Abdou, Hussein, Pointon, John and El-Masry, Ahmed

Neural Net Credit Scoring Models for Egyptian Banks: An Evaluation of Consumer Loans

# Neural Net Credit Scoring Models for Egyptian Banks: An Evaluation of Consumer Loans


**Hussein Abdou**

*University of Plymouth Business School*


University of Edinburgh Management School
Credit Research Centre
Credit Scoring & Credit Control X:
29th & 31st August, 2007

# Outline

This presentation is organized as follows:

- Part one outlines introduction.
- Part two discusses the literature review.
- Part three details the research methodology and data collection.
- Part four explains the results.
- Part five concludes the study results and suggests the area for the future research.

# Introduction

- Banking operations affect and are affected by the natural environment, and consequently the banks might have an important role to play in helping to raise environmental standards. (Thompson, 1998).

- Egyptian banking sector structure includes: first, public sector banks (6 banks), second, private and joint venture banks (28 banks), and third, branches of foreign banks (7 banks).

- Credit scoring is a quantitative evaluation technique employed by financial institutions “banks” to assess the creditworthiness for both individuals and firms that applies for loans (Long, 1973; Thomas et al.,2002).

# Literature Review

- Relevant literature is classified into:

  - Credit scoring models for consumer loans (Orgler, 1971; Banasik et al., 2001; Sarlija et al., 2004; and Lee and Chen, 2005).
  - Credit scoring models in general: including a comparison between different statistical techniques used in building credit scoring models (Desai et al., 1996; Chen and Huang, 2003; and Ong et al., 2005).

- The chosen environment will be the Egyptian banking sector, in which no other studies (in the best of our knowledge) have investigated the use of sophisticated statistical appraisal techniques in credit scoring.

# Methodology

In this part, statistical techniques used in building the scoring models are as follows:

**Conventional techniques:**

- Discriminant analysis model (DA).
- Probit analysis model (PA).
- Logistic regression model (LR).

**Advanced techniques:**

- Probabilistic neural nets (PNNs).
- Multi-layer feed-forward nets (MLFNs).
- Best net search (BNS).

# Conventional techniques:

**First is the DA,** which was first proposed by Fisher (1936) as a discrimination and classification technique.

The general formula of DA is as follows:

$$Z = \alpha + \delta_1 V_1 + \delta_2 V_2 + \ldots\ldots\ldots\ldots\ldots\ldots + \delta_n V_n,$$

where

Z represents the discriminant (zed) score, $\alpha$ is the intercept term, and $\delta_i$ represents the respective coefficient in the linear combination of explanatory variables, $V_i$, for $i = 1$ to $n$ (Lee et al., 2002).

**Second is the PA,** which is also usually used for the purpose of comparing the results (Maddala, 2001; and Pindyck and Rubinfeld, 1997).

The general formula of DA is as follows:

$$\text{Prob } (y = 1) = \Phi\ (\alpha + \delta_1 V_1 + \delta_2 V_2 + \ldots\ldots\ldots\ldots\ldots\ldots + \delta_n V_n),$$

where

y is the zero-one binary outcome for a given set of value. $\Phi$ is the value from the cumulative normal distribution function. $\alpha$ is the intercept term, and $\delta_i$ represents the respective coefficient in the linear combination of explanatory variables, $V_i$ , for i = 1 to n (Abdou et al., 2007).

**Third is the LR,** unlike other statistical techniques, is more suitable for credit scoring problems (Lee and Chen, 2005).

The general formula of DA is as follows:

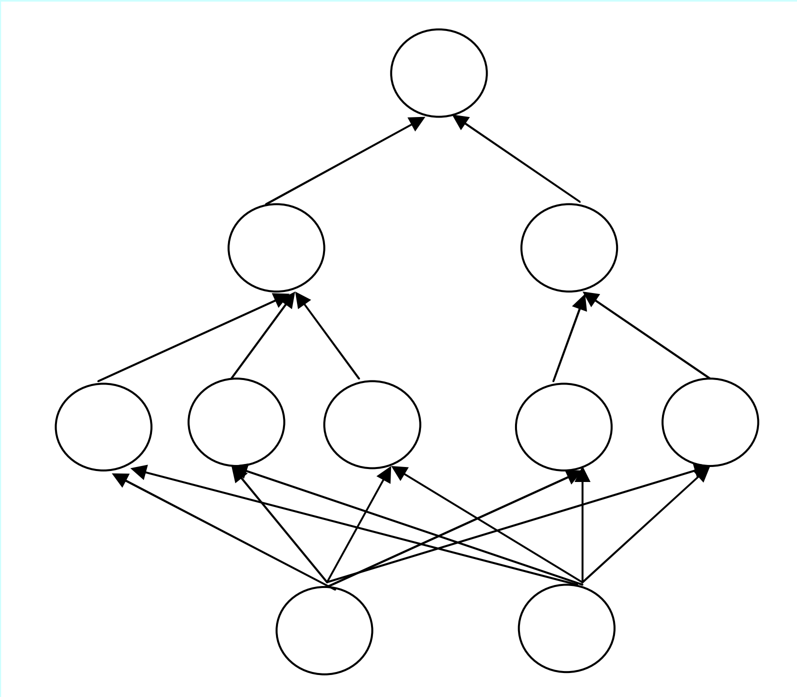
$$\log[p/(1-p)] = \alpha + \delta_1 V_1 + \delta_2 V_2 + \ldots\ldots\ldots\ldots\ldots\ldots + \delta_n V_n,$$

