Small, US Air Carrier Financial Condition: A Back-Propagation Neural Network Approach to Forecasting Bankruptcy and Financial Stress

This article discusses the continuation of our research into the forecasting of air carrier bankruptcy. Prior studies (Chow, Gritta, and Leung 1991; Davalos, Gritta, and Chow 1999) focused on the larger airlines, called majors and nationals by the US Department of Transportation (DOT). These studies utilized both Multiple Discriminant Analysis (MDA) and Neural Network (NN) models to identify financially stressed or failed airlines. The applications of these models to the smaller carriers have not achieved the same results. A preliminary study (Gritta et al. 2000) using only smaller carriers showed promise. Since that study included only nine failed carriers, a larger sample was needed to better gauge the powers of the model. This article presents a study of the larger sample of smaller carriers and the results.

by Richard D. Gritta, Sergio Davalos, Garland Chow, and Marcus Wang

In spite of the past five- to six-year record profit run made by the U.S. airline industry, some carriers remain financially unstable. With the recent additions of Pacific Western Air, Legend Air, PanAm, and several others, the number of bankrupt airlines had risen by mid-2001 to 137 since the deregulation of the airline industry in 1978. Most of these have been the smaller airlines categorized by the US Department of Transportation (USDOT) as large and medium regional air carriers.[1] Should the economy slow, or should interest rates, labor costs or fuel prices rise, more may fail.

The purpose of this article is to continue our research into the forecasting of air carrier bankruptcy. Several prior studies (Chow, Gritta, and Leung 1991; Davalos, Gritta, and Chow 1999) centered on the larger airlines, called majors and nationals by USDOT. These studies utilized both Multiple Discriminant Analysis (MDA) and Neural Network (NN) models to identify financially stressed or failed airlines. Attempts to apply these models to the smaller carriers have not achieved the same success. A preliminary study (Gritta et al. 2000) using only data for these groups of airlines showed promise. That study included only nine failed and nine solvent carriers, however. It was felt that a larger sample was needed to better gauge the powers of the model. The purpose of this article therefore is to significantly expand that study.[2]

Because of its many implications for the flying public, the anticipation of bankruptcy in this industry is important. Given the fragile nature of the airline industry, such a study should be of interest to governmental regulatory authorities, banks, lenders, and stockholders, as well as other parties.

Journal of the Transportation Research Forum. Published in *Transportation Quarterly*,
Vol. 56, No. 2, Spring 2002 (109–123). © 2002 Eno Transportation Foundation, Inc., Washington, DC.

A Review of Prior Financial Distress Research Studies

Traditional balance sheet and income statement ratios have been employed by analysts to measure financial condition. Ground breaking research by Altman (1968) combined the use of predictive statistical models with these traditional financial ratios to create a more powerful approach to assessing strength. Altman used multiple discriminant analysis (MDA), a type of applied multiple regression in which the dependent (or predicted variable) was a cardinal or coded variable, instead of a scalar variable. Known as the Altman Z Score model, it was able to achieve a success rate of 76% in forecasting corporate bankruptcy in advance of the event. Inputs into his model included four categories of ratios; those which measured liquidity (the ability of a firm to pay debts on a timely basis), leverage (the extent to which a firm used debt to finance its asset base), activity or turnover (a gauge of how efficiently a firm was using its assets), and profitability (the firm's profits as a percent of revenue, assets, or stockholders' equity).

Gritta (1982) later applied the Altman Model to the airline industry and successfully predicted the demise of Braniff and Continental in the early 1980s. Further research by Chow, Gritta, and Leung (1991) led to the development of an industry specific MDA model; that is, one designed specifically from air carrier data. Called AIRSCORE, it achieved results on a par with Altman's 76% accuracy rate, but the model showed some bias toward large carriers. In addition, to achieve the 76% success rate, AIRSCORE left some carriers unclassified or in the "zone of ignorance" as Altman (1968) referred to it. Firms in this zone were not classified by the model.[3] For that reason, a decision was made to specify a more accurate model using a newer technique, artificial intelligence. That research (Davalos, Gritta, and Chow 1999) proved fruitful and was able to accurately predict 25 out of 26 instances of carrier failure for the large carriers referred to as "majors." When the network developed from this study was applied to data on smaller carriers, classed as large and medium regional airlines, the model did not perform nearly as well. It was concluded that factors affecting the failure of larger carriers might be different from those affecting the smaller airlines. The success of a preliminary study (Gritta et al. 2000), based on a limited sample of only nine failed and nine nonfailed small airlines, confirmed this hypothesis and it is the motivation for this expanded study.

Overview of the Dynamics of Neural Networks

The conceptual basis for neural networks (NNs) comes from biological research on the neural architecture of the human brain (Caudill 1989; Rumelhart and McClelland 1986). Neural networks are composed of interconnected neurons linked together through weighted directed arcs organized into layers much like the brain. There are many different types of neural networks. They vary in the way nodes are connected and in the way in which weights at each node are updated. The selection of the type of network is a function of judgment and experimentation. Our analysis found that the back-propagation network was the most effective for our purposes. It is the most common and widely studied type of formulation and it used for problems that involve supervised learning (NeuralWare 1995).

