**Association for Information Systems**

**AIS Electronic Library (AISeL)**

ECIS 2000 Proceedings

European Conference on Information Systems (ECIS)

2000

# Prediction of Insolvency of Life Insurance through Neural Networks

Ernest P. Gross

*Creighton University*

George S. Vozikis

*University of Tulsa*

search

# Prediction of Insolvency of Life Insurers through Neural Networks

Ernest P. Goss
Department of Economics
Creighton University
Omaha, NE 68178

George S. Vozikis
Department of Management and Marketing
University of Tulsa
Tulsa, OK 74104

**Abstract**-***Past research studies have documented the failure of the Insurance Regulatory Information System (IRIS) to provide adequate warning of insurer financial distress or insolvency. As a result, scholars have examined alternative parametric and non-parametric models to predict insurer insolvency. This study uses a neural network, a non-parametric alternative to past techniques, and shows how this methodology more effectively predicts insurer insolvency than parametric models***.

## I. Introduction

Researchers often encounter binary dependent variables in the prediction of individual or company credit worthiness. Based upon research of Berkson [10] and Altman [2], analysts have routinely used binary logit regression (BLR) and discriminant analysis (DA) to solve this class of problem. Empirical studies examining these two parametric methodologies have generally concluded that both are effective in predicting business failure or bankruptcy [28] and [7].

However, parametric models, such as BLR and DA, require the researcher to specify, in advance, the functional relationship between dependent and independent variables, and to identify expected interaction effects among independent variables. These limitations can be especially problematic for solvency assessment in the insurance industry where recent high failure rates and low profit margins have intensified practical and academic interest in quantitative and statistical methods of insolvency prediction [4]. Furthermore, the multiple threats of rising health costs and AIDS have increased financial stress among life and health insurers [42]. The increased financial stress has encouraged the federal government to more closely examine insolvency assessment methodologies and to actually consider setting federal insolvency standards for companies involved in the interstate sale of life and health insurance [15].

The heightened importance of insolvency prediction in the insurance industry, and limitations of parametric prediction models, have prompted researchers to test non-parametric, or distribution free, alternatives to BLR and DA [5]. However, to the authors' knowledge, researchers have not tested the effectiveness of neural networks (NNs) in insolvency assessment among life and health insurers, despite NN's demonstrated effectiveness in applications as diverse as aquatic salinity prediction [17], weather forecasting [34], and hospital survival modeling [18] and [13].[1]

The purpose of this study is to construct an empirical NN for predicting life and health insurer insolvency and to compare its forecasting capability to that of parametric models. It is proposed that NN models could assist state regulators in establishing the size of loss reserve funds and in focusing financial audit activities of financially distressed insurers [22]. Using data from an earlier study by BarNiv and Hershbarger [6], the present study will compare insolvency predictions obtained from a NN, BLR and DA.

## II. Insolvency prediction in the insurance industry

In terms of receiving early warning of potential insurer insolvency, insurance purchasers and benefit managers have typically relied heavily on rating services such as the A.M. Best Company [42]. However, A.M. Best currently rates only slightly over 3,000 insurers. Thus, many regulated insurers are not rated by the A.M. Best Company, or other rating services such as Moody's or Standard and Poor's. For

[1] Brockett, Cooper, Golden, and Pitaktong [13] tested a neural network on insolvency prediction in the property-casualty insurance industry in Texas and noted the need for testing a NN on a national sample of life insurers.

example, of the 1,300 life and health insurers examined in 1994, Best placed 294 in the vulnerable category but failed to rate 419 of the insurers [12]. Additionally, hundreds of smaller insurance companies did not meet the requirements to be included in A.M. Best's reports and were likely more susceptible to financial collapse than examined firms.

To aid state regulators in insurer oversight activities and solvency assessment, the National Association of Insurance Commissioners (NAIC) developed the Insurance Regulatory Information System (IRIS) in the 1970s. The IRIS was intended to quickly identify companies that require close surveillance from state insurance department personnel. The IRIS warns regulators that immediate attention is required when four of twelve designated financial ratios of an insurer are outside the specified range [33].

