

Neural network forecasting for airlines: A comparative analysis

Lawrence R. Weatherford,* Travis W. Gentry and Bogdan Wilamowski

Received (in revised form): 16th September, 2002

*University of Wyoming, PO Box 3275, Laramie, WY 82071, USA
Tel: +1 307 766 3639; Fax: +1 307 766 3488; E-mail: lrw@uwyo.edu

Larry Weatherford is Professor at the University of Wyoming. He holds a PhD from the University of Virginia. He has received several outstanding teaching awards and also has a best-selling textbook, *Decision Modeling with Microsoft Excel*, published in 2001. He has published over 20 scholarly articles and presented more than 50 papers on five different continents to professional organisations. He has consulted with such major corporations as Walt Disney World, Hilton Hotels, American Airlines, Northwest Airlines, Lufthansa German Airlines, Swissair, Scandinavian Airlines, Air New Zealand, South African Airlines, Unisys Corporation and Choice Hotels.

Travis Gentry graduated with an MBA in 1995 and a BS in Electrical Engineering in 1993 from the University of Wyoming. Since that time, he has worked as an analyst at InSitu and ProSavvy.

Bogdan Wilamowski is Professor of Electrical Engineering/Computer Science at the University of Idaho. He is President Elect of the IEEE Industrial Electronics Society and the Vice President of the IEEE Neural Network Society. He is also Associate Editor of several IEEE Transactions (including neural networks). He is the author of four textbooks, more than 240 refereed publications and 27 patents.

ABSTRACT

KEYWORDS: forecasting, neural networks, airlines

This paper provides the first published research paper on the technique of neural network forecasting as applied to the airline industry. It compares this new method with the traditional forecasting techniques (moving averages, exponential smoothing, regression, etc.). The data were provided by a major international carrier. All the methods were compared on the basis of a standard error measure — mean absolute percentage error (MAPE). The results of the study are promising. The most basic neural network structures provided better forecasts than traditional forecasting methods.

INTRODUCTION

Having attended the meetings of the Airline Group International Federation of Operations Research (AGIFORS) Study Group on Yield Management and Reservations, the authors realise that there has been a dearth of research on new forecast methods and ideas. Certainly, no topic could have greater impact in this realm than improving forecasting accuracy. All the mathematically sophisticated overbooking models, seat allocation models, network models have at their heart a reliance on an accurate forecast of demand or no-show rates or some other input. It was therefore decided to pursue some research

along the lines of looking for improved forecasting methodology that may have been developed in separate fields, which would have some positive impact on airline forecasts (Gentry and Weatherford, 1995).

Forecasting has been called both an art and a science. It is an ability to recognise patterns through a logical and analytical approach. There is no philosophy more important in business than gaining an advantage or opportunity over the competition. Forecasting permits airlines to estimate unknown variables (eg demand for seats) with a certain degree of reliability.

New technologies and tools are constantly entering the airline industry. For an airline to remain competitive, it must be aware of all opportunities to improve its methods of doing business. One recent technological advancement that has begun to provide airlines with an array of opportunities in pattern recognition is the neural network. The network represents an opportunity to solve many airline-specific problems (including scheduling, optimisation and forecasting) more accurately.

The purpose of this paper is to compare the forecasting ability of traditional forecasting methods and the forecasting ability of various neural network structures to predict the total number of passengers booked on a given flight. The study measured the methods' abilities for both a short-term and a long-term forecast. This next section will describe the data format, provide an overview of the forecasting

methods, and review the accuracy measurement methods used to evaluate the forecasting ability.

THE DATA

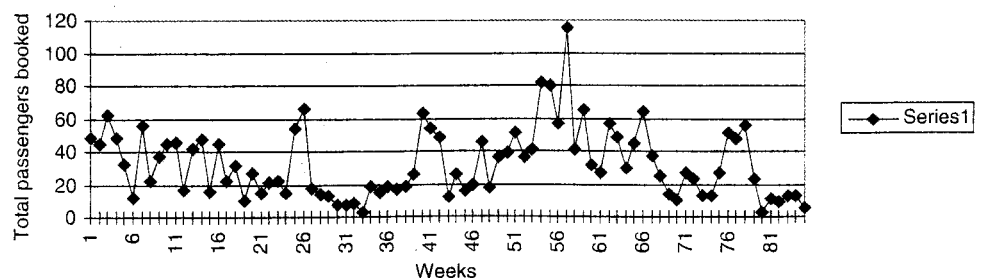
The data were a set of time-series numbers provided by a major international carrier. The information represented the final number of reservations on a specific flight described by two signatures: day of week and passenger fare class. Any distinguishing characteristics about the data's source (exact flight numbers, type of aircraft, origin/destination, specific fare class and time of week, month or year) were disguised. A total of 85 weeks of data (over 1.5 years) were collected. These data are shown in Figure 1.

The scope of the research was narrowed to studying the abilities to forecast the final number of reservations on the flight for both a short-term (one week ahead) and a long-term (three weeks ahead) horizon. The accuracy of each forecasting model will be evaluated based on its performance of predicting the actual values during the holdout sample (weeks 31–85). In the accuracy measurements, the first 30 points are referred to as the 'training set' and the last 55 points are the 'holdout' sample. By observing the graph in Figure 1, it was determined that no obvious seasonality was present in the data.

