2006 (1.6:6.5)	AMC2011 (1.7:15)
$Fold_1$	91.94(114/124)	58.14(50/86)	78.1(164/210)	8.06(10/124)	41.86(36/86)	21.9(46/210)	1.190476	2.652381
$Fold_2$	79.84(99/124)	52.33(45/86)	68.57(144/210)	20.16(25/124)	47.67(41/86)	31.43(66/210)	1.459524	3.130952
$Fold_3$	72.58(90/124)	50(43/86)	63.33(133/210)	27.42(34/124)	50(43/86)	36.67(77/210)	1.59	3.346667
$Fold_4$	69.92(86/123)	56.98(49/86)	64.59(135/209)	30.08(37/123)	43.02(37/86)	35.41(74/209)	1.433971	2.956459
$Fold_5$	74.8(92/123)	51.16(44/86)	65.07(136/209)	25.2(31/123)	48.84(42/86)	34.93(73/209)	1.543541	3.266507
$Fold_6$	75.61(93/123)	53.49(46/86)	66.51(139/209)	24.39(30/123)	46.51(40/86)	33.49(70/209)	1.473684	3.114833
$Fold_7$	73.98(91/123)	56.98(49/86)	66.99(140/209)	26.02(32/123)	43.02(37/86)	33.01(69/209)	1.395694	2.915789
$Fold_8$	78.86(97/123)	55.81(48/86)	69.38(145/209)	21.14(26/123)	44.19(38/86)	30.62(64/209)	1.380861	2.938756
$Fold_9$	75.61(93/123)	50(43/86)	65.07(136/209)	24.39(30/123)	50(43/86)	34.93(73/209)	1.566986	3.330144
$Fold_{10}$	88.62(109/123)	39.53(34/86)	68.42(143/209)	11.38(14/123)	60.47(52/86)	31.58(66/209)	1.724402	3.845933
Mean	**78.18(964/1233)**	**52.44(451/860)**	**67.61(1415/2093)**	**21.82(269/1233)**	**47.56(409/860)**	**32.39(678/2093)**	**1.475914**	**3.149842**

Notation: GG refers to actual good cases, predicted as good cases; BB refers to actual bad cases, predicted as bad cases; TE refers to total error rates (Type I *plus* Type II).

Table 5: Cross-validation results for the 10 Logistic Regression (LR) scoring models (hold-out sub-samples)

LR	Classification results			Error results			Actual Misclassification Costs	
	GG	BB	ACCR%	Type I	Type II	TE	AMC2006 (1.6:6.5)	AMC2011 (1.7:15)
$Fold_1$	88.71(110/124)	61.63(53/86)	77.62(163/210)	11.29(14/124)	38.37(33/86)	22.38(47/210)	1.128095	2.470476
$Fold_2$	72.58(90/124)	54.65(47/86)	65.24(137/210)	27.42(34/124)	45.35(39/86)	34.76(73/210)	1.46619	3.060952
$Fold_3$	67.74(84/124)	48.84(42/86)	60(126/210)	32.26(40/124)	51.16(44/86)	40(84/210)	1.666667	3.466667
$Fold_4$	69.11(85/123)	61.63(53/86)	66.03(138/209)	30.89(38/123)	38.37(33/86)	33.97(71/209)	1.317225	2.677512
$Fold_5$	78.05(96/123)	59.3(51/86)	70.33(147/209)	21.95(27/123)	40.7(35/86)	29.67(62/209)	1.295215	2.731579
$Fold_6$	68.29(84/123)	58.14(50/86)	64.11(134/209)	31.71(39/123)	41.86(36/86)	35.89(75/209)	1.418182	2.900957
$Fold_7$	73.17(90/123)	60.47(52/86)	67.94(142/209)	26.83(33/123)	39.53(34/86)	32.06(67/209)	1.310048	2.708612
$Fold_8$	75.61(93/123)	54.65(47/86)	66.99(140/209)	24.39(30/123)	45.35(39/86)	33.01(69/209)	1.442584	3.043062
$Fold_9$	46.34(57/123)	55.81(48/86)	50.24(105/209)	53.66(66/123)	44.19(38/86)	49.76(104/209)	1.687081	3.264115
$Fold_{10}$	85.37(105/123)	52.33(45/86)	71.77(150/209)	14.63(18/123)	47.67(41/86)	28.23(59/209)	1.412919	3.088995
Mean	**72.51(894/1233)**	**56.74(488/860)**	**66.03(1382/2093)**	**27.49(339/1233)**	**43.26(372/860)**	**33.97(711/2093)**	**1.414421**	**2.941293**