In theory, the IRIS highlights insurers who have a high probability of insolvency and merit closer financial scrutiny from regulatory agencies. However, Thornton and Meador [41] and Hershbarger and Miller [25] concluded that the IRIS tests were ineffective in predicting insurer insolvency and that several of the IRIS ratios provided very little explanatory power in predicting insurer insolvency. One shortcoming of the IRIS approach is that acceptable ranges for the twelve ratios are so wide that few truly vulnerable insurers meet the IRIS conditions for regulator audit activity.

Barrese [11] concluded that it was possible to improve substantially over the IRIS ratios in predicting insolvency using multivariate techniques such as regression. In past studies, researchers have normally used the regression technique of BLR and/or multivariate DA to confirm the inadequacies of the IRIS system and benchmark alternative techniques.

### *A. Parametric Modeling*

Researchers have used the Binary Logit Regression (BLR) model extensively in insurance industry studies predicting insolvency.[2] In cases with non-normally distributed populations, research has found the BLR model to provide classification results superior to that of the Discriminant Analysis (DA) model when the BLR parameters are estimated with maximum likelihood techniques [37] and [31]. One problem with the use of this technique is that in cases where the likelihood function is flat or has more than one peak, the function may diverge or converge to a local maximum depending on the initial or "starting" parameter estimates. As a consequence, different empirical results can be obtained for different starting values of parameter estimates. Typically, statistical packages use ordinary least squares (OLS) estimates as starting values. Thus, if the OLS starting values are far from the global maximum, this methodology can result in convergence to a local maximum rather than the global maximum.

The DA procedure assigns an observation of unknown group membership to one of two or more groups based upon the observation's calculated discriminant score. Fisher's methodology, used in this research study, employs a linear combination of the independent variables, which yields estimated coefficients that maximize the ratio of variance between groups to variance within groups [20].

In the case of insurer insolvency, coefficients of the discriminant function are usually estimated from financial ratios of companies in the training set. The estimated parameters from the DA model are then used to produce a discriminant score for each observation in test or "holdout" data. The DA algorithm then assigns each insurer to the predicted groups based on this discriminant score [36].

Theoretically, DA requires the two populations to be characterized as multivariate normal distributions with equal covariances. Research has concluded that violations of the basic DA assumptions produce parameter estimates and categorization capabilities less than optimal [26]. In the case of bankruptcy prediction, Deakin [16] found fourteen of fifteen financial ratios examined failed to conform to the normality assumption. In order to overcome violations of basic DA assumptions, Lachenbruch, Sneeringer, and Revo [26] recommended the use of quadratic methods. However, Altman, Avery, Eisenbeis, and Sinkey [1] demonstrated that quadratic methods produce poor classification results on test or holdout data. As a result of the limitations of BLR and DA, it is hypothesized that the NN model could potentially provide regulators with a model that provides more accurate predictions when violations of the basic assumptions of the BLR and DA models exist.

### *B. Non-Parametric Methodologies*

Remarkable enhancements in the computational power of digital computers have expanded the researchers' use of non-parametric, or distribution free, techniques which allow the data to dictate functional relationships among variables in the model [32], [8] and [29]. For example, BarNiv and Raveh [8] outlined a non-parametric categorization technique that predicted bankruptcy more accurately than DA in a sample of 200 industrial firms.

Contrary to DA, the BarNiv-Raveh model permits non-normally distributed independent variables. Their procedure maximizes an index of separation that is a monotonic transformation of DA scores derived from Fisher's

[2] See BarNiv and McDonald [5] for a review of empirical studies.

discriminant function. The technique guarantees maximum separation between groups by adjusting the estimated discriminant coefficients so that the number of misclassifications in the training set is minimized. However the method, which is limited to producing ordinal rankings within groups, categorizes observations with no quantification of the prediction error.

Likewise, Messier and Hansen [32] developed a non-parametric technique that categorized bankrupt firms more correctly than parametric techniques such as DA. After scanning training data, the methodology generates a classification tree, or set of rules, which identifies the most effective method of classifying the data into risk groups. These rules, or algorithms, are then used to classify test or "unseen" data. This approach, contrary to other rule-based systems, does not rely on explicitly obtaining the expert's knowledge or decision matrix. However, Messier and Hansen recommend that practitioners limit the use of this technique to relatively small data sets since the entire set of examples must be searched and a new set of rules generated each time a new example is added.