where,

p is the probability of the outcome of interest, α is the intercept term, and $\delta_i$ represents the respective coefficient in the linear combination of explanatory variables, $V_i$ , for i = 1 to n. The dependent variable is the logarithm of the odds, {Log [p/ (1-p)]}, which is the logarithm of the ratio of two probabilities of the outcome of interest (Lee et al., 2002).

# Advanced techniques

## First PNN,



Output Layer

Summation Layer (one neuron per category)

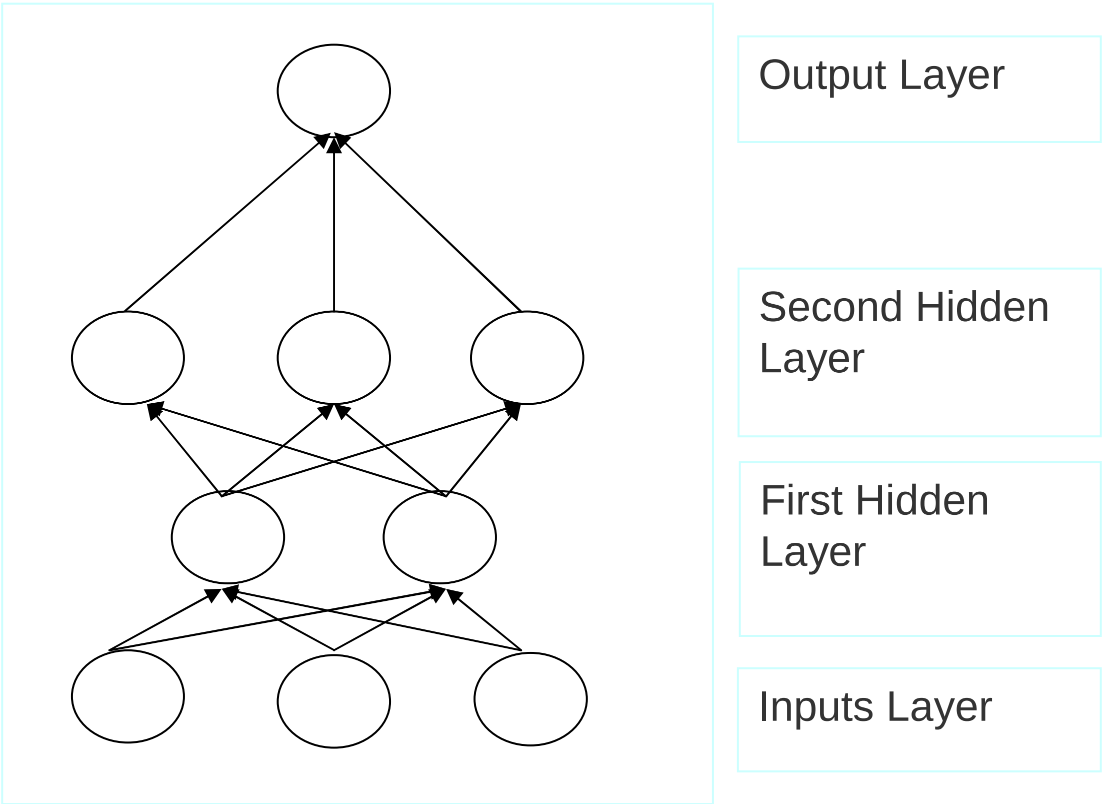
Pattern Layer (one neuron per training case)

Inputs Layer

Basic PNN structure (source: Palisade, 2005)

An implementation of statistical techniques, called kernel discriminant analysis, in which the processes are structured into a multi-layered feed forward net with four layers, is a probabilistic neural net (Bishop,1995).