The goal of this study is to be able to correctly classify and separate different groups; in this case, failed or stressed firms versus nonfailed. The network is presented with both the input and expected output desired. The network learns from experience by modifying the weights of the connections in order to minimize the difference between the expected output and the network's output given the presented input. The network also learns by training through repeated exposure to a set of examples of the object or situation. Supervised learning requires that the network be presented with the correct responses for each input pattern. The term back propagation refers to the dynamic feedback of errors propagated backward through the network. The error values are used in adjusting the connection weights between nodes of the neural network. The back propagation network is guaranteed to converge to a local optimal set of weights (White 1989).

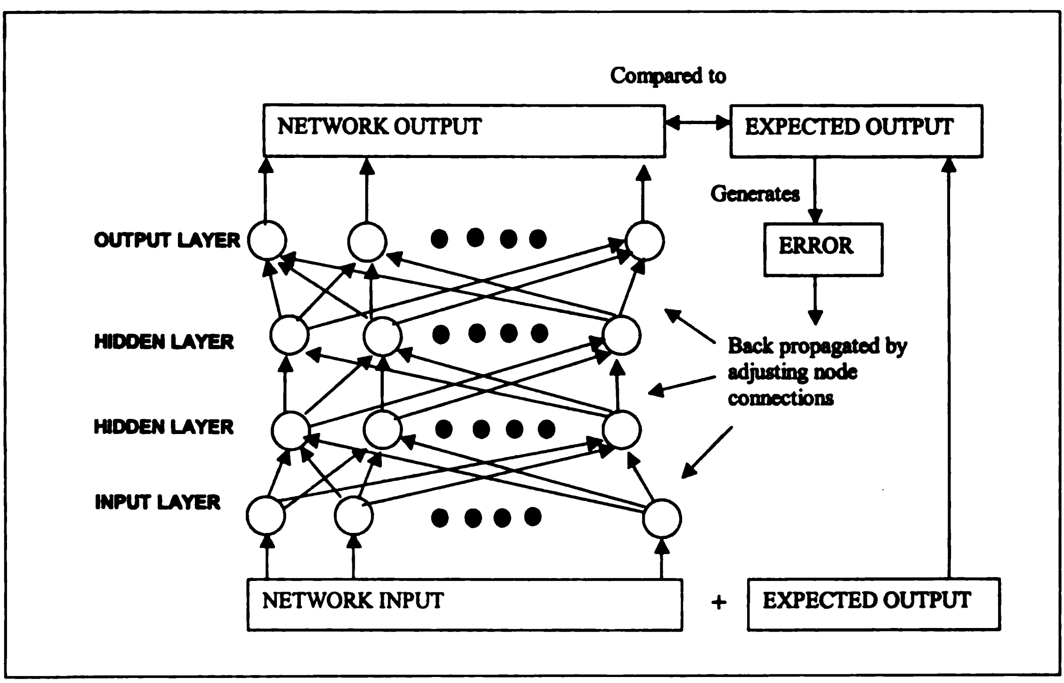
Neural networks have certain strengths not provided by other models (Udo 1993). These include the ability to: Tolerate noise or random data, or missing data, where all the data or rules are not known; self-organize and learn by changing the network connections; train by experience and dynamically adjust to changes in the environment; generalize from specific instances; and find and establish complex relationships among input variables.

Because of the smaller sample sizes and, in many cases, the frustrating problem of missing data or incomplete data (because carriers were delinquent in filing reports or reports filed were not complete), it was felt that the neural network model would be superior to the earlier models discussed above. In addition, there is no "zone of ignorance, " as noted above.

The Back Propagation (B-P) Method

The standard back propagation network has several elements.[4] They are an input layer, an output layer, and at least one hidden layer. Each layer is fully connected to each succeeding layer and each layer can contain any number of neurons or interconnections. Figure 1 depicts the feed forward of node values and the back propagation of error information.

Figure 1: Typical Back Propagation Network



The back propagation (BP) method is based on the gradient descent algorithm, an optimization method finding a global maximum minimum in a problem space. In this case it is the minimum of the global error. In the BP method, except for the input nodes, each node (call it n) generates an output value by applying a function to the weighted sum of the output values from all nodes in a lower layer (s) of nodes linked to node n in the next layer (s+1). Every node in layer s+1 will receive input from every node in layer s. Each node in layer s has a *different* set of weights applied to the input values. This process continues layer by layer to the output nodes. The resulting output from the neural network is thus the function applied to the weighted sum of the output values from the lower layer which are the function of the weighted sum of the output values from the next lower layer and so on.

An error function (it can be any error function) is used to evaluate the difference between the neural network output node values and the expected output. This error is back propagated to adjust the weights associated with each node. Since each node contributes to the global error, the adjustment made to its weights is directly proportional to the magnitude of the weight. Thus, the larger node weights get the most adjustment. The amount of the error of each node is based on the partial derivative of the error with respect to the node's output. This will have the effect of adjusting the weights between the individual nodes locally based on the global error E of the system. The change (delta) in the weight is further adjusted by the use of a factor called the learning coefficient. The learning coefficient affects the rate at which the BP method converges on the configuration of network weights that is the optimum for minimizing the global error.

