
Neural Networks for the Analysis and Forecasting of Advertising and Promotion Impact

Hean-Lee Poh, Jingtao Yao* and Teo Jašić

National University of Singapore, Singapore

ABSTRACT Allocating advertising expenses and forecasting total sales levels are the key issues in retailing, especially when many products are covered and significant cross-effects among products are likely. Various statistical and econometric methods could be applied for such analyses. We explore how well neural networks can be used in analyzing the effects of advertising and promotion on sales in this article. The results reveal that the predictive quality of neural networks depends on the different frequency of data observed, i.e. daily or weekly data models, and the specific learning algorithms used. The study also shows that neural networks are capable of capturing the nonlinear aspects of complex relationships in non-stationary data. By performing sensitivity analysis, neural networks can potentially single out important input variables, thereby making it useful for scenario development and practical use. © 1998 John Wiley & Sons, Ltd.

Keywords: neural networks; sensitivity analysis; sales prediction; advertisement allocation.

INTRODUCTION

This study explores the use of neural networks for marketing decision support in the retail business. Its aim is to capture the complex relationships between marketing factors, such as advertising and promotion strategy, and the total sales levels. The important influence factors could be found after neural network forecasting models are built and sensitivity analysis is then conducted. Faster and better decisions are an important way to assist the businesses to survive in the rapidly changing and competitive business environments. Marketing decision makers are increasingly drawn to computer based Decision Support Systems (DSS) to help

them make informed choices. Standard econometric techniques which deal with such a problem have limited explanatory capabilities because they are based on linear models. Neural network technology has seen many application areas in business especially when the problem domain involves classification, recognition and predictions. With the capabilities of neural networks, hidden trends and relations among data which are previously unseen can be deduced. In other words, this is obtaining information from information (White, 1990). A recent survey research conducted by Wong *et al.* (1995) indicated that at least 127 neural network business application journal papers had been published up to September 1994. Different neural network models have been applied to solving hard real-world problems such as financial forecasting. With their successful application to business problems, intelligent

* Correspondence to: Jingtao Yao, School of Computing, National University of Singapore, Singapore 119260. E-mail: yaojt@comp.nus.edu.sg

decision support systems are seen as a way to overcome ill-structured business strategies problems. Examples of the neural networks used to learn functional relationships from their input variables to predict results can be found in Dutta *et al.* (1994), Hill and Ramus (1994), Poh (1991), Poh and Jašik (1995), White (1990) and Yao *et al.* (1997). One of the key issues concerned in retailing is deciding which products to be advertised and promoted in order to increase the overall store sales. Given the competitive pressure in the retailing sector, the retailers of the future are likely to accept only those promotions which lead to an increase in category and overall store sales. This may be achieved by designing promotional programs which increase category and overall store sales, and by directing promotional efforts towards categories which are more responsive to promotions (Raju, 1992). Cost decisions (i.e. budget allocations), copy decisions (i.e. what message to use), and media decisions (i.e. what media to employ) are three major decision areas for advertising (Aaker and Myers, 1982).

There are two aspects of the given problem: (1) planning and allocating advertising expenses; and (2) forecasting total sales levels across a wide product range where significant cross-effects among products are likely. For a single brand, the optimum level of advertising depends upon its margin and the advertising elasticity of demand. However, for a retailer's product, two additional factors are important as stated by Doyle and Saunders (1990): the cross-effects of the advertised products on other items in the retailer's assortment; and the impact on overall store traffic as the effectiveness of advertising is reflected in the overall store performance rather than in the sales of the promoted product. In addition to forecasting the sales based on advertising, sensitivity analysis could be used to determine which promoted categories have more impact on the sales volume.

