A comparison of corporate distress prediction models in Brazil:
hybrid neural networks, logit models and discriminant analysis

Juliana Yim
School of Economics and Finance, RMIT University
Heather Mitchell
School of Economics and Finance, RMIT University

Key words
hybrid neural networks, corporate failures.

JEL Classification G33, C45.

Abstract

This paper looks at the ability of a relatively new technique, hybrid ANN's, to predict corporate distress in Brazil. These models are compared with traditional statistical techniques and conventional ANN models. The results suggest that hybrid neural networks outperform all other models in predicting firms in financial distress one year prior to the event. This suggests that for researchers, policymakers and others interested in early warning systems, hybrid networks may be a useful tool for predicting firm failure.

Palavras-chave
redes neurais híbridas, falência de empresas.

Classificação JEL *G33, C45.*

Resumo

O presente artigo analisa o desempenho das redes neurais híbridas para prever falência de empresas no Brasil. Esta nova técnica foi comparada com modelos estatísticos tradicionais. Os resultados sugerem que as redes neurais híbridas são superiores as técnicas estatísticas um ano antes do evento. Isto sugere que para pesquisadores, políticos e outros interessados em "early warning systems"*, redes neurais híbridas podem ser uma poderosa alternativa para prever falência de empresas.*

1_ Introduction

One of the main challenges that Brazilian policymakers have been facing is promoting sustainable economic growth. Several economic plans with the aim of controlling the high level of inflation have been developed and adopted during the last two decades, but they lost their credibility during their implementation phase. This provoked an environment of instability in the Brazilian economy. As a consequence, it has been a very difficult task for Brazilian firms to generate their profit margins, to obtain credit and to meet their obligations. Therefore, in order to prevent the failure of Brazilian firms, researchers and policymakers have demonstrated a great interest in investigating the indicators of corporate distress to develop early warning systems.

Most of the previous literature on failures has concentrated on the US and European region. We chose to look at Brazilian firms in financial distress only after 1999 because the 1999 Brazilian currency crisis gave us a good size data set. This paper looks at a relatively new technique, the hybrid neural network, to predict firm failures in Brazil. This type of network is formed by integrating the variables selected by the statistical models and the outputs of statistical models with those of an ordinary neural network to create hybrid models that might be more accurate than either of the techniques used separately. The first objective of this study is to investigate whether hybrid artificial neural networks can outperform traditional statistical models and ordinary ANNs for predicting Brazilian firm failures one year prior to the financial distress. The second objective is to investigate what are the main determinants of Brazilian failures. The third objective of this study is to extend ANN failure prediction literature to non-US countries, as most previous work has used US companies.

Prior studies indicate that researchers generally test and evaluate corporate financial distress models using two popular standard statistical techniques, logit and discriminant analysis (DA). Altman (1968 and 1977), Deakin (1972) and Blum (1974) employed DA, and Ohlson (1980) and Gentry *et al.* (1985) used logit models. Few studies in Brazil investigated the predictive power of statistical models for predicting firm failures. Elizabetsky

(1976) used DA on a sample of 99 Brazilian firms that failed and 274 non-failed firms. The best model correctly classified 63% of the failed firms and 74% of the non-failed firms. Matias (1978) applied DA to a sample of 100 Brazilian firms of which 50 were failed. The best model correctly classified 77% of the failed firms and 70% of the non-failed firms one year prior to the event. Altman *et al.* (1979) used DA on a sample of 23 Brazilian firms that failed during 1975-1977 and 35 non-failed firms. The best model correctly classified 83% of the failed firms and 77% of the non-failed firms. Siqueira and Matias (1996) applied the logit model to a sample of 16 Brazilian banks that failed during 1994-1995 and 20 non-failed banks. The best model correctly classified 87% of the failed banks and 95% of the non-failed banks. Minardi and Sanvicent (1998) applied DA to a sample of 81 Brazilian firms of which 37 failed during 1986-1998. The best model correctly classified 81% of the failed firms and 80% of the non-failed firms one year prior to the event.

DA models require assumptions such as the independent variables must have a multivariate normal distribution and the variance-covariance matrix of the independent variables in each of the groups (failed and non-failed) must be the same Dillon (1984). Logit and Probit models are alternative methods of classification that can be used when the multivariate normal model is not justified. Logit and Probit models are less affected than DA by the covariance inequalities across the groups. When the data do not satisfy these assumptions, both logit and DA provide non-optimal solutions (Altman *et al.*, 1977; Ohlson, 1980). On the other hand, non-parametric and non-linear models, such as artificial neural networks (ANNs), do not rely on these assumptions that are often adopted to make traditional statistical methods tractable. No study known to the authors has used ANN for predicting corporate failure in Brazil.