In summary, the gradient descent method is based on making adjustments in each dimension (determined by the number of input variables) in a direction (partial derivative) that will minimize the global error. The size of the adjustment is based on the learning coefficient. In three dimensions, it is analogous to being on a terrain and moving in the x, y, and z directions that moves to the global minimum point. The learning coefficient is the size of the steps taken. Note that if the size of the steps is too large, the global minimum may be overstepped.

Mathematics of the B-P Model

As the network learns to classify, error information is also propagated back through the nodes and is used to update the weights assigned to each connection. To avoid confusion, clear notation is needed in order to describe the learning rule. Here the superscript is employed to note the layer of the network being considered. The following is the notation:

$x_j^{[s]}$ = current output state of the jth neuron layer

$w_{ji}^{[s]}$ = weight on connection joining the ith neuron in layer (s-1) to jth neuron in layer s

$I_j^{[s]}$ = weighted summation of inputs to jth neuron in layer s

The back-propagation element transfers its inputs as follows:

$$(1)\quad x_j^{[s]} = f\{\Sigma_I(w_{ji}^{[s]} \cdot x_i^{[s-1]})\}$$

$$= f(I_j^{[s]})$$

where f is can be any differentiable function. In this study, a hyperbolic tangent function (TanH) function was selected based on several factors and some trial runs.[5] That function is defined as

$$(2)\quad f(z) = (e^{z} - e^{-z})/(e^{z} + e^{-z})$$

where e is the naperian logarithm. Z is any real value. The network has a global error function, E, which is a differentiable function of all of the connection weights in the network. The actual error function is not as important as the critical parameter that is passed back through the layers. That function is

$$(3) \quad e_j^{[s]} = -\partial E / \partial I_j^{[s]}$$

This is a measure of the local error at processing element j in level s. If the chain rule of differentiation is used twice in succession, the relationship between the local error at a particular processing element at level s (jth node) and all the local errors at the level s+1 (kth nodes) is the following:

$$(4) \quad e_j^{[s]} = f'(I_j^{[s]}) \cdot \sum_k (e_k^{[s+1]} \cdot w_{kj}^{[s+1]})$$

Thus, e_j is the weight sum of the errors of the $e_k^{[s+1]}$ of all the nodes it connects to multiplied by its own slope $f'(I_j^{[s]})$.

If f is the hyperbolic tangent function as defined above, then its derivative can be expressed as a simple function of itself:

$$(5) \quad f'(z) = [1 + f(z)] \cdot [1.0 - f(z)]$$

Given equation (1), equation (4) can be rewritten,

$$(6) \quad e_j^{[s]} = (1.0 + x_j^{[s]}) \cdot (1.0 - x_j^{[s]}) \cdot \sum_k (e_k^{[s+1]} \cdot w_{kj}^{[s+1]})$$

given that the transfer function is hyperbolic tangent. The summation term in equation (6) which is used to back-propagate errors is analogous to the summation term in equation (1) which is used to forward propagate the input through the network. In other words, the inputs are forwarded through the layers to the output layer, errors are determined at the output layer, and then these errors are propagated back through the network to the input layer using equation (6) or (4). The multiplication of the error by the derivative of the transfer function scales the error.

The aim of this process is to minimize the global error E of the system. This is accomplished by modifying the weights locally. Given a set of weights $w_{ij}^{[s]}$, there must be a method to adjust them in order to decrease the global error of the model. A gradient rule is utilized to achieve this:

$$(7) \quad \Delta w_{ji}^{[s]} = -lcoef \cdot (\partial E / \partial w_{ji}^{[s]})$$

where *lcoef* is a learning coefficient. Each weight is thus changed according to the size and the direction of the negative gradient on the error surface.

Finally, the partial derivative in (7) can be calculated from the local error values described above, because, by the chain rule and (1):

$$(8) \quad \partial E / \partial w_{ji}^{[s]} = (\partial E / \partial I^{[s]}) \cdot (\partial I_j^{[s]} / \partial w_{ji}^{[s]})$$
$$= -e_j^{[s]} \cdot x_i^{[s-1]}$$

Combining (7) and (8) together results in the final function

$$(9) \quad \Delta w_{ji} = \text{lcoef } e_j^{[s]} \cdot x_i^{[s-1]}$$

Thus, the change of the connection weights between nodes j and i and levels s and s-1 are based on a learning coefficient which determines size of the error adjustment and the error at node j level s and the output from node i at level s-1. Remember that error j has been back propagated from the global error.

In sum, the lowest (or bottom) layer is called the input layer, and the last layer (or top) is the output layer. The layers in between are called middle layers and there can be any number of such levels. Neurons, the elements, in one layer feed-forward their input to the elements in the following layer. Each connection between the neurons has a weight attached to it. The weight represents the degree of influence of one neuron on other. Depending on the sign, a connection can reinforce or inhibit the link between the two neurons. A positive value indicates reinforcement and vice versa. As training set data is presented to the input layer, the neural network adjusts the connection weights until the error between the expected output and the network output is minimized. Through the connection weights the network memorizes the input.