Notation: GG refers to actual good cases, predicted as good cases; BB refers to actual bad cases, predicted as bad cases; TE refers to total error rates (Type I *plus* Type II).

Table 6: Cross-validation results for the 10 Multi-layer Feed-forward Neural Networks (MLFN) scoring models (hold-out sub-samples)

MLFN	Classification results			Error results			Actual Misclassification Costs	
	GG	BB	ACCR%	Type I	Type II	TE	AMC2006 (1.6:6.5)	AMC2011 (1.7:15)
$Fold_1$	86.29(107/124)	62.79(54/86)	76.67(161/210)	13.71(17/124)	37.21(32/86)	23.33(49/210)	1.12	2.423333
$Fold_2$	73.39(91/124)	61.63(53/86)	68.57(144/210)	26.61(33/124)	38.37(33/86)	31.43(66/210)	1.272857	2.624286
$Fold_3$	75.81(94/124)	45.35(39/86)	63.33(133/210)	24.19(30/124)	54.65(47/86)	36.67(77/210)	1.683333	3.60
$Fold_4$	76.42(94/123)	51.16(44/86)	66.03(138/209)	23.58(29/123)	48.84(42/86)	33.97(71/209)	1.52823	3.250239
$Fold_5$	58.54(72/123)	75.58(65/86)	65.55(137/209)	41.46(51/123)	24.42(21/86)	34.45(72/209)	1.043541	1.92201
$Fold_6$	66.67(82/123)	58.14(50/86)	63.16(132/209)	33.33(41/123)	41.86(36/86)	36.84(77/209)	1.433493	2.917225
$Fold_7$	55.28(68/123)	80.23(69/86)	65.55(137/209)	44.72(55/123)	19.77(17/86)	34.45(72/209)	0.949761	1.667464
$Fold_8$	62.6(77/123)	68.6(59/86)	65.07(136/209)	37.4(46/123)	31.4(27/86)	34.93(73/209)	1.191866	2.311962
$Fold_9$	78.05(96/123)	46.51(40/86)	65.07(136/209)	21.95(27/123)	53.49(46/86)	34.93(73/209)	1.637321	3.521053
$Fold_{10}$	89.43(110/123)	47.67(41/86)	72.25(151/209)	10.57(13/123)	52.33(45/86)	27.75(58/209)	1.499043	3.335407
Mean	**72.26(891/1233)**	**59.77(514/860)**	**67.13(1405/2093)**	**27.74(342/1233)**	**40.23(346/860)**	**32.87(688/2093)**	**1.335944**	**2.757298**

Notation: GG refers to actual good cases, predicted as good cases; BB refers to actual bad cases, predicted as bad cases; TE refers to total error rates (Type I *plus* Type II).

Table 7: Cross-validation results for the 10 Multi-layer Feed-forward Neural Networks (MLFN) scoring models (hold-out sub-samples) random runs