TRADITIONAL FORECASTING MODELS

Three main methods of quantitative forecasting were used: moving averages; exponential smoothing; and regression.

Figure 1: Graphical representation of data set



Moving averages

Two different types of moving averages were studied in this research. The assumption behind these models is that the average performance of the recent past is a good predictor of future performance. These models are good at removing the effects of random fluctuations.

Simple

The simple moving average forecast is shown in equation (1).

$$\hat{X}_{t+1} = \frac{X_t + X_{t-1} + X_{t-2} + X_{t-3} + \dots + X_{t-N+1}}{N} \quad (1)$$

The \hat{X}_{t+1} term is the estimated value of the forecast one period ahead ($t+1$). The other values of X_t are the previous period's actual values up to a given number of N past periods. The key decision in using moving average models is the proper selection of the number of past periods, N , to use. Values of N ranging from two weeks to eight weeks were explored.

Weighted

The weighted moving average permits the forecast to emphasise more recent demand over earlier demands. The sum of the weights in a weighted moving average usually equals one. Equation (2) illustrates the weighted moving average forecast and shows how each historic demand has an associated weight.

$$\hat{X}_{t+1} = W_1X_t + W_2X_{t-1} + W_3X_{t-2} + W_4X_{t-3} + \dots + W_NX_{t-N+1} \quad (2)$$

Exponential smoothing

In the study, single exponential smoothing and exponential smoothing with trend techniques were both used. A major advantage of using exponential smoothing tech-

niques over the moving average techniques is the smaller amount of stored data and calculations required.

Single exponential smoothing

Single exponential smoothing methods are similar to weighted moving average techniques. Exponential smoothing calculates the average of a time series by allocating more weight or importance to recent periods of data. Equation (3) illustrates the single exponential smoothing method (Makridakis *et al.*, 1983).

$$\hat{X}_{t+1} = \alpha X_t + (1 - \alpha)\hat{X}_t \quad (3)$$

The value of alpha (α) can range from zero to one. The higher the value of alpha, the more weight allocated to the most recent data observation.

Exponential smoothing with trend

Exponential smoothing with trend is similar to single exponential smoothing except a trend component is introduced into the forecast. The trend component is defined below in equation (4).

$$T_t = \beta(\hat{X}_t - \hat{X}_{t-1}) + (1 - \beta)T_{t-1} \quad (4)$$

Similar to the alpha value, beta (β) is a smoothing constant between zero and one. A higher value of beta indicates a larger portion of the most recent trend being added to the next period's forecast. The complete formula for exponential smoothing with trend is provided in equation (5).

$$\hat{X}_{t+1} = \alpha X_t + (1 - \alpha)\hat{X}_t + T_t \quad (5)$$

Regression

Regression is a causal forecasting method that fits curves to the entire data set to minimise the forecasting errors. Time (measured in weeks) is the independent variable (x axis) and total passengers booked is the dependent variable (y axis).

This paper looked at three forms of regression techniques: linear, quadratic and cubic.

Linear regression

Linear regression is nothing more than fitting a straight line to the data set. The line consists of two elements, the slope (b) and y intercept (a). The general equation form is given as equation (6).

$$\hat{X}_t = a + bt \quad (6)$$

Quadratic regression

Similar to linear regression, quadratic regression fits a second-order polynomial curve to the data. The model allows for one bend or inflection forming a U(or \cap) shaped curve. The general form is provided below as equation (7).

$$\hat{X}_t = a + bt + ct^2 \quad (7)$$

Cubic regression

The final method used is the cubic regression model. The model fits a third-order polynomial curve to the data. The third-order curve is characterised by two bends or inflections typical in an 'S' shaped curve. The cubic regression curve is described by equation (8).

$$\hat{X}_t = a + bt + ct^2 + dt^3 \quad (8)$$

Measuring forecasting accuracy

Accuracy can be defined as 'goodness of fit' or how well the forecasting model is able to reproduce data that are already known. The error can be defined as the difference between the actual value and the forecasted value as shown in equation (9).

$$e_t = X_t - \hat{X}_t \quad (9)$$

Given this definition of error, there have been many other standard, statistical measures that have been defined: mean error,

mean absolute error, sum of squared errors, mean squared error, standard deviation of errors, etc. This study used mean absolute per cent error (MAPE), as it has been cited by Weatherford and Kimes (2002) as an appropriate error measure in revenue management situations. The MAPE is the absolute difference of the error expressed as a percentage, thus showing a percentage of the error made by the forecast. The MAPE equation is provided in equation (10).

$$MAPE = \frac{\sum_{t=1}^n \left(\frac{|X_t - \hat{X}_t|}{X_t} \right)}{n} \times 100\% \quad (10)$$

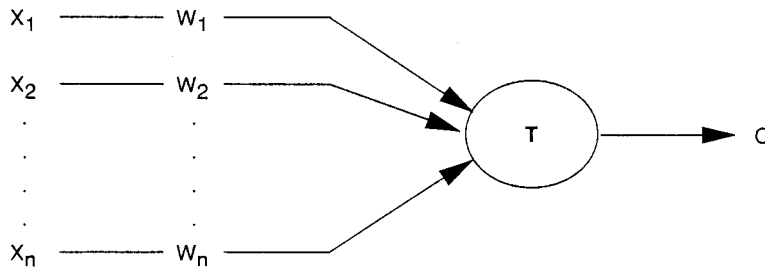
NEURAL NETWORKS

Neural networks represent a promising generation of intelligent machines that are capable of processing large and complex forms of information. The pursuit of the neural network started when McCulloch and Pitts (1943) developed the first formal synthetic neuron model. The model operates by individual inputs entering a system whereby each input (X_i) is multiplied by a unique weight (W_i). The summation of the weighted inputs are compared with the neuron's threshold value (T) to determine whether the neuron is activated or not. Figure 2 illustrates the model.