Because the network connections change with time and experience (input data), it is necessary to define rules for how values going into a neuron affect the amount of signal generated and whether or not a connection is formed. Each neuron has an activation level associated with it. This is the level of signal that must be reached before the neuron emits a signal. If the neuron is at the input layer, the activation level is determined by the input received directly from the environment. For a cell in the middle or output layer, the activation level is determined by taking the weighted sum of the signals from cells connected to it at a lower layer. This computation is known as the transfer function or activation function noted above. The hyperbolic function used in this research generates values between -1 and 1. An alternative function typically used in this situation is the sigmoid function that generates values between 0 and 1. Both of these functions are suited to neural networks (NeuralWare 1995) since: (a) They are smooth continuous functions, (b) the function outputs are within the range 0 to 1 or –1 to 1, (c) they are nonlinear which helps to solve linearly inseparable problems, (d) they are useful in multilayer networks, and (e) they are easy to deal with mathematically.

The Sample

The initial sample consisted of a matched group of bankrupt small air carriers and solvent carriers.[6] The study time horizon encompassed the years, 1982-1998. An insolvent carrier was paired with a solvent carrier of equivalent size (using operating revenues). Of the hundred or so carriers filing for bankruptcy since the year 1982, there was sufficient data for only 32 carriers that actually filed for bankruptcy protection under federal bankruptcy law. There were two main reasons for this; the carrier failed before supplying enough financial information to USDOT or the carrier was exempt from the full USDOT financial filing requirements and thus only filed very abbreviated statements. As noted previously, this limitation confirmed the efficacy of the use of the neural network methodology.

The total number of observations was 65. Initially, the data set consisted of 32 bankrupt and 33 nonbankrupt carriers. Upon examining the random sample of matched solvent carriers, however, it became obvious that some were clearly financially stressed (that is, they had negative equity ratios, and in some cases negative Altman Z Scores—an indication of severe financial stress). The presence of negative equity is defined as "technical insolvency" even though a carrier has not actually filed for bankruptcy.[7] This is the classic problem in this industry. Many "solvent" carriers operate in such a weakened financial condition but manage to survive because their competitors are also weak. It was thus decided to classify these carriers as distressed, or in a "bankruptcy profile," as Altman (1968) refers to them. There were a total of 11 such airlines in the database.

The final sample thus consisted of 43 bankrupt and "distressed" carriers. The rest (22) were solvent and non-distressed. Data gathered for the input level included 21 pieces of financial data from carrier balance sheets and income statements.[8] In addition, several key financial ratios found to be predictive in the earlier noted studies of airlines were included in the training set on the grounds that the more variables provided, the more efficient would be the final neural network.[9] These included the ratios of working capital to total assets—a liquidity ratio, retained earnings to total assets—a profitability ratio, operating profits to total assets, another profitability ratio, and book value of equity to book value of debt, a financial leverage measure, and operating revenues to total assets-a turnover ratio. Also included was the Z score as defined by Altman (1968).

The use of the neural network requires separating the initial sample data set into two distinct data sets: a training set and a holdout sample. The training set is used to train the neural network by repeatedly presenting the network with instances of the target population—in this case, bankrupt and nonbankrupt carriers—and the expected output for each instance. The second set, the holdout sample, is also comprised of examples of the target population. This holdout sample is used to test predictability of the network created by the first set. The use of the holdout sample is an issue on which there is no clear approach. The purpose of the holdout sample is to validate or judge the quality of the model developed. One of the issues to be resolved in using this approach is how to determine the size and makeup of the training set and the holdout sample. In terms of size, the holdout sample should contain enough instances of the different items being classified. In terms of makeup, the holdout sample should contain roughly the same makeup as the target population and as the training set. This requires that each holdout sample be examined to see if it represents the desired makeup. In addition, it is possible that some key training data may only be included in the holdout sample. The leave-k approach addresses these two issues. The leave-k approach is useful on small data sets and uses most of the data for training. This approach is suggested by Hecht-Neilson (1990) and Timothy Masters (1993).

To further validate the robustness of the model, a "jackknife" or Laffenbrach procedure (Swingler 1996) was performed. In this procedure, the sample was randomized and a single observation withheld. The neural network was run on the remainder and then used to classify the single withheld observation. That observation once classified was then reinserted into the set and another selected, and so on. This process was continued until each member of the sample set had been used as the holdout sample set.

Another problem in developing the neural network is determining the number of hidden layers, the number of nodes per hidden layer, and the type of training rule utilized. These questions are usually determined through trial and error by manipulating the number of layers and nodes, and varying the training rules, until there is negligible improvement observed in the performance of the network. The authors examined many different learning rules, hidden layers, and transfer rules until the highest level of accuracy was attained. In each case, the process continued through 150,000 iterations. The best transfer function was a hyperbolic tangent function (TanH) that maps the output to a range of –1 to 1. The optimal output was achieved using only one hidden layer and three hidden neurons.