MLFNran	Classification results			Error results			Actual Misclassification Costs	
	GG	BB	ACCR%	Type I	Type II	TE	AMC2006 (1.6:6.5)	AMC2011 (1.7:15)
Fold1	82.95(107/129)	56.79(46/81)	72.86(153/210)	17.05(22/129)	43.21(35/81)	27.14(57/210)	1.250952	2.678095
Fold2	65.63(84/128)	68.29(56/82)	66.67(140/210)	34.38(44/128)	31.71(26/82)	33.33(70/210)	1.14	2.213333
Fold3	74.81(98/131)	54.43(43/79)	67.14(141/210)	25.19(33/131)	45.57(36/79)	32.86(69/210)	1.365714	2.838571
Fold4	71.76(94/131)	64.10(50/78)	68.90(144/209)	28.24(37/131)	35.90(28/78)	31.10(65/209)	1.154067	2.310526
Fold5	80.87(93/115)	59.57(56/94)	71.29(149/209)	19.13(22/115)	40.43(38/94)	28.71(60/209)	1.350239	2.90622
Fold6	75.83(91/120)	62.92(56/89)	70.33(147/209)	24.17(29/120)	37.08(33/89)	29.67(62/209)	1.248325	2.604306
Fold7	73.44(94/128)	71.60(58/81)	72.73(152/209)	26.56(34/128)	28.40(23/81)	27.27(57/209)	0.975598	1.927273
Fold8	84.21(112/133)	44.74(34/76)	69.86(146/209)	15.79(21/133)	55.26(42/76)	30.14(63/209)	1.466986	3.185167
Fold9	76.98(97/126)	62.65(52/83)	71.29(149/209)	23.02(29/126)	37.35(31/83)	28.71(60/209)	1.186124	2.460766
Fold10	82.17(106/129)	62.50(50/80)	74.64(156/209)	17.83(23/129)	37.50(30/80)	25.36(53/209)	1.109091	2.340191
Mean	**76.85(976/1270)**	**60.87(501/823)**	**70.57(1477/2093)**	**23.15(294/1270)**	**39.13(322/823)**	**29.43(616/2093)**	**1.22471**	**2.546445**

Notation: GG refers to actual good cases, predicted as good cases; BB refers to actual bad cases, predicted as bad cases; TE refers to total error rates (Type I *plus* Type II).

Table 8: Cross-validation results for the 10 Probabilistic Neural Networks (PNN) scoring models (hold-out sub-samples)

PNN	Classification results			Error results			Actual Misclassification Costs	
	GG	BB	ACCR%	Type I	Type II	TE	AMC2006 (1.6:6.5)	AMC2011 (1.7:15)
$Fold_1$	94.35(117/124)	63.95(55/86)	81.9(172/210)	5.65(7/124)	36.05(31/86)	18.1(38/210)	1.012857	2.270952
$Fold_2$	79.03(98/124)	54.65(47/86)	69.05(145/210)	20.97(26/124)	45.35(39/86)	30.95(65/210)	1.405238	2.99619
$Fold_3$	70.97(88/124)	47.67(41/86)	61.43(129/210)	29.03(36/124)	52.33(45/86)	38.57(81/210)	1.667143	3.505714
$Fold_4$	68.29(84/123)	63.95(55/86)	66.51(139/209)	31.71(39/123)	36.05(31/86)	33.49(70/209)	1.262679	2.542105
$Fold_5$	74.8(92/123)	62.79(54/86)	69.86(146/209)	25.2(31/123)	37.21(32/86)	30.14(63/209)	1.232536	2.548804
$Fold_6$	72.36(89/123)	59.3(51/86)	66.99(140/209)	27.64(34/123)	40.7(35/86)	33.01(69/209)	1.348804	2.788517
$Fold_7$	76.42(94/123)	59.3(51/86)	69.38(145/209)	23.58(29/123)	40.7(35/86)	30.62(64/209)	1.310526	2.747847
$Fold_8$	76.42(94/123)	61.63(53/86)	70.33(147/209)	23.58(29/123)	38.37(33/86)	29.67(62/209)	1.248325	2.604306
$Fold_9$	68.29(84/123)	47.67(41/86)	59.81(125/209)	31.71(39/123)	52.33(45/86)	40.19(84/209)	1.698086	3.54689
$Fold_{10}$	87.8(108/123)	48.84(42/86)	71.77(150/209)	12.2(15/123)	51.16(44/86)	28.23(59/209)	1.483254	3.279904
Mean	**76.89(948/1233)**	**56.98(490/860)**	**68.71(1438/2093)**	**23.11(285/1233)**	**43.02(370/860)**	**31.29(655/2093)**	**1.366945**	**2.883123**

Notation: GG refers to actual good cases, predicted as good cases; BB refers to actual bad cases, predicted as bad cases; TE refers to total error rates (Type I *plus* Type II).