In its most basic form, a neural network is nothing more than a powerful classifier of patterns. The network is capable of classifying any linear or non-linear orientation. It is capable of quickly recognising patterns and variable associations. Similarly structured to the human brain, a network is composed of serially arranged neurons connected in parallel. Each network starts with an input layer and ends with an output layer. Depending upon the complexity of the classification, several intermediate or hidden layers may be needed and inserted between the input and output layers.

A single neuron is capable of linear

Figure 2: McCulloch–Pitts neuron model

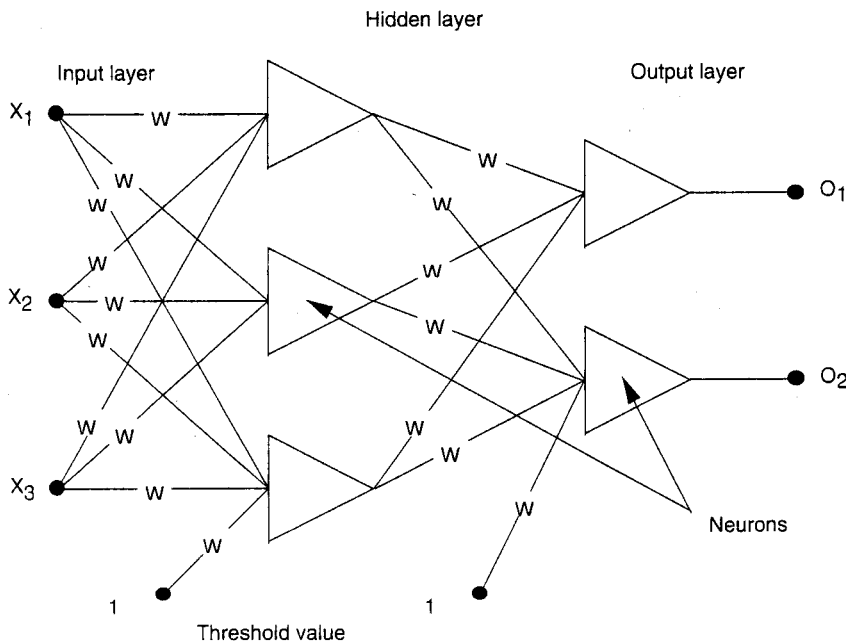


separation. For complex, higher-order classification, multiple neuron and multiple layered networks are required. In these multiple-layered networks, the output of one layer becomes the input to the next layer. As the data are transferred, the network assigns a weight value to each neuron connection.

Figure 3 provides a generalised depiction of a neural network. The model illustrates an input and output layer separated by one hidden layer. Each neuron layer is connected by various, individualised weights

designated by W (the values of W are not necessarily the same). The training of the neural network assigns more positive weights to more important data (excitatory) and assigns more negative weights for less important data (inhibitory) (Hruschka, 1993). Every neuron sums the products of the inputs (X_1, X_2, X_3) and associated weights (W). The sum of each neuron is then compared with its threshold value to determine whether that neuron will be activated. The type of the neuron's activation function (discrete, binary, con-

Figure 3: Generalised neural network model



tinuous, etc.) determines the classification characteristics and abilities of the neural network.

Training

The neural network is a technology that learns: it learns from being trained. Training occurs when patterns of given inputs and known outputs are repeatedly applied to the network. Through the repetition, the network iteratively adjusts each weight until the difference between the desired or expected output and actual output is below a predetermined value. The difference between the desired and actual output values is called the error (Proctor, 1992). Once the learning is complete, the network should be able to identify correctly any pattern close to the pattern it was trained with and match the input pattern to an output pattern. The training technique used in this study was a modified regression method that leads to an immediate solution (Anderson and Wilamowski, 1995). In this study, a Gaussian distribution was used to place more emphasis on the more current data and less emphasis on the older data.

Network structures

There is literally an unlimited number of neural network structures that can be designed. The structures can vary in the number of layers, the number of neurons, the numbers of inputs and outputs, type of activation functions, feedback information, etc. Combinations of these variables give

designers a variety of options to implement the neural network for thousands of different applications.

This study was interested in comparing the basic power of the neural network in forecasting applications by using the simplest network structures. Therefore, only two structures were used: a single-layer, feedforward neural network and a variable-order, feedforward, functional link neural network. Given that the simplest structures were used, even better results than those reported in this paper could be possible with more advanced neural network structures.

Single-layer, feedforward network

The first network used in the study was a simple, one neuron design. The neuron was given eight inputs of past data ($X_n, X_{n-1}, \dots, X_{n-7}$). The structure of the network is given in Figure 4.