Other analysts, using non-parametric categorization techniques, merged rule-based or expert systems with NNs. For example, CREDITVIEW, a hybrid NN and expert system, evaluates the credit worthiness of loan applicants based on past financial data on good and bad obligators, and on industry norms drawn from COMPUSTAT data [29]. Developed by Chase Manhattan Bank, this model produces three-year predictions that are used to assign companies to risk categories of good, problematic or charged-off. Paralleling non-parametric approaches such as Chase's CREDITVIEW, the present study applies a NN to historical financial data to classify life insurance companies into risk categories. This methodology, as opposed to expert systems that rely on the explicit delineation of rules by the expert, depends on "learning" within the system.

### C. The NN Approach

The basis of the neural network approach is the construction of a network of nodes that are interconnected, with weights associated with each connection. Learning algorithms determine the weights for each connection. In order to empirically estimate the weights, a set of representative training cases are first presented to the network. The network "learns" the relationship between the predictor variables and the outcome variable for the training set. When presented with new or test data, the network predicts the outcome variable based upon patterns learned from the training set.

The network is typically composed of an input layer, a hidden layer and an output layer, each of which contains one or more nodes.[3] In standard statistical terminology, nodes in the input layer correspond to the independent variables while nodes in the output layer compare to the dependent variable. The hidden layer simulates interaction effects among nodes in the input layer.

Fig. 1 presents a schematic of a NN with three nodes in the input and hidden layers and with one node in the output layer. Weights to and from each layer, indicated by the solid connecting lines, are empirically estimated to produce an optimal model. The NN package used in this study adds two bias nodes to the basic model. The bias nodes are always "on" supplying a weight similar to a threshold value that is learned just like other weights in the model [39]. The value at the output node is thus a function of the weights of the paths from the hidden nodes and the bias node.

Although published research studies identify a wide variety of NN approaches, researchers increasingly develop applications using back propagation learning algorithms due to their intuitive appeal, high level of development, and ease of use. Kolmogorov [24] through an existence theorem, proved the ability of multi-layer networks to represent non-linear patterns. Applying the back propagation learning algorithm to input data, White [45] concluded that a three layer NN, as presented in Fig. 1, provides an accurate approximation to any function likely encountered as long as the number of hidden units is sufficiently large.

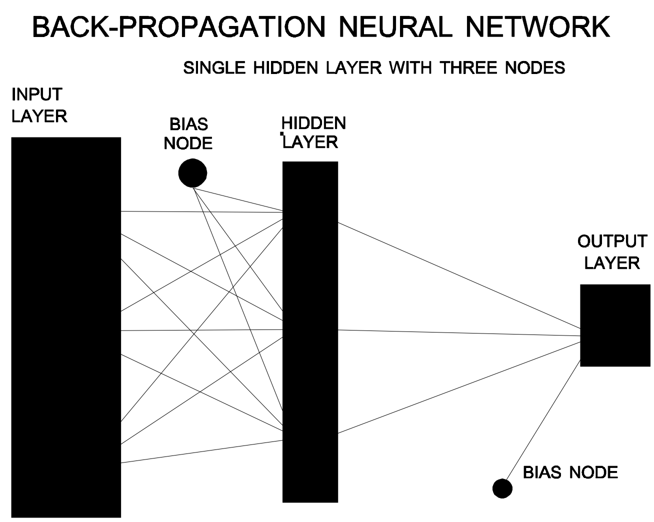


Fig. 1. BPNN with input, hidden, bias, and output nodes

[3] NNs are usually hierarchical and can have multiple hidden layers as well as multiple output nodes. The discussion is limited to NNs with one hidden layer and one output node. Caudill [14] and Lippman [27] recommend the use of single hidden layer NNs unless there are compelling reasons for a four layer NN.

White [45] contended that back propagation and non-linear regression could be viewed as alternative statistical approaches to solving the least squares problem. Rumelhart, Hinton, and Williams [39] provide a more detailed specification of the back propagation learning algorithm. The next section presents the empirical results from applying BLR, DA, and NN models to the sample data.