Results

To classify the carriers, each observation was classified as either (0,1) for a bankrupt or financially stressed carrier or (1,0) for a solvent or healthy carrier. Since the output values of the network are not discrete, they were rounded up or down. That is, any value less than or equal to 0.5 was considered to be "0" and any number above that was rounded to "1."[10] Table I presents the results of the best run of the network. B stands for bankrupt or distressed; NB for nonbankrupt or solvent. Appendix A lists the carriers in the sample. They are coded in the table for legal reasons.

The network classified 77% of the total

Table 1: Results of the Best Run of the Neural Network

Carrier	Classification	Expected Output	Neural Network Output	Neural Network Performance
1	B	01	-0.02, 1.02	Correct
2	NB	10	0.89, 0.11	Correct
3	B	01	0.88, 0.12	Type I
4	B	01	-0.02, 1.02	Correct
5	B	01	0.04, 0.96	Correct
6	B	01	0.75, 0.25	Type II
7	NB	10	0.02, 0.98	Type I
8	NB	10	0.30, 0.70	Type I
9	B	01	-0.01, 1.01	Correct
10	B	01	0.29, 0.71	Correct
11	NB	10	1.08, -0.08	Correct
12	B	01	-0.02, 1.02	Correct
13	B	01	-0.02, 1.02	Correct
14	B	01	-0.03, 1.03	Correct
15	NB	10	0.69, 0.31	Correct
16	B	01	-0.02, 1.02	Correct
17	NB	10	0.71, 0.29	Correct
18	NB	10	0.99, 0.01	Correct
19	B	01	-0.03, 1.03	Correct
20	B	01	-0.02, 1.02	Correct
21	B	01	0.76, 0.24	Type I
22	B	01	-0.01, 1.01	Correct
23	NB	10	0.52, 0.48	Correct
24	B	01	-0.02, 1.02	Correct
25	NB	10	1.12, -0.12	Correct
26	B	01	-0.01, 1.01	Correct
27	B	01	-0.02, 1.02	Correct
28	B	01	-0.02, 1.02	Correct
29	B	01	-0.01, 1.01	Correct
30	B	01	-0.02, 1.02	Correct
31	NB	10	1.01, -0.01	Correct
32	NB	10	0.79, 0.21	Correct
33	B	01	0.20, 0.80	Correct
34	B	01	0.92, 0.08	Type II

Table 1, continued

Carrier	Classification	Expected Output	Neural Network Output	Neural Network Performance
35	B	01	-0.01, 1.01	Correct
36	NB	10	0.85, 0.15	Correct
37	B	01	-0.02, 1.02	Correct
38	B	01	-0.12, 1.12	Correct
39	B	01	-0.02, 1.02	Correct
40	B	01	0.04, 0.96	Correct
41	NB	10	0.05, 0.95	Type I
42	B	01	0.46, 0.54	Correct
43	B	01	0.40, 0.60	Correct
44	B	01	-0.02, 1.02	Correct
45	B	01	1.10, -0.10	Type II
46	B	01	0.16, 0.84	Correct
47	NB	10	-0.01, 1.01	Type I
48	B	01	-0.01, 1.01	Correct
49	NB	10	0.28, 0.72	Type I
50	B	01	0.05, 0.95	Correct
51	B	01	0.06, 0.94	Correct
52	NB	10	1.00, 0.00	Correct
53	NB	10	1.01, -0.01	Correct
54	NB	10	0.90, 0.10	Correct
55	B	01	-0.02, 1.02	Correct
56	B	01	-0.02, 1.02	Correct
57	NB	10	1.06, -0.06	Correct
58	NB	10	0.47, 0.53	Type I
59	NB	10	-0.05, 1.05	Type I
60	B	01	-0.01, 1.01	Correct
61	B	01	0.56, 0.44	Type II
62	B	01	0.96, 0.04	Type II
63	NB	10	1.00, 0.00	Correct
64	B	01	-0.01, 1.01	Correct
65	B	01	0.87, 0.13	Type II

Activation Function Used: Tanh Iterations: 150,000 Number of Layers: 3
Input Layer Nodes: 21
Hidden Layer: 1
Number of nodes: 3 Output layer nodes: 2
B – Bankrupt or distressed carrier
NB – Nonbankrupt or solvent carrier

sample accurately. Two types of errors are present: Type I and Type II. A Type I error occurs when a bankrupt carrier is incorrectly classified as solvent. A Type II error results when a nonbankrupt carrier is classified as failed. Of the total 15 errors, 7 were Type I and 8 were Type II. Thus the successful classification rate for each group of carriers was:

Number Correct	50	77%
Number of Type I Errors	7	11%
Number of Type II Errors	8	12%
Total in Sample	65	100%

As indicated the overall success rate of the model was 77% (50/65). More importantly, the success rate of predicting bankruptcy or distress was 89% (or 100%-11%, the failure rate) and that of solvency 88%.