Functional link network

Single-layer neural networks are easier to train than complex, multi-layer networks. However, recall that single-layer networks can only solve linearly separable problems. To avoid working with complex, multi-layer networks and their training algorithms, Pao (1989) developed the functional link network that still allowed him to introduce non-linear separation abilities. The functional link works by introducing non-linear input terms (eg t^2, t^3) into the network. The network can be reduced to a single layer, which also increases the speed

Figure 4: Single-layer feedforward network

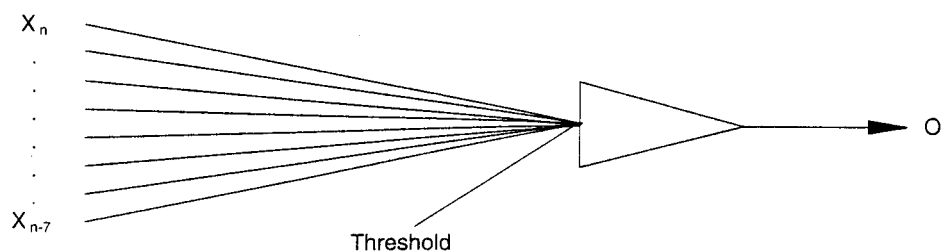
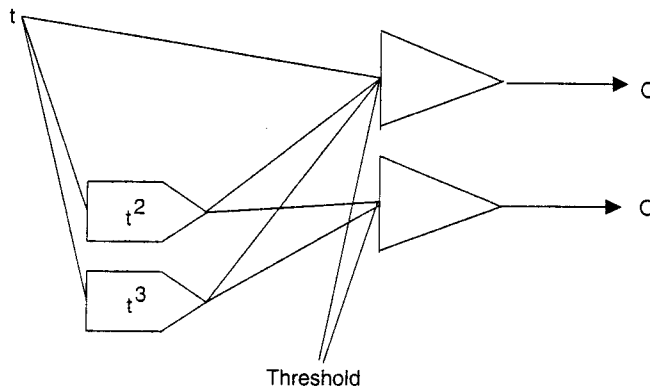


Figure 5: Functional link network



and ease of training. Figure 5 shows the network structure.

RESULTS

The results for the following sections have been divided into two parts, 'training set' and 'holdout sample'. Recall, the training set is the first 30 data points, and the holdout sample is the last 55 data points. The accuracy measurements of each forecasting method will show an error measurement for both the training set and holdout sample.

The study's results will be presented via a graphical representation of the forecasting

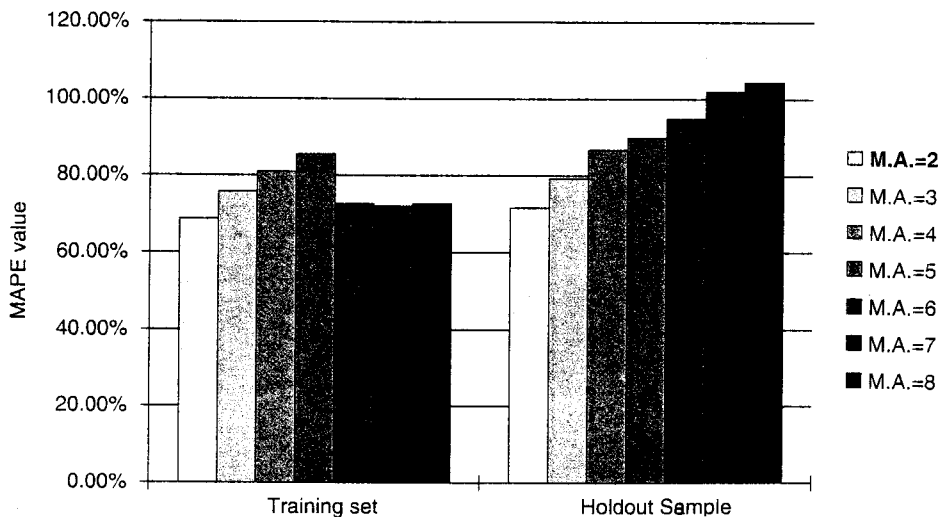
errors. In the graphical representation, the forecasting method that provided the best forecast will be in bold in the legend.

Short-term forecast (one week ahead)

Moving averages

Figure 6 provides a graphical representation of the moving average forecast results. The results labelled 'training set' are the predicted forecasting errors of the first 30 data points or the known, first 30 weeks of data. The 'holdout sample' is the actual

Figure 6: Predicted vs actual forecasting error for moving averages



errors from the last 55 data points or the unknown, last 55 weeks of data.

The two-period moving average provided the best forecast of the seven averages tested. The predicted MAPE based on the training set value for a two-period moving average was 68.82 per cent, and actual MAPE over the hold-out sample was 71.55 per cent.

Another moving average forecast that was performed was the eight-period, weighted moving average. The one-week ahead weighted moving average forecast yielded a 63.8 per cent error over the training set and a 73.1 per cent error over the holdout sample, making it the best in the training set, but worse than the two-period moving average in the holdout sample.

Exponential smoothing

Figure 7 illustrates the exponential smoothing results with and without a trend component. Once an optimum alpha and beta were found by minimising MAPE, those values were used for the entire holdout sample predictions. The values were not updated for each forecasting period.

The best exponential smoothing forecast

yielded a predicted error of 71 per cent and an actual error of 67.8 per cent.