The primary focus of this research is to explore the potential and to investigate the efficacy of neural networks as an alternative to statistical methods such as multiple regression to predict the variable of interest, and to conduct sensitivity analysis of their models. Both

methodologies aim to capture the relationships between the targeted result, the sales volume, and marketing strategies used by the retailers, and in particular, the marketing decision related to advertising and promotion strategies. Although there are several defined econometric models developed for capturing the relationship of these factors in marketing analysis, there are no general rules between strategic promotion decisions and performance. Rules that have been developed are often a result of past experiences as well as an intuitive awareness of local market conditions, rather than a consistently developed systematic approach. Given these circumstances, a neural network technology can shed some light on these complex relationships and potentially be used as a complementary tool, offering some help to managers facing important decisions.

This paper begins with a basic concepts of neural networks used in this study. It is followed by a section on data summary and background information. Subsequently, a section which presents the training and forecasting results, a section which discusses sensitivity analysis and a section which gives major findings follow. The conclusion discusses areas for future research.

CONCEPTS OF NEURAL NETWORKS

A neural network is a collection of interconnected simple processing elements, or neurons. Neural networks are potentially useful for studying the complex relationships between inputs and outputs of a system (White, 1990). The data analysis performed by neural networks tolerates a considerable amount of imprecise and incomplete input data due to the distributed mode of information processing. There are two neural network models investigated in this research: backpropagation networks and counterpropagation networks. There are three major steps in the neural network-based forecasting proposed by this research: preprocessing, architecture, and postprocessing. In *preprocessing*, information that could be used as the inputs and outputs of neural networks are collected. These data are first normalized

or scaled in order to reduce the fluctuation and noise. In *architecture*, a variety of neural network models that could be used to capture the relationships between the data of inputs and outputs are built. Different models and configurations using different training, validation and forecasting data sets are used for experiments. The best models are then selected for use in forecasting based on such measures as out-of-sample hit rates. Sensitive analysis is then performed to find the most influential variables fed to the neural network. Finally, in *postprocessing*, different trading strategies are applied to the forecasting results to maximize the capability of the neural network prediction.

Backpropagation Networks

A multilayer feedforward network with an appropriate pattern of weights can be used to model some mapping between sets of input and output variables. Figure 1 shows an example of a simple feedforward network architecture, with one output unit and one hid-

den layer, which can be trained using backpropagation. The shaded nodes in Figure 1 are processing units. The arrows connecting input and hidden units and connecting hidden units and the output unit represent weights.

The backpropagation learning algorithm (Rumelhart *et al.*, 1986; Werbos, 1974) is formulated as a search in the space of the pattern of weights, W , in order to find an optimal configuration, W^* , which minimizes an error or cost function, $E(W)$. The pattern of weights will then determine how the network will respond to any arbitrary input. The error or cost function is

$$E = \frac{1}{2} \sum_i \sum_p (t_{ip} - o_{ip})^2 \quad (1)$$

This function compares an output value o_{ip} to a desired value t_{ip} over the set of p training samples and i output units. The gradient descent method is used to search for the minimum of this error function through iterative updates:

$$W(k+1) = W(k) - \eta \nabla E \quad (2)$$

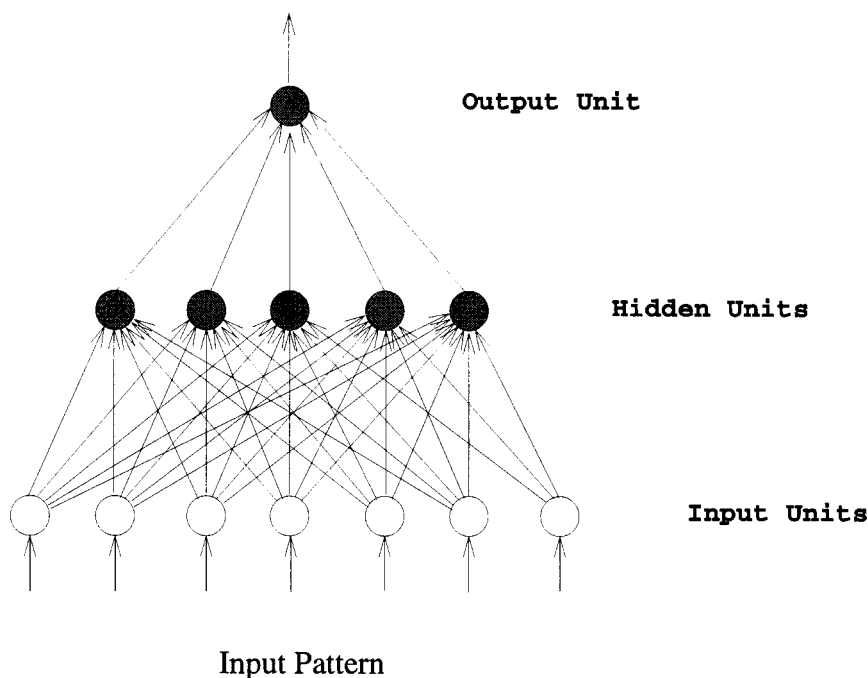


Figure 1 The architecture of a backpropagation network

where η is the learning rate, and ∇E is an estimate of the gradient of E with respect to W .

The algorithm is recursive and consists of two phases: forward-propagation and backward-propagation. In the first phase, the input set of values is presented and propagated forward through the network to compute the output value for each unit. In the second phase, the total-squared error calculated in the first phase is 'backpropagated', layer by layer, from the output units to the input units. During this process, the error signal is calculated recursively for each unit in the network and weight adjustments are determined at each level. The two phases are executed in each iteration of the backpropagation algorithm until the error function converges.

Counterpropagation Network

A counterpropagation network (Hecht-Nielsen, 1987) is a hybrid learning network. It combines a so-called Kohonen layer of unsupervised learning with another layer of supervised learning which uses the basic delta rule. It can speed up learning considerably by training some layers in an unsupervised way (i.e. without a teacher/specified output). This works well in a situation where similar input vectors produce similar outputs. Thus, the aim of Kohonen layer is to categorize the input data into clusters with competitive learning, and then use only the category information for the supervised learning. The classes of similar inputs are defined by a set of prototype vectors. The class of a given input is found by finding the nearest prototype vector using the ordinary (Euclidean) metric. The classes must be formed by the network itself from the correlation of the input data.

The architecture of the counterpropagation network is shown in Figure 2. It consists of three layers of neurons: the input layer where the input pattern is fed, the hidden layer where competitive (Kohonen) learning takes place, and the output layer where supervised delta rule learning occurs. The winning element in the hidden layer sends the value 1 into the third layer. Then the output layer, which is linear, is trained with the usual delta rule

$$\Delta w_{ij} = w_{ij}^{\text{new}} - w_{ij}^{\text{old}} = \eta(t_i - o_i)V_j \quad (3)$$

where $w_{ij}^{\text{new}}, w_{ij}^{\text{old}}$ are the weights between elements i and j , V_j is the output of the winning element (usually set to 1), η is the learning rate, O_i is the output of the network and t_i is the desired output (target).

The outputs of the other hidden elements are set to zero. It should be noted that an underlying assumption of such a network is that input vectors of similar Euclidean distances will give similar outputs. This condition may cause problems in situations where the output of the data is sensitive to small changes in the input vector space. Such situations are common in real-world business and forecasting problems (e.g. time-related patterns of financial data series), making the suitability of such networks dependent on the type of data used. An appropriate data analysis is desirable before using this type of network.

VARIABLE SELECTION AND DATA SUMMARY

Data Sample

The data used in this study are collected from a retailer in Singapore who has a chain of six department stores at various locations. The company is centrally operated with inclination toward a top-down product budgeting approach. Most of the promotion and advertising decisions are made by the headquarters manager. The advertising and promotion campaign data cover various product categories, different media, expenditure levels and sales data are collected for the period between April 1989 and March 1991. A campaign here can be defined as the advertising expenditure aimed at either a specific merchandise or a general sales event. It is possible for campaigns to be overlapping. The data for competitive positions, i.e. market share, prices and costs relative to competition, advertising and promotion strategy of competitors, are not available. This set of data should capture important linkages between marketing strategies and affect sales over the time and hence the overall profitability