# Second MLFN with four nodes,




Basic MLF nets architecture (source: Palisade, 2005)

It may be advisable to model a system using multi-layer feed-forward nets (multi-layer perceptron networks), in situations of complex relationships between variables (Masters, 1995; Palisade, 2005).

Third BNS,

- From multi-layer feed-forward net with two to six nodes and from probabilistic neural net as well, was an option selected in the current Neural Tools package.

- Particular attributes of <u>multi-layer feed-forward nets</u> include reliability outside the training data range, compactness in size, an excellent classifier, and with a capability to generalize results from small training data. By contrast, <u>probabilistic neural nets</u> are particularly fast, they do not require a number of hidden layers and nodes, they have a parallel structure, and they classify and return probabilities for different dependent categories, and guarantee convergence to the optimal case (Masters, 1995; Palisade, 2005).

# Data Collection

- In order to build the proposed credit scoring models, a personal loans data-set was provided by one of the commercial banks in Egypt. This consists of 581 personal loans with 433 good loans and 148 bad loans.

- Each bank customer in this data-set is linked to 20 independent variables, in addition to the dependent variable, Selected variables for the proposed models were reduced to 12 variables, as follows: loan amount, company, gender, marital status, age, monthly salary, additional income, house status, telephone, education, loans from other banks, and corporate guarantee.

# Results

## Whole Sample Credit Scoring Models

Classification results for DA, $DA_1$, PA, $PA_1$, LR, $LR_1$, PNN, and MLFN

| Model | ACC Rate % | Model | ACC Rate % |
|---|---|---|---|
| DA | | $DA_1$ | |
| Good | 85.91 | Good | 85.91 |
| Bad | 89.19 | Bad | 89.86 |
| Total | 86.75 | Total | 86.92 |
| PA | | $PA_1$ | |
| Good | 94.00 | Good | 93.07 |
| Bad | 69.59 | Bad | 70.27 |
| Total | 87.78 | Total | 87.26 |
| LR | | $LR_1$ | |
| Good | 94.00 | Good | 93.76 |
| Bad | 71.62 | Bad | 70.95 |
| Total | 88.30 | Total | 87.95 |
| PNN | | MLFN | |
| Good | 97.92 | Good | 96.07 |
| Bad | 88.51 | Bad | 85.81 |
| Total | 95.52 | Total | 93.46 |

Cut-off point 0.50, only applicable for conventional techniques.

Models in this section were built using the whole sample, including the neural net models.

# Validated Scoring Models

Classification results using the RDA, RPA,RLR, RPNN and RMLFN : predictions (in columns) versus observations (in rows).

| Sample<br>Model | Hold-out sample<br>ACC Rate % | Training sample<br>ACC Rate % | Overall sample<br>ACC Rate % |
|---|---|---|---|
| RDA | | | |
| Good | 87.95 | 86.00 | 86.37 |
| Bad | 81.82 | 88.70 | 87.16 |
| Total | 86.21 | 86.67 | 86.57 |
| RPA | | | |
| Good | 95.18 | 94.00 | 94.23 |
| Bad | 45.45 | 72.17 | 66.22 |
| Total | 81.03 | 88.60 | 87.09 |
| RLR | | | |
| Good | 95.18 | 93.71 | 94.00 |
| Bad | 42.42 | 74.78 | 67.57 |
| Total | 80.17 | 89.03 | 87.26 |
| RPNN | | | |
| Good | 96.39 | 98.00 | 97.69 |
| Bad | 57.58 | 81.74 | 76.35 |
| Total | 85.34 | 93.98 | 92.25 |
| RMLFN | | | |
| Good | 95.18 | 99.14 | 98.38 |
| Bad | 42.42 | 76.52 | 68.92 |
| Total | 80.17 | 93.55 | 90.88 |

Cut-off point 0.50, only applicable for conventional techniques.

For the purpose of making a fair comparison between conventional techniques and neural net techniques, and to avoid a sample bias which might happen in the above whole sample analysis, we apply a simple validation technique by dividing the data-set into a training sample and a hold-out (testing) sample.