Regression

Of the three regression techniques used, the cubic regression provided the best forecast. Figure 8 illustrates this. The predicted MAPE of the cubic regression was 52.1 per cent, and the actual MAPE was 84.7 per cent. In a similar manner to the exponential smoothing analysis, the linear, quadratic and cubic coefficients of the regression analyses were determined based on the training set and held constant over the holdout sample.

Neural networks

Figure 9 shows the forecasting results of both network structures. The single-layer, multiple-input neural network yielded a predicted MAPE of 61.2 per cent and an actual MAPE of 60.96 per cent.

Overall comparison

The following results take the best of each forecasting technique and presents them in a manner to compare which method yielded the best overall results.

Figure 7: Predicted vs actual forecasting error for exponential smoothing

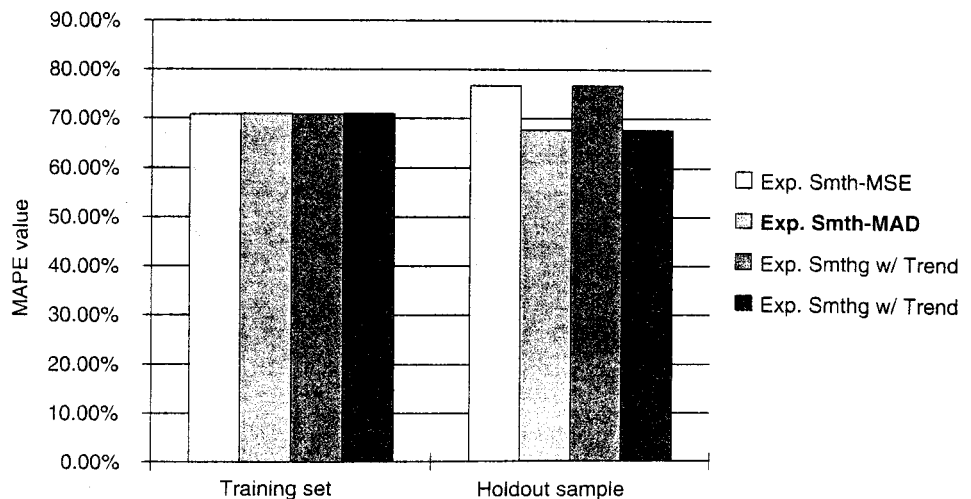
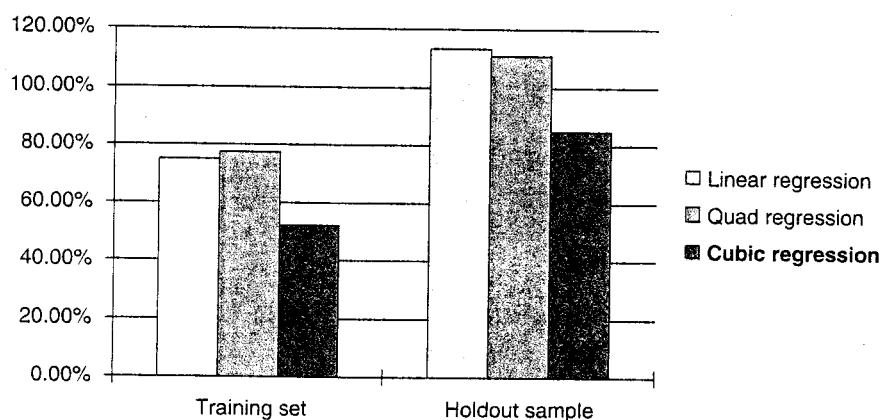


Figure 8: Predicted vs actual forecasting error for regression



Of all four general forecasting methods (moving average, exponential smoothing, regression and neural networks), the single-layer multiple-input neural network provided the best one-week-ahead forecast. Figure 10 illustrates the results. The predicted MAPE was 61.2 per cent, and the actual MAPE was 60.96 per cent.

Long-term forecast (three weeks ahead)

To test the accuracy of the methods beyond a one-week ahead forecast, a three-week ahead forecast was used.

Moving averages

Figure 11 provides a summation of the moving average forecasting results. The eight-period moving average was best of all moving average periods with a 84.04 per cent predicted MAPE and 109.8 per cent actual value over the holdout sample. An eight-period *weighted* moving average forecast was also tested. The results showed a MAPE of 87.5 per cent during the training set and a MAPE of 93.7 per cent over the holdout sample. The results were among the worst in the training set and the best in the holdout sample.