Table 9: Cross-validation results for the 10 Probabilistic Neural Networks (PNN) scoring models (hold-out sub-samples) random runs

PNNran	Classification results			Error results			Actual Misclassification Costs	
	GG	BB	ACCR%	Type I	Type II	TE	AMC2006 (1.6:6.5)	AMC2011 (1.7:15)
Fold1	79.84(103/129)	59.26(48/81)	71.90(151/210)	20.16(26/129)	40.74(33/81)	28.10(59/210)	1.219524	2.567619
Fold2	81.43(114/140)	55.71(39/70)	72.86(153/210	18.57(26/140)	44.29(31/70)	27.14(57/210)	1.157619	2.424762
Fold3	79.31(92/116)	67.02(63/94)	73.81(155/210)	20.69(24/116)	32.98(31/94)	26.19(55/210)	1.142381	2.408571
Fold4	81.36(96/118)	59.34(54/91)	71.77(150/209)	18.64(22/118)	40.66(37/91)	28.23(59/209)	1.319139	2.83445
Fold5	78.46(102/130)	59.49(47/79)	71.29(149/209)	22.31(29/130)	40.51(32/79)	29.19(61/209)	1.217225	2.532536
Fold6	78.63(92/117)	65.22(60/92)	72.73(152/209)	21.37(25/117)	34.78(32/92)	27.27(57/209)	1.186603	2.5
Fold7	82.81(106/128)	59.26(48/81)	73.68(154/209)	17.19(22/128)	40.74(33/81)	26.32(55/209)	1.194737	2.547368
Fold8	80.00(88/110)	61.62(61/99)	71.29(149/209)	20.00(22/110)	38.38(38/99)	28.71(60/209)	1.350239	2.90622
Fold9	86.21(100/116)	60.22(56/93)	74.64(156/209)	13.79(16/116)	39.78(37/93)	25.36(53/209)	1.273206	2.785646
Fold10	87.90(109/124)	63.53(54/85)	77.99(163/209)	12.10(15/124)	36.47(31/85)	22.01(46/209)	1.078947	2.34689
Mean	**81.60(1002/1228)**	**61.27(530/865)**	**73.20(1532/2093)**	**18.49(227/1228)**	**38.73(335/865)**	**26.85(562/2093)**	**1.213962**	**2.585406**

Notation: GG refers to actual good cases, predicted as good cases; BB refers to actual bad cases, predicted as bad cases; TE refers to total error rates (Type I *plus* Type II).

Table 10: General linear model results for error rates and AMC for different scoring models

Criterion		Sum of squares	df	Mean square	*F*	*P*-value
Type I error	Intercept	48290.755	1	48290.755	390.656	0.000
	Error	1112.532	9	123.615		
Type II error	Intercept	175890.052	1	175890.052	2292.760	0.000
	Error	690.439	9	76.715		
TE	Intercept	90876.017	1	90876.017	2466.941	0.000
	Error	331.538	9	36.838		
AMC 2006	Intercept	175.210	1	175.210	4096.574	0.000
	Error	0.385	9	0.043		
AMC 2011	Intercept	781.822	1	781.822	3279.850	0.000
	Error	2.145	9	0.238		

Table 11: Average variable impact for each variable under each of the scoring models