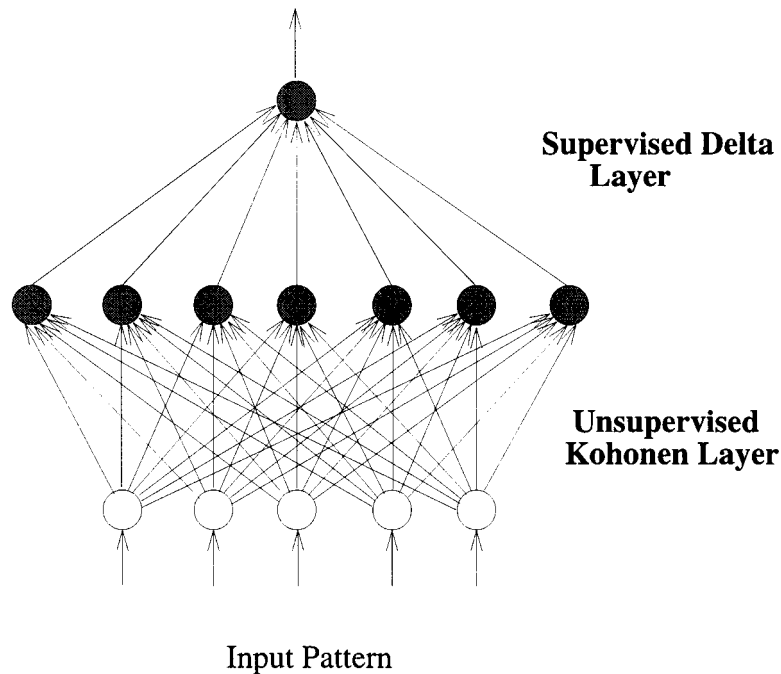


Figure 2 Structure of a counterpropagation network

and the growth of organization, sales promotion, special events, marketing services and visual merchandising. There are four sections in the advertising and promotion department of the retailer organization under study. The sales promotion and special events section make the core decision while the other two play support roles. The manager's decision in advertising or promotion depends on the records of past advertising and promotion campaigns, the latest trend, competitors' activities, budgets, inventory data, sales figures, profit loss statement and other raw data. To establish budgets, a simple historical method, adjusted from previous-year budgets, is used.

The original form of data is the daily model, and the weekly model is easily derived from it. The daily data model takes into account a very short lag effect of advertising with observation of an immediate sales response. Sales results are taken one day after the associated advertisement placement. The weekly data model is expected to aggregate this effect at

the macro level or seasonal effects, resulting in fewer training sets.

The graphs of daily sales and weekly sales and advertising expenditures are shown in Figures 3 and 4, respectively. In Figure 3, the graph of the observed daily sales over the specified period is a series of discrete peaks, which represent sales during the weekends (Saturday and Sunday). However, in some periods, the graph allows the capture of continuous trends, while giving rise to chaotic regimes in others due to seasonal effects. Figure 4 shows data sets organized on a weekly basis. As this pattern represents an aggregation of daily values, the seasonal effects are now more obvious, although the trend and the cycle cannot be detected *a priori*. It is interesting to observe the dynamics of input variables, such as advertising expenditure levels versus sales levels. At times it constitutes a uniform policy, while at other times, it is a pulsing policy. The high peak in July 1989 is due to a massive promotion campaign for a special anniversary of the retail