Figure 9: Predicted vs actual forecasting error for neural networks

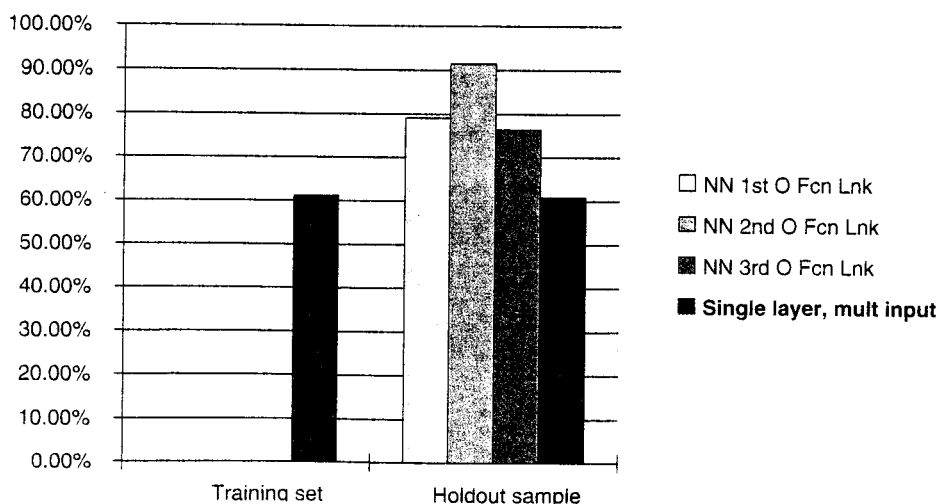
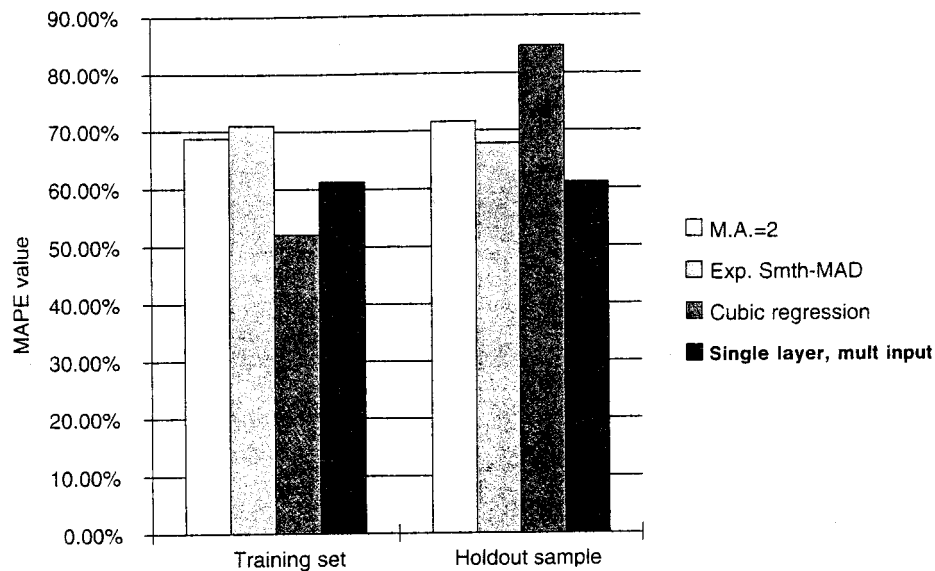


Figure 10: Predicted vs actual forecasting error for best of all methods



Exponential smoothing

The three-week-ahead exponential smoothing results, shown in Figure 12, show that an exponential smoothing technique with a trend component (that seeks to minimise the MAD) provides the best forecast in the holdout sample. The forecast error in the training set was 86 per cent, while the forecast error in the holdout sample was 103 per cent.

Regression

Figure 13 summarises the results of the regression techniques for a three-week-ahead forecast. The cubic regression yielded the best forecast in the holdout sample with a MAPE of 84.5 per cent, and 51.6 per cent in the training set.

Neural networks

Of the two neural network structures

Figure 11: Predicted vs actual forecasting error for moving averages (three weeks ahead)

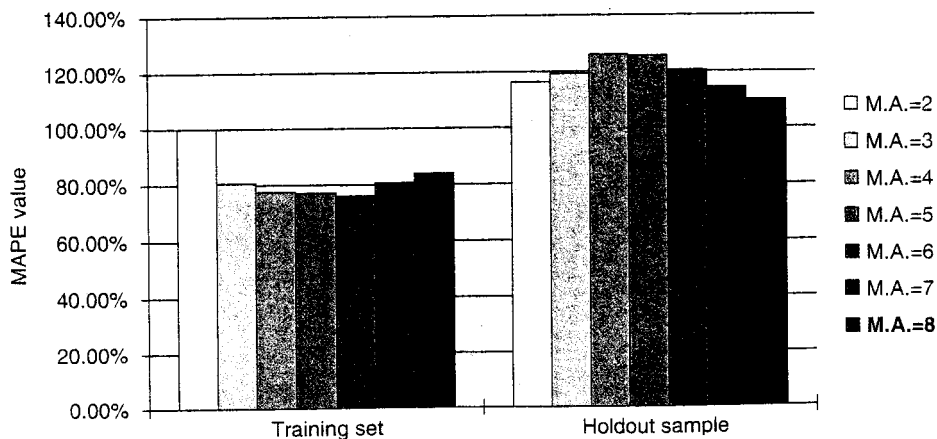


Figure 12: Predicted vs actual forecasting error for exponential smoothing (three weeks ahead)