Model / Variable	DA	LR	MLFN	MLFNran	PNN	PNNran
	AVI	AVI	AVI	AVI	AVI	AVI
AGE	15.142	0.077	7.630	6.872	4.421	5.459
EDU	0.067	0.415	1.959	2.563	1.395	1.579
EMI	0.105	0.324	15.404	14.029	4.574	1.623
GEN	1.183	1.585	2.768	2.363	4.648	4.506
JOB (14)	0.323	0.217	1.941	2.078	0.073	0.075
LAMT (3)	0.110	0.175	11.857	13.492	33.823	36.561
LPRP	0.276	2.255	7.110	6.541	4.424	3.650
MRST (2)	30.068	23.066	11.010	11.029	11.405	11.085
NINC (1)	42.851	50.514	18.571	20.079	18.844	18.887
OTLO	0.866	17.018	9.123	9.280	10.039	10.287
OVEH (12)	0.159	0.271	1.679	1.842	0.856	1.106
PEMP	6.848	0.699	2.604	2.038	0.120	0.102
STATE	1.748	3.287	6.048	5.361	5.322	5.005
TERM (13)	0.255	0.098	2.294	2.434	0.056	0.075
∑	100	100	100	100	100	100

Notation: Each cell represents an average of 10 numbers obtained from 10 scoring models across 10-folds. DA = Discriminant Analysis; LR = Logistic Regression; MLFN = Multi-Layer Feed-Forward Neural Network; MLFNran = Multi-Layer Feed-Forward Neural Network random folds; PNN = Probabilistic Neural Network; PNNran = Probabilistic Neural Network random folds; AVI = Average Variable Impact; AGE = Actual age of the client; EDU = Educational level; EMI = Equated Monthly Instalment; GEN = Gender; JOB = Client current job; LAMT = Actual loan amount in Rupees ₹ crore; LPRP = Loan Purpose; MRST = Marital Status; NINC = Actual Net Income in Rupees ₹ crore; OTLO = Other Loans; OVEH = Vehicle Ownership; PEMP = Previous Employment; STATE = State of residence; TERM = Loan duration.

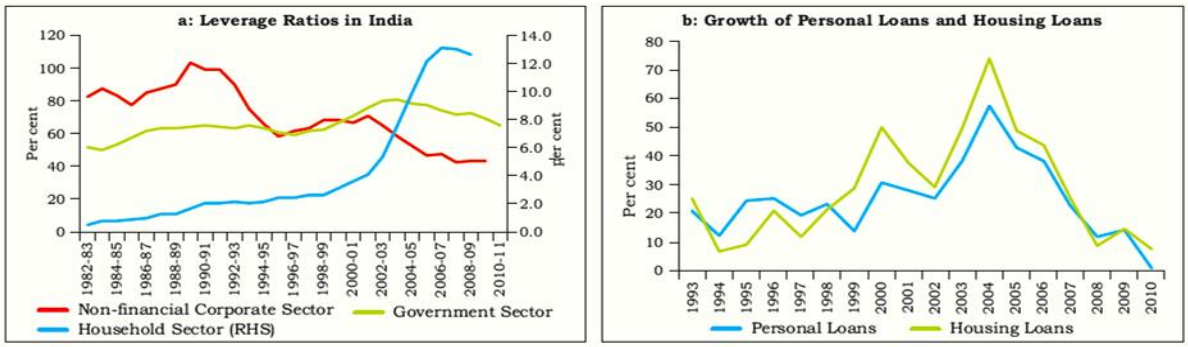
Table 12: Average variable impact for each variable under each of the scoring models for default cases

Model / Variable	DA	LR	MLFN	MLFNran	PNN	PNNran
	AVI	AVI	AVI	AVI	AVI	AVI
AGE	8.301	7.645	11.194	8.682	1.470	3.754
EDU (14)	0.509	0.338	3.118	3.155	2.425	1.259
EMI (2)	8.878	3.015	11.311	11.509	16.984	24.426
GEN	4.590	7.186	2.669	4.013	2.313	1.956
JOB	10.531	10.449	4.380	3.779	2.161	1.202
LAMT (3)	10.754	4.123	10.512	11.612	13.132	3.565
LPRP	1.846	6.958	9.107	8.983	6.912	9.235
MRST	5.935	4.039	5.305	5.379	1.946	8.092
NINC	6.546	1.549	9.659	8.084	2.836	0.690
OTLO	1.061	6.584	6.734	6.603	3.756	4.064
OVEH (12)	3.135	3.614	3.483	3.627	2.034	1.295
PEMP (13)	3.833	1.228	3.363	3.746	1.459	0.274
STATE (1)	32.576	41.856	12.834	14.422	39.621	37.889
TERM	1.505	1.416	6.329	6.406	2.951	2.300
∑	100	100	100	100	100	100