To further validate the results, a nonparametric binomial distribution test was performed. The probability of getting 7 or fewer errors, the Type I error rate, was found to be equal to .0004%. The probability of getting 8 Type II errors or less was slightly higher, 0.143%, but still statistically significant. The neural network thus provides an accurate classification of air carriers into the two categories.[11]

In a comparative study, research by Coats and Fant (1993) demonstrated the superiority of neural networks over MDA and logit/probit models in forecasting insolvency across different lines of industry. This study confirmed that superiority. The small carrier network model developed here clearly outperformed the authors' MDA AIRSCORE Model. The neural network's success rates of 77% overall, with 89% for bankruptcy and 88% for solvency, exceeded that of the MDA model's 73%, 76%, 72%, respectively, with a "zone of ignorance" of 20% (Chow, Gritta, and Leung 1991, Table 3). The neural network also outperformed the basic Altman Model, whose 76% success rate is deemed to be a benchmark in the financial literature.[12]

Conclusion

This article has applied a powerful technique, a neural network, to the problem of identifying financially distressed smaller carriers, known as large and medium regional airlines. The neural network has several advantages not provided by discriminant models, such as MDA and logit/probit. It can tolerate noise and missing data, self organize and learn by changing network connections, generalize from the specific to the general, and establish complex relationships among input variables.

The airline industry has seen its share of financial distress in the past. The number of bankruptcies among major carriers is noteworthy (Braniff, Continental, Eastern, TWA, PanAm, and the near bankruptcy filings by Northwest and USAir). The failure of so many smaller carriers, however, is just as disconcerting. Most of the carriers have had troubled histories due in large part by high operating volatility and excessive debt finance (Gritta, Freed, and Chow 1998). This fact alone makes the separation into the strong and the weak difficult. The neural network, however, has proven to be a powerful tool in this endeavor as the current study has documented. The "black box" nature of the neural network, however, does limit our understanding or knowledge regarding just how the method solves a particular problem, such as forecasting financial stress. Further light may be shed on this by examining the network connection weights to determine the contribution of each node to the output and manipulating these weights.

Promising areas of new research include chaos theory and survival analysis, which includes environmental factors and biological mechanisms. The ability of a firm to survive in the airline industry might also be explored using evolutionary-based computational models such as genetic algorithms. The authors intend to explore these models in future research as a means of further increasing forecasting accuracy.

Appendix A

Bankrupt Air Carriers	Non-Bankrupt Air Carriers
Air Florida	Air Cal
Air North	Air Midwest
Air One	Air Wisconsin
Altair	Alaska International Air
American International	Arrow
Apollo	Aspen Air
Bar Harbor Air	Big Sky Airlines
Business Express	Empire Airlines
Capitol Air	Evergreen Air
Cascade Air	Florida Express
Cochise Air	Great American Air
Flagship Express	Great Northern Airlines
Golden Gate	Hawaiian Air
Golden West	Hughes Air
Grand Airways	Imperial Air
Gulf Atlantic Air	Jet America Air
Hermans Air	Kodiak Air
Imperial Airlines	Mississippi Valley Air
Kiwi International	Munz Northern Air
L'Express	Muse Air
Mark Air	North Central Air
Metro Airlines NE	Pacific Express
New York Air	Pacific Southwest Express
Northwest Executive	Pilgrim Airlines
Pan Am Express	Reeve Air
States Air West	Reno Airlines
Sun West Airlines	Rocky Mountain Airlines
Swift Air	Skywest Air
US Africa Airways	Southern Air
Virgin Island Air	Sun World
Wien Air Alaska	Trans States Air
Wright Air	Zantop Air

Endnotes

1. USDOT classifies carriers by groups based on total dollar operating revenues. Major carriers have revenues of $1.0 billion or larger, nationals, $100.0 million to $1.0 billion, large regionals from $20.0 million to $100.0 million, and medium regionals from $0 to $20.0 million. This study centers on the latter two groups.

2. A preliminary version of this article was presented at the 4th Air Transport Research Group Conference (World Conference on Transportation Research) in Amsterdam, The Netherlands, in July 2000.

3. The "zone of ignorance" contains both failed and nonfailed firms, but it is difficult to separate them. The accuracy of the model (either Altman or AIRSCORE) can be increased by increasing this zone. To do so, however, results in fewer firms being classified. Thus, there is a trade-off. With AIRSCORE, the researchers varied the "zone of ignorance" in order to achieve Altman's standard of 76% success in forecasting insolvency, as that rate has become a benchmark in the finance literature. For details, the interested reader is referred to (Altman 1968) and (Chow, Gritta, and Leung 1991). The neural network does not suffer from this limitation. There is no "zone of ignorance." This is a major reason for using the methodology.

4. This section of the article relies heavily on the back propagation algorithms developed for artificial neural networks research. See Haykin (1998) and Swingler (1996) for details. What follows is a summary of the pertinent models.