Some work has been done using ANNs for markets outside Brazil, most notably the US. Most of the papers that have been published on the comparison of ANNs and statistical models for firm failure prediction indicated that ANNs outperformed statistical models (Odom and Sharda, 1990; Coats and Fant, 1993; Wilson and Sharda, 1994). One study presented findings that were contrary to those established by previous comparative studies. This was the study by Altman *et al.* (1994), which compared ANN models with DA for Italian industrial firms using a large sample size. The results indicated that DA outperformed ANN models. Altman *et al.* (1994) suggested that the performance of neural network might be improved by integration with other methods for optimizing the topology of the neural networks.

A considerable amount of architectures and methods for optimizing both the size and generalization capabilities of neural networks could have been proposed in order to solve the problem of the choice of the topology in the application of the neural networks. Among these techniques, there are the pruning methods (Weigned and Neuneier, 1995) and hybrid neural networks (Han *et al.*, 1996). The most popular pruning method is the simple weight elimination. This method aims to reduce the size of the neural network and preserve generalization (Weigend *et al.*, 1991; Bebis, 1997; Cunha, 2000). The alternative pruning method consists of removing unnecessary nodes and weights without losing its generalization capacity, using genetic algorithm (Miller *et al.*, 1989, Bebis *et al.*, 1997; Yao 1997). There are several papers that combine ANN with pruning methods for predicting firm bankruptcy (Back, 1996).

However, the authors are interested in integrating ANN with statistical models because we believe that using the same variables as a linear and parametric model, such as DA, or a non-linear and parametric model, such as the logit model, as inputs to a non-linear and non-parametric model, such as ANN, may not only reduce the size of the networks, but may also improve the classification results. In addition, when the results from the DA, the logit and ANN models are compared, they can misclassify different cases. So each approach would be using the information in different ways. In this case, the combination of statistical models and ANN may produce better results.

It has been suggested that combining forecast models should become part of the mainstream of forecasting practice to achieve more accurate results (Clement, 1989; Chandler, 1990). Given that the neural network is considered to have great potential as a powerful forecasting instrument, its combination with other statistical techniques should improve the overall performance of integrating neural networks with statistical techniques to predicting business failures.

There are few time series papers that discuss the integration between statistical models. Tang *et al.* (1991) compared neural networks and Box & Jenkins models (Box & Jenkins, 1976), using international airline passenger traffic, domestic sales and foreign car sales in USA. They found that Box & Jenkins results were superior to ones from the ANN models in short-term. However, ANN models were superior to Box & Jenkins in the long-term. Lachtermacher and Fuller (1995) used Box and Jenkins to identify the lag components of the series, to determine a compact network structure. The series used were annual river flow and annual electricity consumption. This model minimized the size of the network and the data required to train the network. The results showed that hybrid network is a very helpful instrument for time series prediction. Drossu and Obradovic (1996) investigated the possibility of integrating stochastic models and ANN models in order to speed up the design process of an appropriate predictor. They used entertainment video traffic data. They found that the stochastic models provided useful information on selecting the number of inputs and initial weights of the neural network.

There are few studies in Brazil that investigated the integration between statistical models and ANN. Portugal (1995) compared Box & Jenkins, Structural models and ANN models for predicting the monthly gross output of Rio Grande do Sul (Brazil). The components of the series identified by Box and Jenkins model and from the Structural models (Harvey, 1993) were used to provide insights for the architecture of the network. Andrade *et al.* (1999) used a hybrid ANN, which integrates the Box & Jenkins methods and ANN for a stock market time series prediction. The data used were the stock market of time series of the Brazilian Telecommunication Company (TELEBRAS). The Box and Jenkins was used to identify the lag components of the series, to determine a compact network structure. This combination can support the neural network design in order to obtain more accurate predictors.

Han *et al.* (1996) introduced hybrid neural networks which combine neural network models with other statistical or artificial intelligence models, and found that hybrid neural network models are very powerful for bankruptcy prediction. Markham and Ragsdale (1995) combined output estimated by DA with ANN and concluded that the hybrid network performed better than the DA and ANN used individually. The general conclusion of the above literature is that neural networks have potential as a forecasting tool and their integration with other statistical techniques might improve their overall performance. The authors are not aware of any study which has used ANN for predicting Brazilian firm failures.

This paper is organized as follows. Section 2 introduces the hybrid neural networks. Section 3 describes the data. Section 4 presents an evaluation and a comparison of neural network models, hybrid models and statistical models. Section 5 presents concluding comments.

2_ Hybrid neural network model

This section introduces a method, hybrid neural network to be applied to the classification problems by integrating the variables selected by the statistical models and the outputs of statistical models with those of an ordinary network to create hybrid models that might be more accurate than either of the techniques when considered individually. The ordinary ANN used in this study uses a multilayer perceptron network (MLP), trained by a gradient descent algorithm called backpropagation (Rumelhart, 1986). The MLP is the most common type of formulation and is used for problems that involve supervised learning.