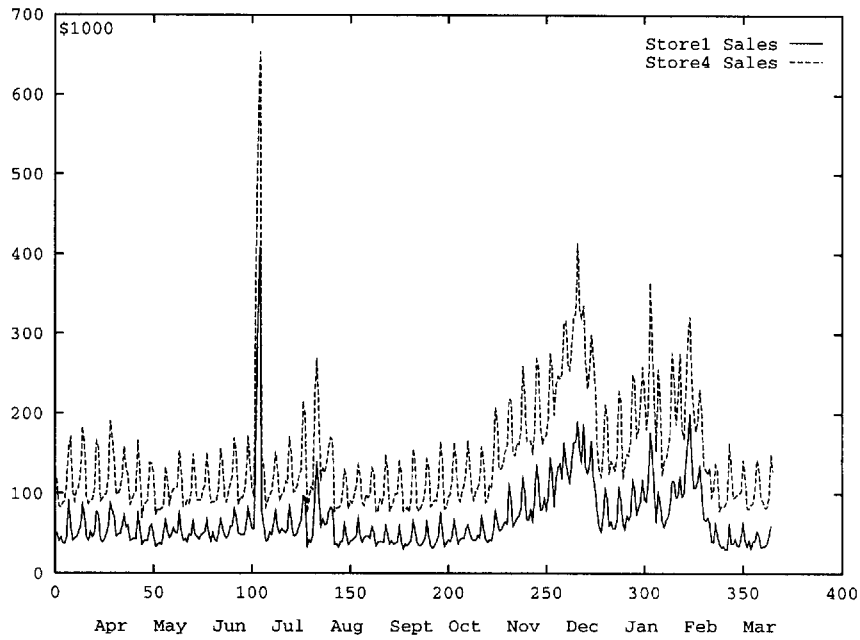


Figure 3 Sales data of stores 1 and 4 on a daily basis for the period: April 1989 to March 1990 (The units are in 1000 Singapore dollars, the solid line refers to sales in store 1 while the dotted line refers to store 4)

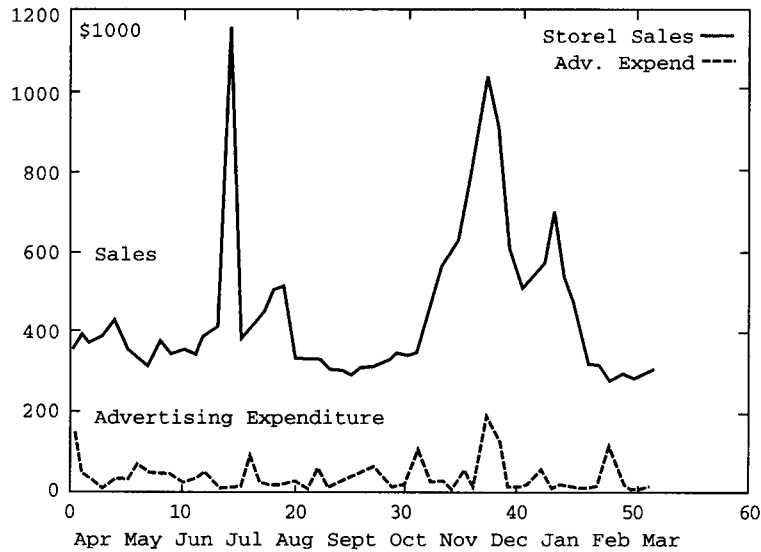


Figure 4 Sales data of store 1 and advertising expenditure on a weekly basis for the period: April 1989 to March 1990 (units are in 1000 Singapore dollars)

organization combined with the regular annual summer sale.

Classification of Variables

There are 22 input variables extracted for use in this study. They follow the existing retailer's classification of advertising and promotion campaign practices. These input variables are roughly divided into three major groups: (1) variables related to the categories of products which have been advertised (Var_1 to Var_{13}); (2) variables related to different media being used for advertising and promotion (Var_{14} to Var_{20}); and (3) variables related to the cost of advertising and promotion activities with respect to the group of stores to which the budget is allocated (Var_{21} , Var_{22}). The classification of input and output variables in the daily data model together with their mean and standard deviation is given in Table 1.