## III. Empirical results

Data used in the subsequent empirical analysis comes from the BarNiv and Hershbarger study [6]. This is one of a very few studies examining insolvency in the life insurance industry. BarNiv and Hershbarger's selection criteria produced a total of twenty-eight insolvent life insurance companies matched with twenty-eight solvent insurers based on state of domicile, firm size and time frame. They defined insolvent insurers as those companies declared insolvent by their respective state insurance commissioners and subsequently also reported insolvent by the A.M. Best Company. BarNiv & Hershbarger's definition, selection criteria and data are also used in the present study. However, BarNiv and Hershbarger included eight life insurance firms for which data were not provided by *A.M. Best's Insurance Reports*. Since being listed in the *A.M. Best's Insurance Reports* is itself an indicator of insurer financial solvency, these eight firms were omitted from the analysis.[4] Applying this additional condition to the BarNiv and Hershbarger dataset reduced this study's dataset to twenty insolvent and twenty matching solvent life insurance firms.

Though the dataset is somewhat smaller than that used by BarNiv and Hershbarger, past research has generally demonstrated that small samples generally favor parametric models over NN in classification comparisons. For example, Soulie, Gallinari, Cun, and Thiria [39] found that as the size of the training set increases, the ability of the neural network to accurately predict the test set increased. Generalizing to other studies, they found that for sample sizes less than 100, the relative accuracy of the NN deteriorated dramatically. On the other hand, Wann, Hediger, and Greenbaum [43] found that larger training sets do not guarantee better NN performance. To the authors' knowledge, no study to date has demonstrated that sample size adversely affects prediction comparisons between NNs and parametric models.

Despite the paucity of variables, the authors, for several reasons, rejected adding other explanatory variables. First, Frydman, Altman, and Kaod [21] concluded that the use of fewer financial ratios and equal group sizes produces less biased results in failure prediction. Second, adding other financial ratios to the model, in general, provides negligible explanatory power since financial ratios of life insurers are usually highly correlated.

BarNiv and Hershbarger [6] used a four variable model to estimate insolvency in the life insurance industry and found that the selected IRIS ratios had significant discriminatory power.[5] In the subsequent analysis, this same four variable model is used. A description of each variable, along with group means, is presented in Table 1.

## IV. Classification methodology

In order to compare the relative effectiveness of the empirical models, this study uses the jackknife method, as outlined by Lachenbruch [26] and applied to the insurance industry by Ambrose and Seward [3]. This method produces sample distribution data from which one obtains confidence intervals for the probability of misclassification. This validation technique supplies the researcher with a valuable tool for comparing the effectiveness of alternative models when the number of observations is relatively small.

The jackknife technique has an advantage over re-substituting the sample values back into the optimal model, since re-substitution underestimates the probabilities of misclassification. Another testing technique, dividing the sample into training and testing sets, requires fairly large samples and has the disadvantage that the empirical results could be biased by non-random selection of the training and test sets [26].

Using the jackknife method, the researcher first develops an optimal model from a training set obtained by omitting one observation from the full data set. Using the empirical model generated from the training data, the researcher classifies the omitted observation. Next the analyst adds the previously omitted observation to the training set, omits a new observation and re-estimates the model. The researcher uses the new optimal model to classify the omitted case. The researcher repeats this process until all observations in the training set have served as test data and have been classified. The analyst then computes the overall percent of cases misclassified. A misclassification is determined by comparing the actual group membership (solvent or insolvent) to the predicted group membership for each life insurer in the test set. For BLR and NN models, the predicted group membership is determined by comparing the estimated

[4] Six of these eight insurers subsequently failed.

[5] Using the jackknife method, Ambrose and Seward [3] tested a four independent variable model on a sample of 58 property/casualty insurers and correctly classified almost 80% of the test set correctly.

probability of insolvency ($O_k$) to the cutoff value (*CV*). Thus if $O_k \geq CV$, the test case is classified as insolvent, and if $O_k < CV$, the test case is classified as solvent. In cases with matched data whereby half of the full data set fails, one normally uses 50% as the *a priori* probability of failure or *CV*. BLR produces a predicted probability of insolvency and the NN produces an output score that can be interpreted as the probability of insolvency.