Notation: Each cell represents an average of 5 numbers obtained from 5 scoring models across 5-folds. DA = Discriminant Analysis; LR = Logistic Regression; MLFN = Multi-Layer Feed-Forward Neural Network; MLFNran = Multi-Layer Feed-Forward Neural Network random folds; PNN = Probabilistic Neural Network; PNNran = Probabilistic Neural Network random folds; AVI = Average Variable Impact; AGE = Actual age of the client; EDU = Educational level; EMI = Equated Monthly Instalment; GEN = Gender; JOB = Client current job; LAMT = Actual loan amount in Rupees ₹ crore; LPRP = Loan Purpose; MRST = Marital Status; NINC = Actual Net Income in Rupees ₹ crore; OTLO = Other Loans; OVEH = Vehicle Ownership; PEMP = Previous Employment; STATE = State of residence; TERM = Loan duration.

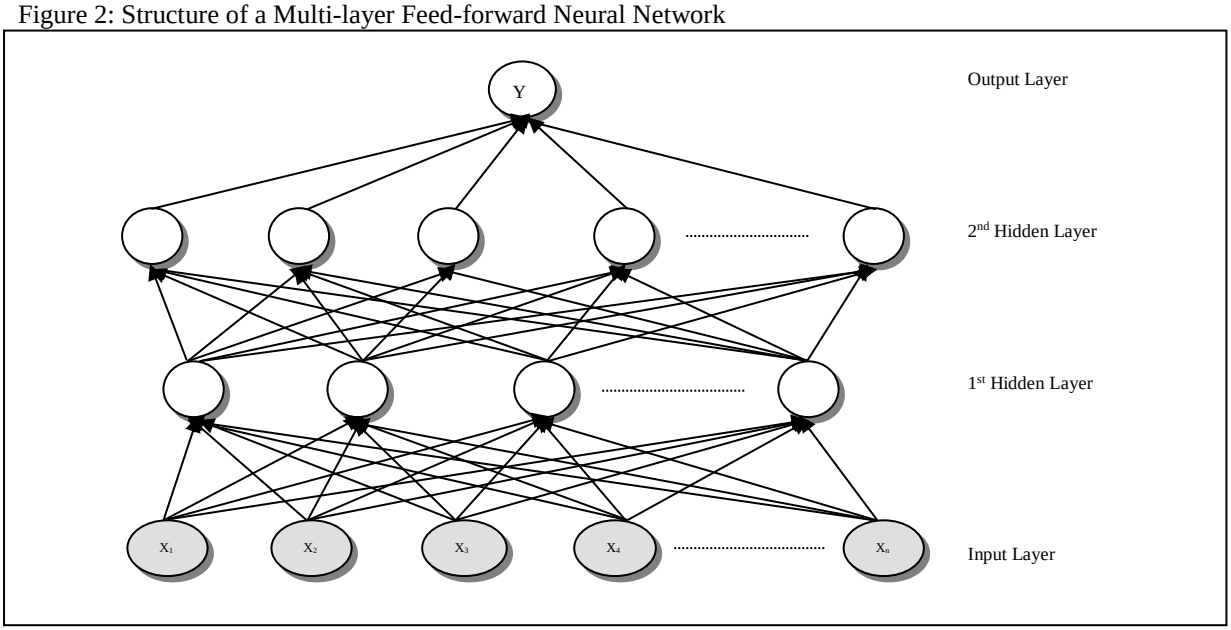
FIGURES

Figure 1: Indebtedness in India



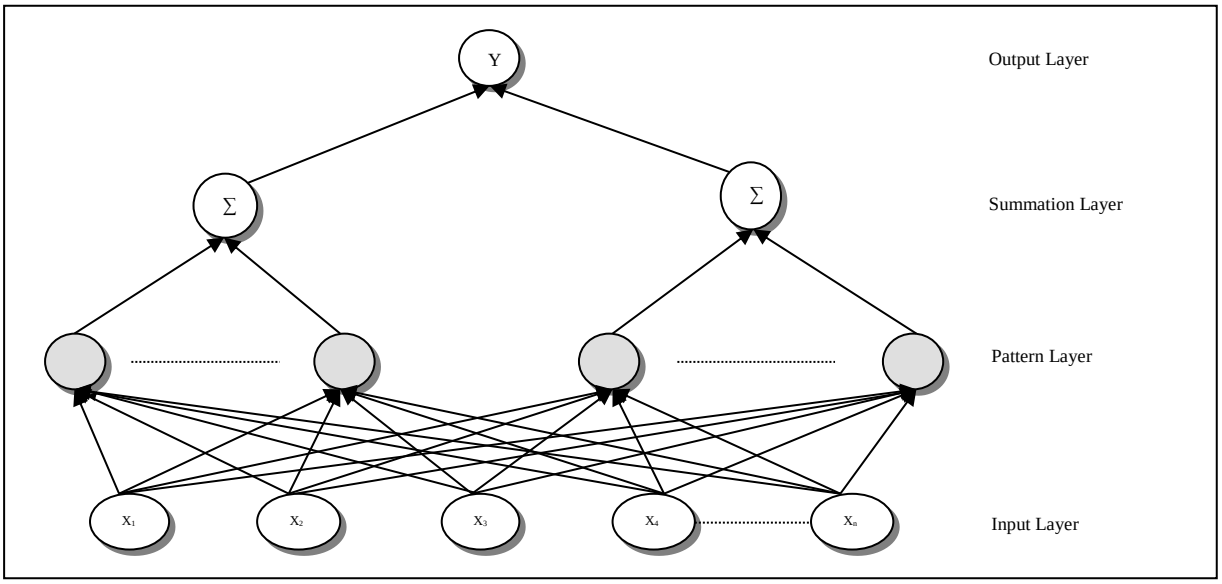
Source: RBI (2014), p. 100.