5. The choice of the transfer function can affect the convergence of the neural network to a solution and the resulting algorithm that is used to implement the mathematics of the neural network. The candidate functions are the sigmoid and hyperbolic tangent (TanH). The TanH function was selected for a variety of reasons (Swingler 1996): (1) The range of values of the TanH function is –1 to +1. Data can be normalized to fall within this range with a zero mean and unit standard deviation, (2) the TanH function can be computed faster than the logistics function, (3) TanH leads to faster learning than the logistics function, (4) it is a continuous, real-valued function whose domain is real, with a positive derivative and with a bounded range, and (5) values at the extreme ends of the input range have less impact than values near the mid-range where the derivative of the function reaches its maximum. In addition, it provided the best overall results when run on the data.

6. Using all carriers as the base data set would undermine the validity of the discriminator. For instance, if the percentage of financially healthy carriers out of the total population were 90%, a discriminator that classified every carrier as financially healthy would be accurate 90% of the time. There would be no guarantee that the outcome of any high scoring discriminator is valid.

7. Technical insolvency is normally the first stage in bankruptcy. A firm has liabilities (debt) that exceed assets, and hence has a negative equity position. This can prompt creditors to file claims against the firm, thus precipitating a bankruptcy filing. The reclassification of these 11 carriers was necessary in that the neural network was not performing as expected. The NN was being presented with contradictory instances of data using the bankruptcy criteria as a determination of financial health. This affects the outcomes of training and the reclassification of the 11 carriers improved significantly the performance of the network. It should be noted that this reclassification does not affect how the neural network learns (trains). The learning method remains the same: back propagation. It does, however, affect the values of the connection weights within the network. The authors decided to continue using the 65 carriers and not to add more to the sample in order to accommodate a wider range of financially distressed carrier profiles and counteract any tendency of the network to overfit the data.

8. Income statement items included as variables were operating revenues, depreciation, total operating expenses, earnings before taxes and interest, income before taxes, income after taxes, and net income. Balance sheet accounts included were cash, receivables, current assets, total assets, current liabilities, taxes, total noncurrent liabilities, total deferred credits, total liabilities, retained earnings, and equity. These financial statistics were selected because they are the most useful income statement and balance sheet items necessary to an in-depth financial analysis of any airline. From this data, virtually any ratio that measures financial condition can be constructed. In addition, these were the data gathered for the major carrier model (Davalos, Gritta, and Chow 1999). There are, of course, other factors that could increase the explanatory ability of the NN. Operating statistics (such as load factor, stage length, and other traffic data) may be added in future research by the authors.

9. A heuristic in the development of neural networks is that the more information provided to the neural network, the better is the performance. While the authors were aware that several variables were key in contributing to configuration of the neural network and the outcome of the network, they also wanted to identify and evaluate the contributing effect of other nonkey variables. With respect to the accuracy of the prediction, the neural network is stopped when either a target RMS (root mean squared) value is reached or the designated number of iterations is reached. In all cases the RMS value was no greater than 0.0125. Linearity is not assumed with neural networks, therefore it is possible that several input variables can have combined effects upon the outcome variable. It is the intention of the authors to accommodate such possible interactions by using a larger set of variables. Examination of the neural network will allow the determination of which variables contributed to the outcome and such variables can either be eliminated or their input connections pruned from the network. This will result in increasing the performance and generalization of the neural network. The performance is improved by eliminating extra calculations and the generalization is improved by minimizing the ability of the network to memorize the correct responses to the input.

10. Since the neural network uses the Tanh activation function, which produces continuous values, it is not possible to produce the output values of 0,1 and 1,0. Thus, the output values are decimal-valued. However, the output values do not necessarily correspond to the degree of financial health. The neural network learns from the input, and internal weights are adjusted to produce an output that matches the expected output. Since the neural network adjusts its internal weights after every data point presented, the neural network reflects a state of weighted connections that can best approximate the expected output. It is not necessarily true that decimal-valued output reflects degrees of financial health, the output generated is based on how the neural network reacts to the different patterns of input. It is possible for a financially healthy carrier to generate output values that are closer to 0,0 than to 0,1 if its pattern of input variables is substantially different from the other financially healthy carriers.

11. The results obtained were similar to those found in the preliminary study noted above (Gritta et al. 2000). That study achieved an overall success rate of 88%. While that rate exceeds that in the current study, the authors feel that the very small sample size (only nine bankrupt and nine nonbankrupt carriers) limits the validity of that prior study to some degree. The 88% success rate at forecasting financial stress was about the same as the prior study's 91%. The authors are more confident in the model developed with this larger sample and the statistical tests employed.

12. When the small carrier data was input to the large carrier NN (Davalos, Gritta, and Chow 1999), the authors were never able to achieve a predictive rate above that due to chance. The accuracy rate was only 54%. This research thus confirms the necessity of developing a separate model for the small air carriers. At this time, the authors are unable to explain the failure of the original large carrier NN to predict small carrier financial stress. The smaller carriers face different competitive environments, use different aircraft types, etc., and these factors may explain the failure of the large carrier model.

References

Altman, E.I. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *Journal of Finance* 23(4), (1968): 589-609.

Caudill, M. "Neural Network Primer." *AI Expert* 4, (1989): 61-67.