Hecht-Nielsen [24] used Kolmogorov's theorem to show that any real mapping can be exactly performed by a three layered neural network of limited size and that (2*i*+1) hidden nodes are required for a single hidden layer network, where *(i)* is the number of input nodes. However, Pao [35] and Maren, Jones, and Franklin [30] recommended a smaller number of hidden nodes in order to more fully exploit regularities of the sample.

Table 1. Mean value of predictor variables

| Variable | Insolvent Firms | Solvent Firms | All Firms |
|---|---|---|---|
| I2-IRIS ratio of net gain to income | 0.066 | 0.095 | 0.081 |
| I3-IRIS ratio of commission and other expenses to premium income | 9.086 | 0.472 | 4.780 |
| I4-IRIS ratio of net invested income to net invested assets | 0.081 | 0.094 | 0.088 |
| I11-IRIS ratio of change in income invested in assets[6] | 0.096 | 0.073 | 0.084 |
| At the 95% level of confidence, differences between ratios of solvent and insolvent insurance companies were not statistically different. | | | |

Yoon, Swales, and Margavio [47] using a NN to predict stock prices, found improved performance for a three layered network when the number of hidden units is increased up to a certain limit. More recent research has demonstrated that increasing the number of nodes in the hidden layer usually improves performance on training data but diminishes classification performance on test data as the ability of the NN to generalize is reduced [27].

However, Rumelhart, Hinton, and Williams [39] contended that there is no objective methodology to generate the optimal NN configuration and that determination of the proper number of nodes in the hidden layer is more art than science. In order to gauge the impact of hidden layer size on classification effectiveness in the present investigation, the number of nodes in the hidden layer is varied from (0*i*+1) to (3*i*+1), or from one hidden node ($NN_1$) to thirteen hidden nodes ($NN_{13}$).

During the NN training process for the thirteen models, the learning rate ($\alpha$) was gradually reduced from 0.9 to 0.01 to hasten learning.[7] This methodology is supported by White [45] who suggested that the learning rate should decline to a vanishing point in order to generate optimum results and to avoid "getting stuck" in a local optimum. The learning rate determines the effect of past weight changes on the current direction of movement in weight space. Lower learning rates tend to avoid oscillation but produce longer training times [43].

Initial experimentation revealed no significant reduction in error below an average of 3% per observation in the training set, even when the neural network was allowed to "learn" for eight hours on an Intel 80486, 33 MHz microcomputer. Subsequently, the learning process was stopped when the average error in the training set descended to 3% per observation. When the error was greater than 4%, classification effectiveness of the test set was found to be impaired.

Misclassifying insolvent insurers as solvent is usually assumed to cost much more than misclassifying solvent insurers as insolvent [3]. However, if an otherwise solvent firm is predicted to fail, this could have the impact of damaging the financial condition of the firm and encouraging insolvency. Thus, dropping the *CV* to increase the likelihood that insolvent insurance firms are highlighted could be very costly for firms that would otherwise survive financial problems. Moreover, if the *CV* is lowered in order to reduce the proportion of misclassified insolvent insurers, $NN_4$ continues to outperform BLR and DA in classification

[6] NAIC more broadly defines this IRIS ratio as the change in overall assets.

[7] In addition to employing alternative hidden layer size, the authors experimented with alternative learning rates.

accuracy.

Fig. 2 presents the classification results for each of the thirteen configurations using the methodology outlined above and using the four ratios listed in Table 1 as input nodes.

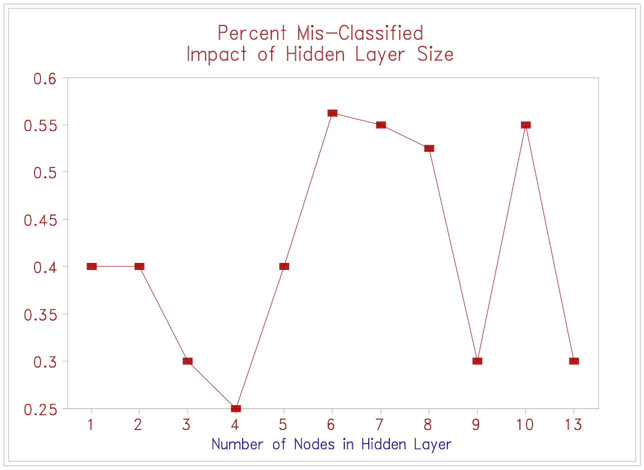


Fig. 2. Impact of hidden layer size on percent Missclassified

As shown in Fig. 3, the $NN_4$ provides superior results for cutoff values less than 60%. Above a ***CV*** of 60%, the $NN_4$ produces a proportion of misclassifications between that of BLR and DA.