The neural network "learns" by adjusting its connections weights. These weights are adjusted to minimise the mean square errors between the actual and predicted output values. Using backpropagation[1] the weights between the last two layers are updated first, then the next two layers, until the first layer is reached.

The standard ANN has two main problems when dealing with a large number of variables. These are the time taken for selection and the possibility of overfitting. By combining statistical models with ANNs to create hybrid models, these problems are reduced and the classification accuracy may improve in the following ways:

_ Using statistical models to preselect variables reduces the risk of overfitting and also reduces the time taken to select the model.

[1] Details of the backpropagation method see Appendix A.

_ Using output from a statistical model as input to an ANN efficiently condenses information.

2.1_ Types of hybrid neural networks

In this study we will consider three different approaches to hybrid models. The first approach uses statistical models to select the variables to be used as inputs to the ANN. The second is to use statistical output, such as an estimated probability, as an input to a neural net. The last approach is to use both statistical model to select the variables and an output estimated by a statistical model as an input to the ANN. The probit models were not used for the hybrid models because the estimated probit models used the same variables as the logit models. Also, both models misclassified the same firms and the estimated probabilities of failure were almost the same. This is not unexpected, as results from these two models are unlikely to differ for the sample size used here (Madalla, 1983). So it was decided to consider only the probability of failure from the logit models to build the hybrid networks.

To denote the various hybrid models we use the following notation. If a statistical model is used as a pre-processor for selecting the variables we add its name to ANN. So ANN-DA is a hybrid artificial neural network that used Discriminant Analysis to select the input variables. When we wish to indicate the probability from a statistical model is used as input, we put a "P" before its name before adding it to ANN. The two effects can also be used in combination, so ANN-Logit-PDA is a hybrid ANN which used a logit model to preselect the variables and has the probability from a DA model as an additional input.

2.2_ Formulation of hybrid neural networks

Following Markham and Ragsdale (1995), the output from a neural network can be written as:

$$P(ANN) = f(x_1, x_2, \dots, x_m)$$

where $P(ANN)$ is the probability estimated by the ANN and $(x_1, x_2, \dots, x_m)$ are the inputs.

The output from the hybrid neural network (ANN_H) which uses for example the probability estimated by DA as new inputs to the network, can be written as follows:

$$P(ANN_H) = f(x_1, x_2, \dots, x_m, PDA)$$

where PDA, $P(ANN)$ and $P(ANN_H)$ are the probabilities estimated by DA, ANN and ANN_H models, respectively.

The hybrid ANNs were training using the backpropagation algorithm which was described in the Appendix A.

3_ Description of the data

The objective of this study is to compare DA, logit and probit models and ANN for predicting Brazilian firm failures. The sample consists of a total of 121 companies, 29 of which were in financial distress between 1999 and 2000. There are 4 failed firms from the mining industry, 7 from the energy industry, 5 from the construction industry, 4 from the machinery and equipments industry, 4 from the telecommunication industry, 2 from the textile industry and 3 from the food industry.

There are 22 ratios[2] available for the types of company used here and all of those were considered when setting up the failure prediction models. As can be observed in Table 1 of the Appendix B, these variables are classified in four standard ratios categories, profitability, liquidity, asset utilization and structure. Not all variables were available for all companies. It was decided to use financial ratios rather than absolute account data to remove the size effects of the firms. Table 1 show the descriptive statistics of the financial ratios. The Jarque-Bera statistic indicates that the most of the financial ratios are non-normal. Only the solvency ratio is normal.

[2] Values of ratios for each company can be obtained from the authors by request.

Table 1_ Descriptive statistics of financial ratios

Financial ratio	Failed companies			Non-failed companies		
	Mean	SD	JB	Mean	SD	JB
Return on Shareholders Funds (%)	–76.76	79.23	10.16	16.2	22.45	6910
Return on Total Assets (%)	–17.31	15.2	99.09	7.21	7.11	820.19
Return on Capital Employed (%)	–8.29	22.14	24.94	19.22	16.46	361.58
Current ratio	0.68	0.38	1.92	2.11	5.49	27693
Net Assets Turnover	0.74	0.69	10.25	1.09	1.03	131.47
Shareholders Liquidity ratio	1.37	2.07	244.9	7.25	18.75	6834.6
Solvency ratio (%)	27.67	18.93	0.21	52.89	17.92	1.23
Gearing (%)	2.55	4.56	285.5	0.64	0.82	735.18

The successful firms were matched to the failed firms by randomly selecting firms with same asset size. The ratios for the same year as the failed firms were used. After the initial groups are defined and firms were selected, balance sheet and income statement data were collected. The information was obtained from the CD-Osiris database 2002[3]. Data used for the failed firms is from the last financial statements issued before the firms failed. The firms were classified as failed based on the information provided by the Troubled Company Reporter collected from www.bankrupt.com.