Figure 2: Structure of a Multi-layer Feed-forward Neural Network



Notation: this Figure presents a structure of a number of independent predictor variables for MLFN. This network is configured to have a larger number of nodes in the second hidden layer compared to the first hidden layer. The output at a given layer (for example, second hidden layer) may be expressed as a connection-weighted summation of outputs from the previous layer (for example, first hidden layer) plus a neuron-bias (a parameter assigned to each neuron). Arriving at a neuron in the output layer, the value from each hidden layer neuron is multiplied by a weight, and the resulting weighted values are added together. Then, a conversion function for the output layer produces *Y* values as outputs of the network (Abdou, 2009a, p. 101).

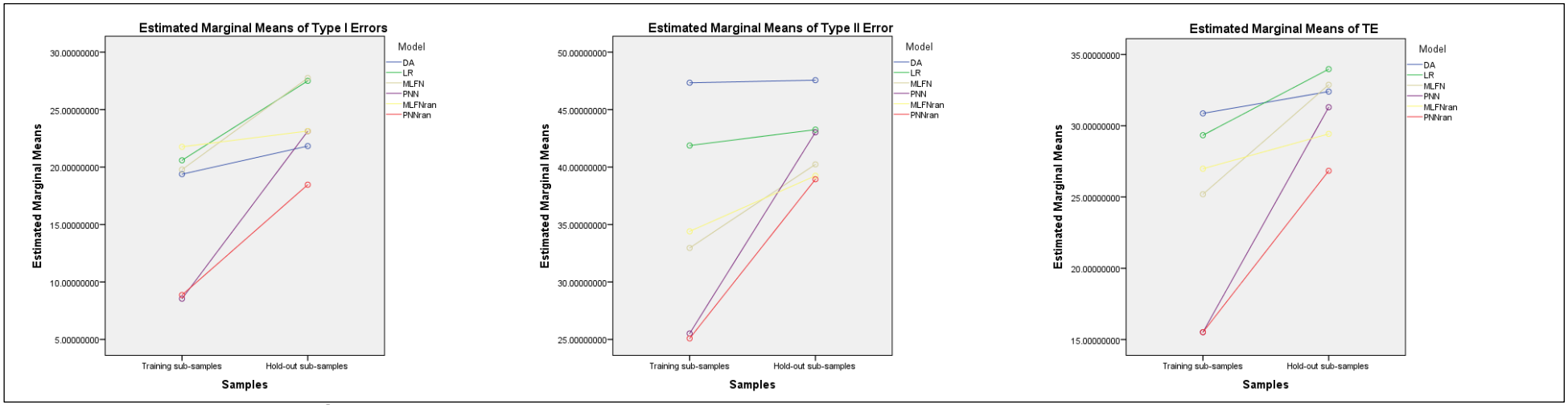
Figure 3: Structure of a Probabilistic Neural Network