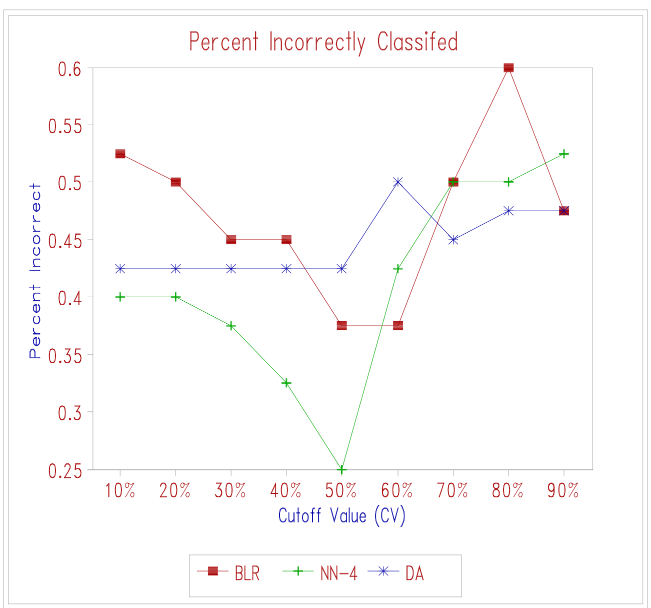


Fig. 3. Comparison of Incorrectly Classified Life Insurers for Alternative Models

Compared to other non-parametric techniques tested previously, the $NN_4$ generates promising classification results on life and health insurance companies. For example, the $NN_4$ developed in this study compares favorably to BarNiv and McDonald's [5] exponential generalized beta distribution of the second kind (EGB2) in classification success. Despite EGB's lower expected cost of misclassification in comparison to DA, it did not generate classification results that were statistically different from DA.

On the other hand, the classification success of each model tested in this study was somewhat less than that obtained by BarNiv and Hershbarger [6], because by excluding firms not listed in the *A.M. Best's Insurance Reports*, a lower success rate was already expected. It was hypothesized however, that these unlisted insurers were much smaller, more likely to fail, and more accurately classified than listed life and health insurers.

## V. Summary and Conclusions

The NAIC created the IRIS to quickly identify companies that require close surveillance from state insurance department personnel. Past studies have documented the failure of IRIS ratios to predict insolvency in the insurance industry and have likewise criticized the system's arbitrary range of predictor variables. This study examined a neural network for predicting insolvency and showed how it can effectively assist regulators in predicting insolvency.

The NN tested provided a lower percentage of life and health insurance firms incorrectly categorized than either BLR or DA and a smaller prediction error than BLR. In addition to providing superior empirical results, the NN methodology does not require the developer to make assumptions regarding the underlying parameter distributions nor does it require the developer to specify potential interactions among independent variables. And contrary to DA, NNs produce a continuous scoring system between zero and one for comparisons of observations within the same group.

However, several limitations exist which may restrict the use of neural networks for prediction purposes. First, there is no formal theory for determining optimal network typology. Thus, decisions such as the appropriate number of layers, hidden layer size, and the appropriate learning rate must be determined experimentally. Second, the interpretation of neural network output results requires more expertise from the user than traditional statistical models.

Nonetheless, results from this study indicate promising prospects of the NN model for insolvency prediction in the life and health insurance industry, and demonstrate an alternative to BLR, DA and the IRIS methodology. Results show that state insurance regulators could potentially use a NN to more adequately warn of likely financial difficulties among insurers.

## REFERENCES


[1] Altman, E. L., R. Avery, R. Eisenbeis, and J. Sinkey. *Applications of Classification Techniques in Business Banking and Finance*. JAI Press, Greenwich, CT, 1981.