The prediction model was applied to an independent sample of failed and non-failed firms to test the validity of the model. We used a total of 29 Brazilian firms, 9 of which were in financial distress between 1999 and 2000.

[3] This database covers 90 countries and includes 22,000 publicly listed companies. The information from in Osiris is obtained from several sources such as Fitch, Reuters News, Standard & Poor's, Mood's, JCF International, Bridge Information Systems and Bureau van Dijk.

[4] Estimated using the software package SPSS 11.5.

Table 2_ The test of equality of group means

Variables	Wilks' lambda	F	Sig.
Return on capital employed	0.697	51.734	0.000
Solvency ratio	0.737	42.520	0.000
Gearing	0.889	14.851	0.000

4_ Empirical investigation: Brazilian corporate failure models

4.1_ Discriminant analysis

The first part involves estimating the discriminant function and determining whether or not they are statistically significant. The second part evaluates further the predictive accuracy of the discriminant function.

First we estimate an optimum discriminant function and determine whether or not it is statistically significant. Several different models[4] were estimated by experimentation as the automatic stepwise procedure did not provided good results. So, in this study, the criteria for selecting the final variable set were based on the high significance of the variables for the model, the correct sign of the coefficient in the model and a high level of prediction accuracy for the training sample. The best model was found to contain three variables: returns on capital employed, solvency and gearing. The model is given below.

$$Z = 2.45 - 0.046RCE - 0.041SOL + 0.041GEA$$

The test of equality of group means is shown in the Table 2 the Wilks' lambda statistic indicates that returns on capital employed, solvency and gearing are very significant.

The assumption of equal covariance matrices is tested using Box's M[5] test, which tests the equality of the determinants. The value of this statistic is M = 229.891, p-value = 0.000 using the F approximation, so the assumption of constant variance is not satisfied. This means that the tests of model adequacy which follow may not be reliable. Next we carry out various tests on the adequacy of the model. The canonical correlation is used because it is a measure of the degree of association between the discriminant scores and the groups. The discriminating function for the best model is highly significant and displays a canonical correlation of 0.715. One interprets this correlation by squaring it $(\text{canonical correlation})^2 = 0.51$, and concluding that 51% of the variance in the dependent variable can be accounted for by this model. The Wilks lambda test is used for analysing the significance of each discriminant function and produced a score of 0.489, which indicates that the selected discriminant function is significant.

The second part consists of evaluating the predictive accuracy of the discriminant function. For this study, the critical cutting score is zero, so a firm is classified as non-failed if its discriminant score is negative and as failed if its discriminant score is positive. The model correctly classified 79.3% of the failed firm and 91.3% of the non-failed firms. For the holdout sample, the success rate of predicting failure was 67.0% and that of success, 100.0%. The test for the statistical significance of the hit rate[6] shows that the results in the holdout sample are satisfactory at level of 5%. So, the DA model produced satisfactory results.

4.2_ Logit and probit models

This section analyses the predictive ability of logit and probit model for Brazilian firm failures. Several different models[7] were estimated by experimentation, using maximum likelihood. The best variables were selected using the same criteria as for DA. The variables selected were returns on capital employed, net asset turnover, solvency and gearing. The variables which are significant at a level of 5%, and the final logistic and probit functions are given by:

5 For full details of this procedure see Norusis (1990).

6 For full details of this test see Franses (2000).

7 Estimated using the software package Eviews 4.0.

$$Pr\,(failure) = \frac{1}{1+e^{-Z_i}}$$

$$Z_i = \underset{(0.0001)}{0.191RCE} - \underset{(0.0043)}{0.026SOL} + \underset{(0.0084)}{0.598GEA} + \underset{(0.0347)}{0.824NAT}$$

$$Pr(failure) = \int_{-\infty}^{X_i\beta} \frac{1}{\sqrt{2\pi}} exp\left(\frac{-z^2}{2}\right) dz$$

$$X_i\beta = \underset{(0.000)}{-0.097RCE} + \underset{(0.0099)}{0.331SOL} + \underset{(0.0164)}{0.509GEA} - \underset{(0.0005)}{0.0017NAT}$$

Note: Figures in brackets are p-values.

The Hosmer and Lemeshow test (1989) was used to compare the fitted expected values to the actual values by group. If these differences are large, we reject the model as providing an insufficient fit to the data. For the logit and probit models, the Hosmer and Lemeshow goodness-of-fit test have a p-value of 0.48 and 0.34, respectively, which is greater than 0.05, implying that the model's estimates fit the data to an acceptable level.