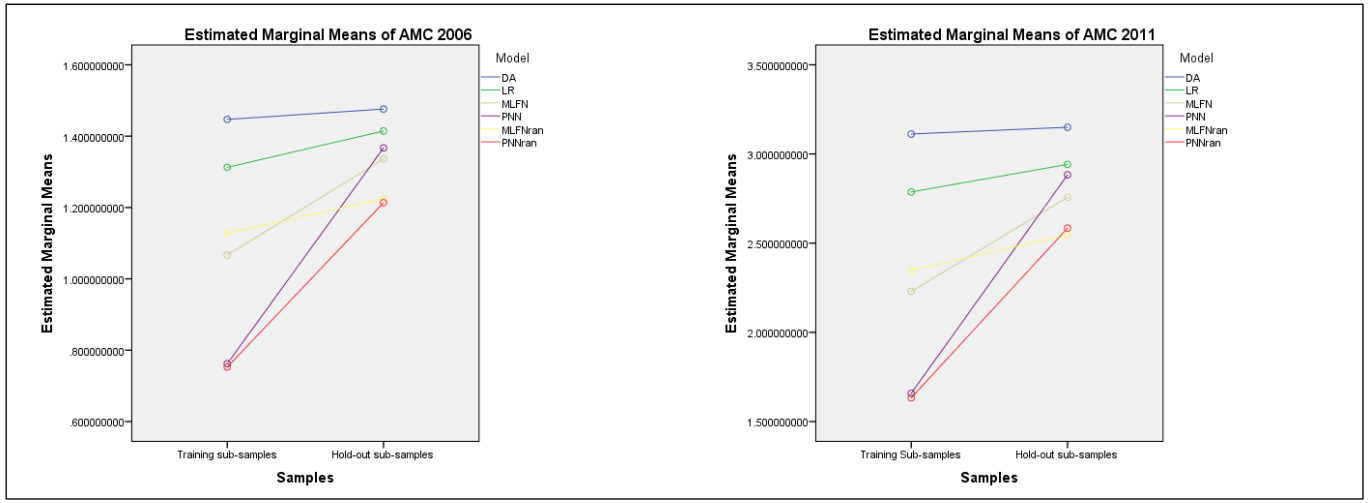
Notation: this Figure presents a structure of a number of independent predictor variables for PNN. Each node in the pattern layer measures the distance between each of the input values and the training values reintroduced by each of the node. Then, each of these values pass to each of the nodes in the summation layer, which is a function of the distance in the smoothing factors. One node per dependant variable is in the summation layer, each node computes a weighted average using the training cases in that category. The summation layer output values can be interpreted as a probability weighting associated with each class. Finally, the output node selects the category with the highest probability weighting as the predicted category (Abdou, 2009a, p. 99).

Figure 4: graphical presentation of the General Linear Model for Type I, Type II, TE, AMC 2006 and AMC 2011

2.a 2.b 2.c



2.d 2.e



Notation: Figures 2.a to 2.e illustrate the General Linear Models for Type I, Type II, TE, AMC 2006 and AMC 2011. The right-hand sides in each sub-figure present the hold-out sub-samples results for different scoring models in contrast to the training sub-samples results on the left-hand sides.