[2] Altman, E. L. "Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy." *Journal of Finance*, (September 1968): pp. 589-609.

[3] Ambrose, J.M. & A. J. Seward. "Best's Rating, Financial Ratios and Prior Probabilities in Insolvency Prediction." *The Journal of Risk and Insurance*, 55, (July 1988): pp. 229-244.

[4] American Council of Life Insurance. *1992 Life Insurance Fact Book*, Washington, D.C., 1992.

[5] BarNiv, R., and J. B. McDonald. "Identifying Financial Distress in the Insurance Industry: A Synthesis of Methodological and Empirical Issues." *The Journal of Risk and Insurance*, LIX (1992): pp. 543-574.

[6] BarNiv, R., and R. A. Hershbarger. "Classifying Financial Distress in the Life Insurance Industry." *The Journal of Risk and Insurance*, 57 (Spring 1990): pp. 110-136.

[7] BarNiv, R. "Accounting Procedures, Market Data, Cash-Flow Figures, and Insolvency Classification: The Case of the Insurance Industry." *The Accounting Review*, 65 no. 3(1990): pp. 578-605.

[8] BarNiv, R. and A. Raveh. "Identifying Financial Distress: A New Non-Parametric Approach." *Journal of Business Finance and Accounting*, 16 no. 3 (1989): pp. 361-383.

[9] BarNiv, R. and M. Smith. "Underwriting, Investment and Solvency." *Journal of Insurance Regulation*, (1987): pp. 409-428.

[10] Berkson, J. "Maximum Likelihood and Minimum Chi-Square Estimates of the Logistic Function." *Journal of the American Statistical Association*, 50 (1955): pp. 130-162.

[11] Barrese, James. "Assessing the Financial Condition of Insurers. *CPCU Journal*, 43 no.1 (1990), pp. 37-46.

[12] Best, A. M. Company. *Trend Report-Life and Health Edition*. Oldwick, N.J.: A.M. Beast Company, 1994.

[13] Brockett, P.L. W.W. Cooper, L.L. Golden and U. Pitaktong. "A Neural Network Method for Obtaining an Early Warning of Insurers Insolvency." *Journal of Risk and Insurance*, 61 (1994), pp. 402- 424.

[14] Caudill, M. "Using Neural Nets." *AI Expert*, (December 1989), pp. 34-61.

[15] Christine, B. "On Raters, Ratings, and New Legislation." *Risk Management*, 39 no. 8 (1992): pp. 63.

[16] Deakin, E. B. Distributions of Financial Accounting Ratios: Some Empirical Evidence, *Accounting Review*. (January 1976): pp. 90-96.

[17] DeSilets, L. B. Golden, Q. Wang, and R. Kumar. "Predicting Salinity in the Chesapeake Bay Using Back Propagation." *Computers & Operations Research*, 19 no. 3 (1992): pp. 277-285.

[18] Ebell, M. H. "Artificial Neural Networks for Predicting Failure to Survive Following In- Hospital Cardiopulmonary Resuscitation." *The Journal of Family Practice*, 36 no. 3 (1993): pp. 297-303.

[19] Efron, B. "Regression and ANOVA with Zero-One Data: Measures of Residual Variation." *Journal of the American Statistical Association*, (March 1978): pp.113-121.

[20] Fisher, G. R. "The Use of Multiple Measurement in Taxonomic Problems." *Annals of Eugenics*, 7 (1936): pp. 179-188.

[21] Frydman, H., E. Altman, and E.l. Kaod. "Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress." *Journal of Finance*, 40 no. 1 (1985): pp. 269-291.

[22] Harrington, S.E. and J.M. Nelson. "A Regression-Based Methodology for Solvency Surveillance in the Property-Liability Insurance Industry." *Journal of Risk and Insurance*, LIII (1986): pp. 119-134.

[23] Hecht-Nielsen, R. *Neurocomputing*, New York, NY: Addison-Wesley 1989.

[24] Hecht-Nielsen, R. "Kolmogorov's Mapping Neural Network Existence Theorem." *Proceedings of the First IEEE International Conference on Neural Networks*, IEEE Service Center, Piscataway, New Jersey, (1987): pp. 11-14.