For the both these models, the institutions were classified as failed if the probability of failure exceeds 0.5. The logit and probit models produced very similar values for the probabilities and classified all firms into the same categories. The model correctly classified 72.4% of the failed firms and 97.8% of the non-failed firms. For the holdout sample, the success rate of predicting failure was 67.0% and that of success, 95.0%. The test for the statistical significance of the hit rate shows that the results in the holdout sample are satisfactory at the 5% level. So, the results from the logit and probit models were satisfactory and they were inferior to DA models for predicting non-failed firms.

4.3_ ANN models

This section analyses the predictive ability of ANNs[8] for Brazilian firm failures. The best specification consisted of four variables in the input layer. The input variables used were returns on total assets, net asset turnover, solvency and gearing. The number of neurons in the hidden layer was selected by experimentation. At the first stage, we tried to use one up to six neurons in a single hidden layer, then two and three hidden layers. The results indicated that two neurons was best choice. Also, the results from the networks with more than one hidden layer were over-fitted. So, the best ANN only used one hidden layer. The final topology chosen for each of the hybrid network are given in Table 2 of the Appendix B.

[8] Estimated using the software package Neuroshell 2.

For ANN models, the firms were classified as failed if the probability of failure exceeds 0.5. The success rate of predicting failure was 79.3% and that of success, 98.0%. For the holdout sample, the success rate of predicting failure was 67.0% and that of success, 95.0%. The test for the statistical significance of the hit rate shows that the results in the holdout sample are satisfactory at the 5% level. So, the results from the ANNs are inferior to DA, for predicting non-failed firms, and as good as the logit and probit models for predicting all firms.

4.4_ Hybrid ANN models

This section analyses the predictive ability of hybrid ANNs for Brazilian firm failures. When we compare the results from the DA, logit and ANN models, we see that each technique misclassified different firms. So, each approach must be using the information in different ways. For that reason, the combination of statistical models and ANN may produce better results. The ANN models were combined with the best statistical models estimated in previous section to produce hybrid neural networks. Firstly, we used statistical models to pre-select the variables used in the ANNs. Secondly, the outputs from the statistical models were used as inputs into a network. Finally, we combined the preselected variables from the statistical models and the output from the statistical models. The final topology chosen for each of the hybrid network are given in Table 2 of the Appendix B.

The best probit models used the same variables as the logit models. Also, both models misclassified the same banks and the estimated probabilities of failure were almost the same. So, it was decided to consider only the probability of failure from the logit models to build the hybrid networks. The institutions were classified as failed if the probability exceeds a cutoff point of 0.5. The classification for the best hybrid networks is shown in Table 3 the best hybrid model is ANN-Plogit networks. The success rate of predicting failure was 93.0% and that of non-failed, 100.0%. Table 3 shows that in the holdout sample the best hybrid model correctly classified 89.0% of the failed firms and 100% of the non-failed firms The test for the statistical significance of the hit rate shows that the results in the holdout sample are satisfactory at level of 5%. So, the results from hybrid ANN model were superior to all other models in the holdout sample.

Table 3_ Classification accuracy for the best hybrid network models

(in %)

Model	In-sample		Holdout sample	
	Failed firms correctly classified	Non-failed firm correctly classified	Failed firms correctly classified	Non-failed firm correctly classified
ANN-DA	79.3	97.0	67.0	95.0
ANN-Logit	79.3	98.0	67.0	95.0
ANN-Logit-plogit	83.0	97.0	78.0	100.0
ANN-Logit-PDA	79.3	98.0	67.0	95.0
ANN-DA-PDA	83.0	98.0	67.0	100.0
ANN-DA-Plogit	83.0	97.0	67.0	100.0
ANN-PDA	90.0	100.0	89.0	100.0
ANN-Plogit	93.0	100.0	89.0	100.0

4.5_ Comparison of the models

From the statistical models, the DA model correctly classified more non-failed than the logit models in sample, but correctly classified less non-failed firms in the holdout sample. The results from ANN were the same as the logit or probit models in the holdout sample. The performance of the ANN was improved when the hybridization with DA and logit models was considered. According to Table 4, the best models were ANN-Plogit. This model correctly classified 93.0% of the failed firms and 100% of the non-failed firms. For the holdout sample, 89.0% of failed firms and 100% of non-failed firms were accurately predicted.

The test for the statistical significance of the hit rate strongly rejects the null hypothesis that the model predicts no better than a non-informative model at level of significance of 5% for all the hybrid models.

The variables selected by each model were not identical, however every model used the solvency and gearing ratios. Only Logit model used returns on total assets. Only DA and ordinary ANN models used returns on capital employed and net asset turnover. As all these variables are highly correlated, the final selection of variables will depend on the functional form of the model.