Chow, G., R. Gritta, and E. Leung. "A Multiple Discriminate Analysis Approach to Gauging Air Carrier Bankruptcy Propensities: The AIRSCORE Model." *Journal of the Transportation Research Forum* 31(2), (1991): 371-377.

Coats, P. and F.L. Fant. "Recognizing Financial Distress Patterns Using a Neural Network Tool." *Financial Management* 22(3), (1993): 142-155.

Cox, E. "Applications of Fuzzy Systems Models." *AI Expert* 7(10), (1992): 34-39.

Davalos, S., R. Gritta, and G. Chow. "The Application of A Neural Network Approach to Predicting Bankruptcy Risks Facing the Major U.S. Carriers: 1979-1996." *Journal of Air Transport Management* 5, (1999): 81-86.

Dutta, S. and S. Shekhar. "Bond Ratings: A Non-Conservative Application of Neural Networks." *Proceedings of the IEEE International Conference on Neural Networks* 2, (1998): 443-450.

Gritta, R. "Bankruptcy Risks Facing the Major U.S. Air Carriers." *Journal of Air Law & Commerce* 48, (1982): 89-108.

Gritta, R., E. Freed, and G. Chow. "Measuring the Degrees of Operating and Financial Leverage." *Transportation Law Journal* 26(1), (1998): 51-71.

Gritta, R. et al. "Forecasting Small Air Carrier Bankruptcies Using a Neural Network Approach: A Preliminary Study." *Journal of Financial Management and Analysis* 139(10), (2000): 44-49.

Haykin, Simon S. *Neural Networks: A Comprehensive Foundation.* Prentice Hall, 1998.

Hecht-Neilson, R. *Neurocomputing.* Addision Wesley, 1990.

Kamijo, K., T. Tanigawa. "Stock Price Pattern Recognition: A Re-Current Neural Network Approach." *Proceedings of the International Joint Conference.* San Diego, (1990): i215-i221.F

Lee, K.C., I. Han, and Y. Kwon. "Hybrid Neural Network Models for Bankruptcy Prediction." *Decision Support Systems* 18, (1996): 63-72.

Masters, Timothy. *Practical Network Recipes in* C++." Morgan Kaufmann. 1993.

NeuralWare. *Advanced Reference Guide: Software Reference for Professional II/PLUS and NeuralWorks Explorer.* NeuralWare Inc., Technical Publications Group, Pittsburgh, PA, 1995.

Rumelhart, D.E. and J.L. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* MIT Press, Cambridge, MA, 1986.

Russell, R., and R. Marks. *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks.* MIT Press, Cambridge, MA, 1999.

Swingler, K. *Applying Neural Networks: A Practical Guide.* Academic Press, 1996.

Tam, K., and M.Y. Kiang. "Managerial Applications of Neural Networks: The Case of Bank Failure Prediction." *Management Science* 38, (1992): 926-947.

Udo, G. "Neural Network Performance on the Bankruptcy Classification Problem." *Computer and Industrial Engineering* 25, (1993): 377-380.

White, H. "Some Asymptotic Results for Learning in Single Hidden Layer Feedforward Network Models." *Journal of the American Statistical Association* 84, (1989): 1003-1013.

Richard D. Gritta is professor of finance and transportation at the Robert B. Pamplin Jr., School of Business Administration, University of Portland, Portland, OR. He holds the following degrees: BBA, University of Notre Dame, MBA, Indiana University, and Ph.D., University of Maryland. He has published numerous articles in transportation, finance, economics, and international journals. He is currently associate editor-air transportation for the Journal of the Transportation Research Forum. *He was a Senior Fulbright Fellow at the Transportation Institute-Warsaw School of Economics in 1996, and advises Sen. Ron Wyden (D-OR) who currently serves on the US Senate Subcommittee on Aviation Affairs. His research interests include the financial and economic problems in air transportation, airline financial analysis, and airline financial management strategies.*

Sergio Davalos, currently a visiting professor at Portland State University, is associate professor of computer science at the University of the Virgin Islands, having taught at the University of Portland and Portland State University in Oregon. He holds BBA, MBA, and Ph.D. degrees from the University of Arizona. He has published more than a dozen articles in transportation, finance, and computer science journals. He has taught courses in artificial intelligence and directs the Caribbean CSC Asynchronous Distance Learning Program at UVI.

Garland Chow is associate professor of logistics and transportation at the University of British Columbia, Vancouver, B.C. He holds BBA and MBA degrees from the University of Maryland, and a Ph.D. from Indiana University. Chow has also been a visiting professor at the Warsaw School of Economics, Arizona State University, and the University of Sydney in Australia. He has published numerous articles in transportation, logistics, and finance journals. Chow is on the editorial boards of the Transportation Research-Part E, *and the* International Journal of Physical Distribution. *He has international consulting experience in the area of transportation and business logistics. His interests include the financial and economic problems in air transportation, air carrier financial analysis, motor carrier economics and policy, and third party logistics.*

Marcus Wang has been a research associate at the University of Portland. He holds a BS from Tung Ji University in China and MBA from the University of Portland, and he has been accepted into the Ph.D. program at the University of Oregon. He has published several articles in transportation and finance journals.