[25] Hershbarger R. and R. Miller. "The NAIC Information System and the Use of Economic Indicators in Predicting Insolvencies." *The Journal of Insurance Issues and Practices*, 9, (June 1986): pp. 21-43.

[26] Lachenbruch, P.A., C. Sneeringer, and J.T. Revo. "Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Non-normality." *Communication Statistics*, 1, (1973): pp. 39-56.

[27] Lippman, R.P. "An Introduction to Computing with Neural Nets." *IEEE, ASSP*, 4, (1987): pp. 4-22.

[28] Lo, A. W. "Logit versus Discriminant Analysis: A Specification Test and Application to Corporate Bankruptcy." *Journal of Econometrics*, 31 (1986), pp.151 178.

[29] Marose, R. A. "A Financial Neural-Network Application," *AI Expert*, (May 1990): pp. 50-53.

[30] Maren, A. J., D. Jones, and S. Franklin. *Handbook of Neural Computing Applications*, San Diego, CA: Academic Press, 1990.

[31] McFadden, D. "A Comment on Discriminant Analysis versus Logit Analysis."*Annals of Economic and Social Measurement*, (1976): pp. 511-523.

[32] Messier, W. F. and Hansen. "Inducing Rules for Expert Systems Development: An Example Using Default and Bankruptcy Data." *Management Science*, 34 no. 12 (1988): pp.1403-1415.

[33] National Association of Insurance Commissioners. *NAIC-IRIS Tests Using the NAIC Insurance Regulatory Information System*, Life and Health Edition, Madison, WI: NAIC, 1986.

[34] Ono, T. "Chaos Engineering Coming in 1993." *Japan Times Weekly* International Edition, 33 no. 4 (1993): pp. 16.

[35] Pao, Y. H. *Adaptive Pattern Recognition and Neural Networks*. Reading, MA: Addison-Wesley, 1989.

[36] Pinches, G. and J. Trieschmann. "Discriminant Analysis, Classification Results and Financially Distressed P-L Insurers." *The Journal of Risk and Insurance*, 44 (1977): pp. 289-298.

[37] Press, S. J. and S. Wilson. "Choosing Between Logistic Regression and Discriminant Analysis." *Journal of American Statistical Association*, (December 1978): pp. 699-705.

[38] Radcliffe, Steve. "Where Have All the Margins Gone. "*Best's Review, Life/Health Ed.* (April 1982): pp. 10-15.

[39] Rumelhart, D. E., G.E. Hinton and R.J. Williams. "Learning Internal Representations by Error Propagation." in *Parallel Distributed Processing Exploration in the Microstructure of Cognition*, 1, 318-362, (ed.), D.E. Rumelhart and J.L. McClelland, Cambridge, MA: MIT Press, 1986.

search

[40] Soulie, F., P. Gallinari, Y. Cun, and S. Thiria. "Evaluation of Network Architectures on Test Learning Tasks." *IEEE First International Conference on Neural Networks*, II (1990): pp. 653-660.

[41] Thornton, H.H. and J.W. Meador. "Comments on the Validity of the NAIC Early Warning System for Predicting Failures among P-L Insurance Companies." *CPCU Annals*, (1977): pp. 191-211.

[42] Vogel, G. L. "An Analyst's View of Insurance Insolvency. "*Journal of the American Society of CLUs and ChFCs*, 46 no. 3 (1992): pp. 52-58.

[43] Wann, M. T. Hediger and N. Greenbaum. "The Influence of Training Sets on Generalization in Feed-Forward Neural Networks." International Joint Conference on Neural Networks, III (1990): pp. 137-142.

[44] White, H. "Neural Network Learning and Statistics." *AI Expert*, (December 1989), pp. 48-52.

[45] White, H. "Consequences and Detection of Misspecified Non-Linear Regression Models." *Journal of the American Statistical Association*, 76 (1981): pp. 419-433.

[46] Widrow, B. and M.E. Hoff. "Adaptive Switching Circuits." 96-104, New York: IRE, WESCON Convention Record, 1960.

[47] Yoon, Y., G. Swales and T. M. Margavio. "A Comparison of Discriminant Analysis versus Artificial Neural Networks." *Journal of Operational Research Society*, 44 no. 1 (1993): pp. 51-60.