# Powerful Neural Net Models

Classification results for the best five PNNs, MLFNs and BNSs

| Sample Neural Net | Hold-out sample | Training sample | Overall sample |
|---|---|---|---|
| | ACC Rate % | ACC Rate % | ACC Rate % |
| PNN$_2$ | 90.52 | 95.91 | 94.84 |
| PNN$_7$* | 90.52 | 97.63 | 96.21 |
| PNN$_{11}$ | 89.66 | 96.77 | 95.35 |
| PNN$_{14}$ | 85.34 | 97.63 | 95.18 |
| PNN$_{15}$ | 84.48 | 97.85 | 95.18 |
| MLFN$_2$ | 81.90 | 95.91 | 93.12 |
| MLFN$_4$ | 84.48 | 95.70 | 93.46 |
| MLFN$_9$* | 86.21 | 95.91 | 93.98 |
| MLFN$_{12}$ | 85.34 | 95.27 | 93.29 |
| MLFN$_{20}$ | 84.48 | 95.27 | 93.12 |
| BNS$_1$-PNN | 84.48 | 97.42 | 94.84 |
| BNS$_8$-PNN* | 88.79 | 96.56 | 95.01 |
| BNS$_{16}$-MLFN-5N | 83.62 | 97.63 | 94.84 |
| BNS$_{18}$-MLFN-5N | 89.66 | 95.91 | 94.66 |
| BNS$_{19}$-MLFN-5N | 85.34 | 97.20 | 94.84 |

*The best model under each net is highlighted.

Looking for a powerful model using the neural nets' capabilities, we ran the PNNs and the MLFNs again with the BNSs. The experiment was repeated 20 times with a different hold-out (testing) sub-sample each time and the remaining data-set was the training sample. The reason for repeating the process was to investigate whether different results, in terms of average correct classification rate, were being achieved because of the ***random*** selection procedure as part of the software design.

For the purpose of comparing results of all models developed in this research, and in order to evaluate the overall credit scoring capability and effectiveness:

- The misclassification costs, besides average correct classification rates, have been taken into account, in order to find the minimum expected misclassification cost in a credit scoring model (West, 2000). Since the cost associated with type I errors differ from those associated with type II errors.

# Comparison of results of different credit scoring models

Comparing ACC Rates, and estimated misclassification costs for the selected techniques

| Scoring model | ACC Rate % (Overall sample) | Estimated misclassification cost |
|---|---|---|
| DA | 86.75 | 0.2428 |
| $DA_1$ | 86.92 | 0.2343 |
| RDA | 86.57 | 0.2653 |
| PA | 87.78 | 0.4324 |
| $PA_1$ | 87.26 | 0.4307 |
| RPA | 87.09 | 0.4737 |
| LR | 88.30 | 0.4065 |
| $LR_1$ | 87.95 | 0.4169 |
| RLR | 87.26 | 0.4582 |
| PNN | 95.52 | 0.1620 |
| RPNN | 92.25 | 0.3187 |
| $PNN_7$* | 96.21 | 0.1482 |
| MLFN | 93.46 | 0.2102 |
| RMLFN | 90.88 | 0.4083 |
| $MLFN_9$ | 93.98 | 0.2188 |
| $BNS_8$-PNN | 95.01 | 0.1740 |

*Best model amongst all models.

Note: for all models Type II errors exceeded Type I errors except for the first three conventional models: DA, $DA_1$, RDA.

# Powerful Neural Net Models

There is evidence of significant differences between the neural nets models.

A comparative statistical evaluation of the powerful neural net models

| | NN Models [comprising 20 PNN, 20 MLFN, and 20 BNS] |
|---|---|
| Count | 60 |
| Average (Mean) | 93.4653 |
| Standard deviation | 1.13189 |
| ANOVA F-Ratio | 12.73*** |
| Fisher's least significant difference test: | |
| PNN-MLFN | 1.5225** |
| PNN-BNS | 0.6530** |
| MLFN-BNS | -0.8695** |
| | |
| Cochran's C Test: | 0.526505* |
| Bartlett's Test: | 1.10247* |
| Levene's Test: | 3.42494** |
| | |
| Kruskal-Wallis Median Test Statistic: | |
| Average Rank | - |
| Test Statistic | 19.8774*** |

*, **, and ***denotes a statistically significant difference at 10, 5 and 1 per cent level, respectively.

# Conclusion and Area of Future Research

- This study presents an evaluation of personal loans to help strengthen the credit risk evaluation process in the Egyptian banking sector using four credit scoring statistical techniques: DA, PA, LR, and NNs.

- As to the ranking of the models using the highest average correct classification rate, $PNN_7$ is preferred. This model also had the lowest misclassification cost amongst all selected models.

- Although when the analysis was extended to all powerful NN trials, the ranking changed and $BNS_{19}$-MLFN-5N was chosen according to the lowest misclassification cost.

## Conclusion continued

- Results so far have revealed that NNs models gave a better average correct classification rate than the conventional techniques.

- Future studies should aim to use other advanced statistical scoring techniques, such as genetic algorithms, besides the neural nets and traditional scoring models which were used in the current paper, and perhaps integrated with other techniques, such as fuzzy discriminant analysis.

- In addition to this, the plan is to collect more data and employ more variables that might increase the accuracies of the scoring models. Finally, future research would use more than one bank's data-set.