Table 4_ Classification accuracy for the best hybrid network models

(in %)

Model	In-sample		Holdout sample	
	Failed firms correctly classified	Non-failed firm correctly classified	Failed firms correctly classified	Non-failed firm correctly classified
DA	79.3	91.3	67.0	100.0
Logit	72.4	98.0	67.0	95.0
ANN	79.3	98.0	67.0	95.0
ANN-Plogit	93.0	100.0	89.0	100.0

5_ Conclusion

This study investigated whether two artificial neural networks, multilayer perceptron and hybrid networks, can outperform traditional statistical models for predicting Brazilian firm failures one year prior to the financial distress. The statistical models and the ordinary ANN worked well for predicting Brazilian firm failures. The most important finding is based on the fact that the best hybrid ANNs were able to provide better results than all other models one year before failure. According to the results, return on capital employed, return on total assets, net asset turnover, solvency and gearing are the most important determinants of Brazilian firm failures.

The gearing ratio is defined as the debt per equity. This ratio shows how much of the company's financial structure is debt and how much is equity. A high ratio indicates greater leverage. Although financial leverage is desirable up to a certain point, too much debt is dangerous, because it limits the company's ability to finance additional growth. The higher the level of debt, the higher the level of financial risk due to the increased volatility of profits. The profitability ratios selected were returns on capital employed and returns on total assets. The first ratio tells how much the firm is earning on shareholder investment. It is a measure of overall efficiency, and a reflection on financial as well as operational management. The second ratio measures the efficient utilization of the company's assets in generating profits. As we would expect, low profitability ratio is associated with high probability of failure. The solvency ratio is the total of shareholders' funds per total assets. Failed firms had a low solvency ratio because it implies that these firms are predominantly financed with debt. The lower the level of solvency, the lower the chances of the firm to meet its obligations. The asset management ratio is the net asset turnover. This measures the company's effectiveness in using its total assets and is calculated by dividing total assets into sales. This ratio tells us how many dollars of sales have been generated for every one dollar of asset employed. As we would expect, low activity ratio is associated with high probability of failure.

In summary, the Brazilian firm failures are best explained by how much and how long they borrow, how well they manage their funds and how fast they can pay their obligations. Therefore, the challenge of policymakers is to use appropriate prudent supervision in order to regulate their industry. This supports the conclusion that for researchers, policymakers and others interested in early warning systems, hybrid networks would be useful.

References

ALTMAN, E. Discriminant analysis and prediction of corporate bankruptcy. *Journal of Finance*, p. 589-609, 1968.

ALTMAN, E. I.; BAIDYA, T.; RIBEIRO-DIAS, L. M. Assessing potential financial problems of firms in Brazil. *Journal of International Business Studies*, Fall, 1979.

ALTMAN, E.; MARCO,. G.; VARETTO, F. Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking and Finance*, n. 18, p. 505-529, 1994.

ANDRADE, L.; ANDRADE, M.; CARVALHO, A. *Architecture design of artificial neural networks based on Box & Jenkins models for times series prediction*. IEEE/ International Conference on Computational Intelligence and Multimedia Applications, v. 1, p. 29-35, 1999.

BACK, B.; SERE, K.; WEZEL, M. C. Choosing the best set of bankruptcy predictors. Technical Report. *Turku Centre for Computer Science*, Finland, 1996.

BEBIS, G.; GEORGIOPOULOS, M.; KASPARIS, T. Coupling Weight elimination with genetic algorithm to reduce network size and preserve generalization. *Neurocomputing*, v. 17, p. 167-194, 1997.

BLUM, M. P. Failing company discriminant analysis. *Journal of Accounting Research*, v. 12, 1974.

BOX, G. E. P.; JENKINS, G. M. *Time series analysis:* forecasting and control. 2th edition. Holden Day: San Francisco-California, 1976.

CHANDLER, J. S.; HAN, I.; LIANG, T. Integrating Statistical and Inductive Learning Methods for Knowledge Acquisition. *Expert Systems with Applications*, v. 1, p. 391-401, 1990.

CLEMEN, R. T. Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting,* v. 5, p. 559-583, 1989.

COATS, P. K.; FANT, L. F. Recognizing financial distress patterns using neural network tool. *Financial Management* , v. 22, p. 142-155, 1993.

CUNHA, A. G. *Algoritmo de poda em redes neurais:* um estudo de hipertensão arterial. 2000. Dissertação (Mestrado) – Departamento de Engenharia e Produção, Universidade Federal Fluminense, 2000.

DEAKIN, E. B. A discriminant analysis of predictors business failure. *Journal of Accounting Research,* p. 167-179, Spring, 1972.

DILLON, W.; GOLDSTEIN, M. *Multivariate analysis:* methods and applications. New York: Wiley, 1984.

DROSSU, R.; MARTIN, R. D. Rapid design of neural networks for time series predictions. *IEEE Computational Sciences and Engineering,* v. 3, n. 2, p. 78-89, 1994.