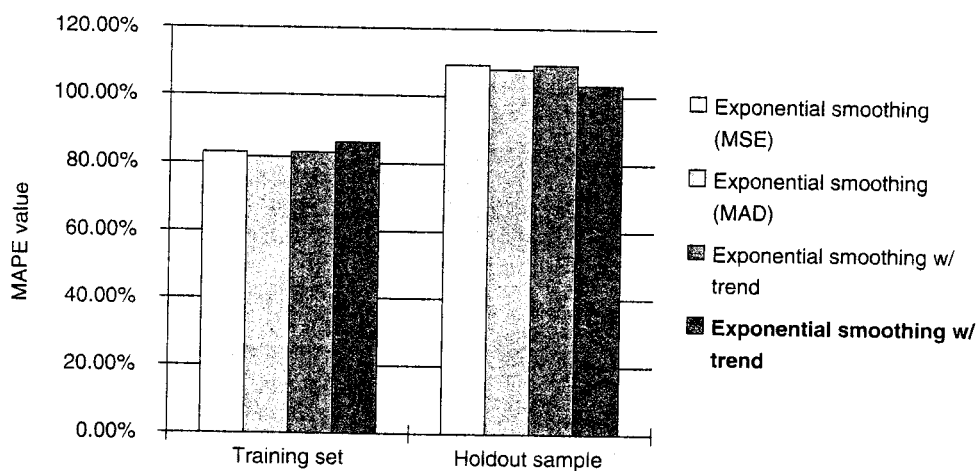


Figure 13: Predicted vs actual forecasting error for regression (three weeks ahead)

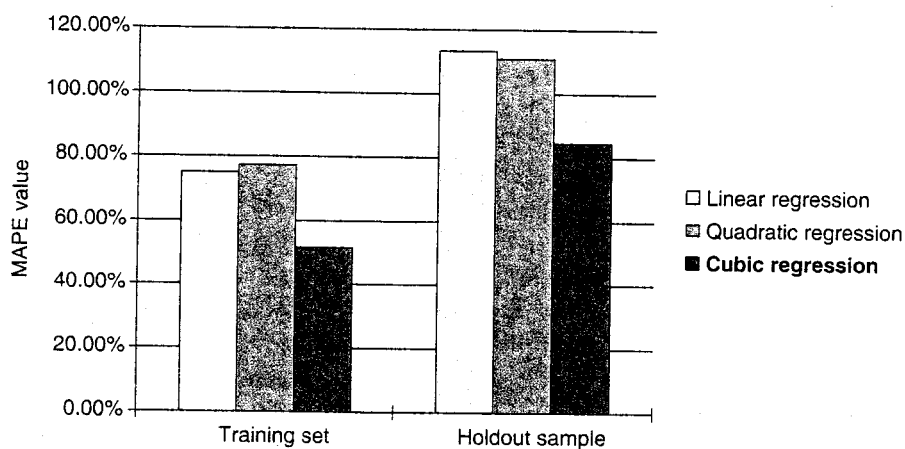


Figure 14: Predicted vs actual forecasting errors for neural networks (three weeks ahead)

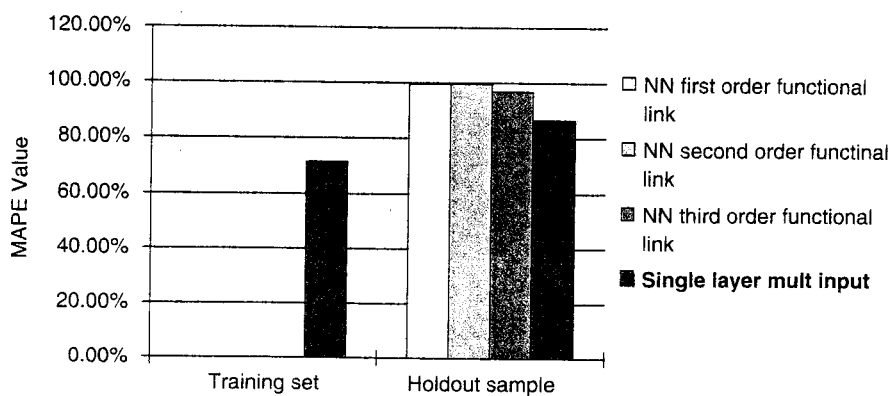
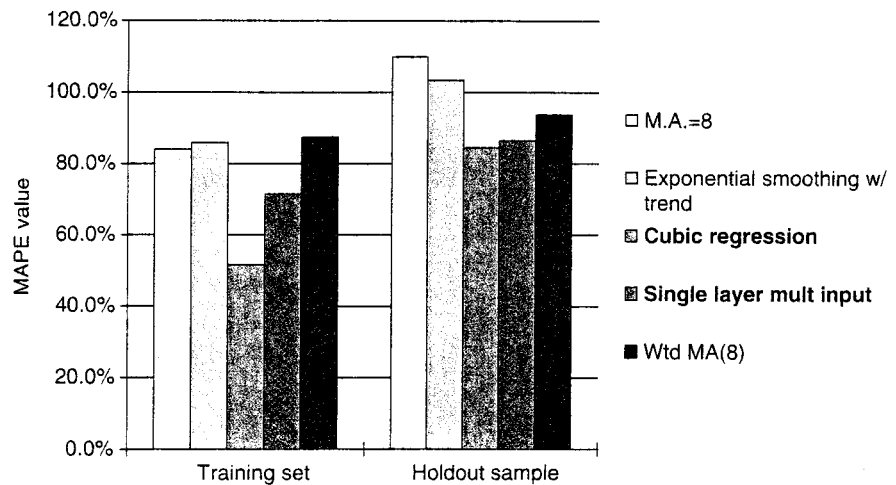


Figure 15: Predicted vs actual forecasting error for best of all methods (three weeks ahead)



tested, the single-layer, multiple-input network provided the best three-week-ahead forecast. As shown in Figure 14, the single-layer, multiple-input structure had a training set MAPE of 71.4 per cent and a holdout sample MAPE of 86.6 per cent.