ELIZABETSKY, J. R. *Um modelo matemático para a decisão no banco comercial.* 1976. Trabalho apresentado ao Departamento de Engenharia e Produção da Escola Politécnica da USP.

FRANSES, P. H. A test for the hit rate in binary response models. *International Journal of Market Research,* v. 42, n. 2, p. 239-245, 2000.

GENTRY, J. A.; NEWBOLD, P.; WHITFORD, D. T. Classifying bankruptcy firms with funds flow components. *Journal of Accounting Research,* p. 146-160, 1985.

HAN, I.; KWON, Y.; LEE, K. C. Hybrid neural network models for bankruptcy predictions. *Decision Support Systems*, v. 18, p. 63-72, 1996.

HARVEY, A . C. *Times Series Models.* Second Edition. MIT Press, 1993.

HOSMER, D. W.; LEMESHOW, S. *Applied Logistic Regression.* John Wiley & Sons, 1989.

JINN, T.; NAM, J. H. Bankruptcy prediction: evidence from Korean listed companies during IMF Crisis. *Journal of International Financial Management and Accounting*, v. 11, n. 3, p. 179-197, 2000.

LACHTERMACHER, G.; FULLER, J. D. Backpropagation in Time-series forecasting. *Journal of Forecasting*, v. 14, p. 381-393, 1995.

MADALLA, G. S. *Limited-dependent and qualitative variables in Econometrics.* Cambridge University Press, UK, 1983.

MARKHAM, I.; RAGSDALE, C. Combining neural networks and statistical prediction to solve the classification problem in discriminant analysis. *Decision Sciences*, v. 26, n. 2, p. 229-241, 1995.

MATIAS, A. B. *Contribuição as técnicas de análise financeira:* um modelo de concessão de crédito. 1978.Trabalho apresentado ao Departamento de Administração da Faculdade de Economia e Administração da USP.

MILLER, G.; TODD, P.; HEDGE, S. *Designing neural networks using genetic algorithms.* Third International Conference of Genetic Algorithm and their Applications, p. 379-384, 1989.

MINARDI, A. F.; SANVICENT, A. Z. *Identificação de Indicadores contábeis significativos para previsão de concordata de empresas.* IBMEC, 1998. (Financelab Working Paper, n. 3)

NORUSIS, M. J. *SPSS advanced statistics user's guide.* SPSS, 1990.

ODOM, M. D.; SHARD, R. *A Neural Network Model for Bankruptcy Prediction.* International Joint Conference on Neural Networks, San Diego, CA, v. 2, p. 163-167, 1990.

OHLSON, J. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, v. 18, n. 109-131, 1980.

PORTUGAL, M. Neural Networks versus Time Series Methods: a Forecasting Exercise. *Revista Brasileira de Economia*, v. 49, n. 4, p. 611-629, 1995.

RUMELHART, D.; MCCLELLAND, J.; PDP Group 1986. *Parallel distributed processing.* Exploration in the Microstructure of Cognition. v.1: Foundation. Cambridge, Mass.: MIT Press, 1986.

SILVA, J. P. *Administração de Crédito e Previsão de Insolvência.* São Paulo: Editora Atlas, 1983.

SIQUEIRA, J. O.; MATIAS, A. B. Risco Bancário: modelo de previsão de insolvência de bancos no Brasil. *Revista de Administração*, v. 31, p. 19-28, 1996.

TANG, Z.; ALMEIDA, C.; FISHWICK, P. A. Time Series forecasting using neural networks versus Box-Jenkins methodology. *Simulations,* p. 303-310, 1991.

WARNER, J. B. Bankruptcy costs: some evidence. *The Journal of Finance*, v. 32, p. 337-347, 1977.

WEIGEND, A. S. E., NEUNEIER, H. G. R. *Clearing.* Technical Report. University of Colorado, 1995.

WEIGEND, A. S.; RUMELHART, D. E.; HUBERMAN, B. A. *Generalization by weight-elimination applied to currency exchange rate prediction.* IEEE/International Joint Conference of Neural Networks, v. 1, p. 837-841, 1991.

YAO, X. *Evolutionary System for Evolving artificial Neural Networks.* IEEE Trans. Neural Networks, p. 694-713, 1997.

Author's e-mail
• juliana.yim@ems.rmit.edu.au

Appendix A
Backpropagation algorithm

The first phase of the backpropagation algorithm consists of repeatedly presenting the network with examples of input and expected output. Suppose that the q^{th} neuron of the hidden layer receives the activation signal, H_q, given by:

$$H_q = \sum_j v_{qj} x_j$$

where x_j is the signal to the input neuron j and v_{qj} is the weight of the connection between input neuron j and the hidden neuron q.