Overall comparison

A comparison of the best of all methods, as shown in Figure 15, illustrates that the cubic regression and single-layer, multiple-input neural network structure provided the best three-week-ahead forecasts (84.5 per cent MAPE and 86.6 per cent MAPE, respectively), and they could be chosen as the most robust forecasts. The neural network showed less deterioration in going from the training set to the holdout sample (71.4 per cent \rightarrow 86.6 per cent) compared with cubic regression (51.6 per cent \rightarrow 84.5 per cent).

CONCLUSIONS

The results of the forecasting comparisons illustrate how the most basic neural network (a single neuron) can outperform the traditional forecasting methods of moving averages, exponential smoothing and regression. Based on MAPE error value,

the single-layer, multiple-input neural network provided the best short-term and the second-best long-term forecast.

The single-layer, multiple-input neural network's success was probably due to its superior ability to combine data from the past eight periods (exponential smoothing, moving average (8) and regression also used data from the last eight periods, but not as well). The neural network was able to provide consistent results for either a one-week-ahead forecast or a three-week-ahead forecast.

A comparison of the predicted MAPE error values of the training set and the MAPE error values of the holdout sample shows that a single-layer, multiple-input neural network had the most forecasting consistency over all other forecasting methods. If one forecast method was to be selected over all others regardless of forecast period, the single-layer, multiple-input neural network would be the best choice.

The neural network has proven itself to be a formidable substitute in many traditional problematic situations ranging from medicine to forecasting (Hill *et al.*, 1996). The general challenge with neural networks is finding the proper structure, the

correct number of variables, the right data transformation to use, etc.

This paper is only the beginning. A multivariate approach can be initiated to predict the total number of passengers booked in a certain fare class on a specific flight (ie a five-week-ahead forecast of the number of passengers booked in First Class on an early morning flight from Denver to Los Angeles). If additional signatures were provided such as competitor information, economic indicators and an array of other information, the neural network structure could be designed to provide a more accurate and comprehensive forecasting system. The neural network structure is able to perform complex forecasts much more easily than any commonly used forecasting technique.

Given the good performance of relatively simple neural networks that has been demonstrated in this paper, one has to wonder why neural networks have not made greater progress into the airline industry as a forecasting tool? To the authors' knowledge, only one major US airline has adopted such a forecasting technique. Certainly, part of the reason for the reluctance could be the increased data storage requirements and increased computational resources required. Perhaps airlines consider it too great an investment/change in exchange for too small an improvement in forecasting accuracy.

REFERENCES

- Anderson, T. and Wilamowski, B. M. (1995) 'A Modified Regression Algorithm for Fast One Layer Neural Network Training', World Congress on Neural Networks, Washington, DC.
- Gentry, T. W. and Weatherford, L. R. (1995) 'Neural network forecasting: A comparative analysis', in Corriveau, Y. (ed) *Proceedings of the AGIFORS Yield Management and Reservations Study Group*, Washington, DC: AGIFORS.
- Hill, T., O'Connor, M. and Remus, W. (1996) 'Neural network models for time series forecasts', *Management Science*, **42**, 7, 1082-1092.
- Hruschka, H. (1993) 'Determining market response function by neural network modeling: A comparison to econometric techniques', *European Journal of Operational Research*, **66**, 28.
- Makridakis, S., Wheelwright, S. C. and McGee, V. E. (1983) *Forecasting: Methods and Applications*, John Wiley, New York.
- McCulloch, W. and Pitts, E. (1943) 'A logical calculus of the ideas immanent in nervous activity', *Bulletin of Math Biophysics*, **5**, 115-133.
- Pao, Y. H. (1989) *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, Reading, MA.
- Proctor, R. A. (1992) 'Marketing decision support systems: A role for neural networks', *Marketing Intelligence and Planning*, **10**, 1, 22.
- Weatherford, L. R. and Kimes, S. E. (2003) 'A comparison of forecasting methods for hotel revenue management', *International Journal of Forecasting*, forthcoming.

Journal of

Revenue and Pricing Management

Volume One Number Four

Henry Stewart Publications
Russell House, 28/30 Little Russell Street,
London, WC1A 2HN, UK
Tel: + 44 (0)20 7404 3040; Fax: + 44 (0)20 7404 2081;
Website: www.henrystewart.com

Henry Stewart Publications
Subscription Office
PO Box 10812
Birmingham
AL 35202-0812, USA
Tel: 800 633 4931; Fax: 205 995 1588; E-mail: HSP@ebsco.com

Journal of Revenue and Pricing Management
Volume One, Number Four, January 2003
© Henry Stewart Publications 2003
All Rights Reserved
ISSN 1476-6930

Journal of Revenue and Pricing Management is abstracted in International Abstracts
in Operations Research

No part of this publication may be reproduced or transmitted in any form or by
any means, including photocopy and recording, without the written permission
of the publishers. Such written permission must also be obtained before any part
of this publication is stored in a retrieval system of any nature.

Typeset by Acorn Bookwork, Salisbury, Wiltshire
Printed by Cambrian Printers, Dyfed

Journal of

Revenue and Pricing Management

Volume One *Number Four*



Now available online