This activation signal is then transformed by a transfer function f in the hidden layer to give the output

$$h_q = f(H_q)$$

The output neuron i now receives the activation signal, O_i, from the hidden nodes given by

$$O_i = \sum_q w_{iq} h_q$$

where w_{iq} is the weight of the connection between hidden neuron q and output neuron i. This is transformed again to give the output signal

$$o_i = f(O_i)$$

This is then compared with the desired, or actual value of the output neuron, and the function of squared errors for each node, which is to be minimized, is given by

$$E(w) = \frac{1}{2} \sum_i (d_i - o_i)^2$$

In the second phase the weights are modified to reduce the squared error. The change in weights, Δw_{iq}, used by the backpropagation is given by

$$\Delta w_{iq} = -\gamma \frac{\partial E(w)}{\partial w_{iq}}$$

where $0 < \gamma < 1$ is the learning rate. Using the chain rule, it can easily be shown that

$$\Delta w_{iq} = -\gamma \frac{\partial E}{\partial o_i} \frac{\partial o_i}{\partial O_i} \frac{\partial O_i}{\partial w_{iq}} = \gamma (d_i - o_i) f'(O_i) h_q = \gamma \delta_{oi} h_q$$

where δ_{oi} is the error signal of neuron i and is given by

$$\delta_{oi} = (d_i - o_i) f'(O_i)$$

To avoid oscillation at large γ, the change in the weight is made dependent on the past weight change by adding a momentum term

$$\Delta w_{iq}(t+1) = \gamma \delta_{oi} h_q + \alpha \Delta w_{iq}(t)$$

where α is a constant chosen by the operator. Similarly it can be shown that the change in the weight between the hidden neuron i and the input neuron j, Δv_{ij}, is given by

$$\Delta v_{qj} = \gamma \delta_{hq} x_j$$

where δ_{hq} is the error signal of neuron q and is given by

$$\delta_{hq} = f'(H_q) \sum_i \delta_{oi} w_{iq}$$

As before a momentum term can be used to prevent oscillation.

Appendix B

Table 1_ List of financial ratios

Category	Financial Ratio	Definition	Code
Profitability	Profit Margin (%)	P*/L** before tax/operating revenue (turnover) × 100	PRM
	Return on Shareholders Funds (%)	P/L before tax/Shareholder's Funds × 100	RSF
	Return on Total Assets (%)	P/L before tax/total assets × 100	RTA
	Return on Capital Employed (%)	(P/L before tax-int expense)/(Shareholder's Funds + non-current liabilities)	RCE
	Cost of Empl./Op. Revenue (%)	(-Cost of employees/operating revenue) × 100	ORE
	Operat. Rev. per Employee th USD	(Operating revenue/number of employee)	ARE
	Aver. Cost of Employee/Year th USD	(-Cost of employees/number of employee)	ACE
	Profit per Employee th USD	P/L before tax/number of employee	PPE
	Cash Flow/Oper. Revenue (%)	(Cash flow/operating revenue) × 100	CFO
Liquidity	Current ratio	Current assets/current liabilities	CUR
	Liquidity ratio	(Current assets-stock)/current liabilities	LIQ
	Interest Cover	Operating P/L/interest expenses	INT
	Collection Period	(debtors/operating revenue) × 360	COP
	Credit Period	(creditors/operating revenue) × 360	CRP
Asset Utilization	Share. Funds per Employee th USD	Share. Funds/number of employees	SFE
	Working Cap. per Employee th USD	Working Cap. /number of employees	WCE
	Total Assets per Employee th USD	Total Assets/number of employees	TAE
	Stock Turnover	Operating revenue/stocks	STT
	Net Assets Turnover	(Operating revenue)/(Shareholder's Funds + non-current liabilities)	NAT
Structure	Shareholders Liquidity ratio	Shareholder's Funds/non-current liabilities	SHL
	Solvency ratio (%)	(Shareholder's Funds/total assets) × 100	SOL
	Gearing (%)	(non-current liabilities + loans)/Shareholder's Funds	GEA

Notes: (*) P: Profit
(**) L: Loss

Table 2_ Topology from the best ANN models

Topology	Best network	Learning rate	Momentum
ANN	$4 \times 3 \times 1$	0.5	0.5
ANN-DA	$3 \times 2 \times 1$	0.5	0.7
ANN-Logit	$3 \times 2 \times 1$	0.5	0.7
ANN-Logit-plogit	$4 \times 3 \times 1$	0.7	0.9
ANN-Logit-PDA	$4 \times 2 \times 1$	0.5	0.7
ANN-DA-PDA	$4 \times 2 \times 1$	0.5	0.7
ANN-DA-Plogit	$4 \times 3 \times 1$	0.7	0.9
ANN-PDA	$5 \times 4 \times 1$	0.5	0.5
ANN-Plogit	$5 \times 4 \times 1$	0.5	0.7