The effect of advertising and promotion is observed in an aggregate model comprising a different categories of products, and the observed sales volume is the sum of sales of all items. This focus is relevant from a retailer's perspective because the retailer's revenues are more closely linked to the overall category and store sales than to the sales of any particular brand or product category. For the first group of variables (categories of products), there can be various advertisements related to different categories of items appearing in different media on a daily basis. The variable for the product category therefore assumes the value which is equal to the number of advertisements on each day.

Variables var_8 (Gift Ideas), var_{10} (Storewide Advertising and Promotion Campaign), var_{11} (Stock Clearance Announcement), and var_{13} (Fashion Shows) are not confined to any specific category of products and are characterized as special events. These events can be grouped together as they do not coincide at the same time and represent promotional sales of different ranges of products across the stores. The incorporation of these events is important because they can have a strong influence on the 'normal' pattern of sales.

For the second group (media), the value is

the number of advertisements appearing in that medium. As a few different advertisements can be present in the same medium on the same day, the value of the variables in this group is sometimes greater than one. For the weekly data model, these values would be aggregated.

Variable var_{14} (Newspaper 1) is the English newspaper with the largest circulation, while Newspaper 2 (var_{15}) is the local language newspaper with major circulation. The other newspaper media, Newspaper 3 (var_{16}) and Newspaper 4 (var_{17}), do not feature advertisements of the retailer so often.

STATISTICAL MODELS FOR IMPACT OF ADVERTISING AND PROMOTION

Two multiple linear regression models are built to evaluate the impact of advertising and promotion. The models can be formulated as follows:

$$SaleVolume = \beta_0 + \sum_{i=1}^n \beta_i var_i + \mu \quad (4)$$

where $n=22$ for daily model and $n=19$ for weekly model; $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ are regression coefficients; $SaleVolume$ is the Sales Volume to be forecast; and μ is the error term that cannot be explained by independent variables.

The number of input units in the weekly model is reduced to 18 by performing regression analysis and excluding variables which are found not significant, such as the variable var_{11} (Stock Clearance Announcement). Furthermore, the variable var_2 (Weekend Indicator) is excluded because in the weekly model only seasonal effects at the macro level are present. Two variables describing newspaper media in the daily model (var_{16} /Newspaper 3 and var_{17} /Newspaper 4) can be combined together. Finally, two variables related to the cost of advertising (var_{21} and var_{22}) in the daily model are also combined into one. Therefore, the definition of var_i may differ from the daily model and thus a bracketed term is provided after each variable.

The models aim to seek the least mean squared error between the actual and the forecast sales volumes so as to determine the coef-

Table 1. Data summary of training data in the daily data model (var_2 is a binary code to indicate the weekend, var_{18} is the number of video clips or spots televised, var_{20} to var_{23} are in thousand Singapore dollars and others are number of times)

Variable	Meaning	Mean	Std. Dev.	Max. value
var_1	Man Fashion	0.05	0.2114	1.00
var_2	Weekend Indicator	—	—	—
var_3	Ladies' Fashion	0.1085	0.3123	1.00
var_4	Ladies' Accessories	0.0233	0.1513	1.00
var_5	Toys/Stationery	0.0155	0.1240	1.00
var_6	Children's wear Products	0.0155	0.1240	1.00
var_7	Household	0.0233	0.1513	1.00
var_8	Gift Ideas	0.0310	0.1740	1.00
var_9	Cosmetics	0.5814	0.5687	3.00
var_{10}	Storewide Sales Advert.	0.4031	0.5379	2.00
var_{11}	Stock Clearance Announcement	0.01	0.088	1.00
var_{12}	Seasonal Fashion Statements	0.062	0.2421	1.00
var_{13}	Fashion Shows	0.0775	0.2685	1.00
var_{14}	Newspaper 1	1.2016	0.6541	3.00
var_{15}	Newspaper 2	0.4264	0.4965	2.00
var_{16}	Newspaper 3	0.0543	0.2274	1.00
var_{17}	Newspaper 4	0.2868	1.7862	20.00
var_{18}	TV (3 Channels)	2470.931	7665.473	66000.00
var_{19}	Mailers	0.0465	0.2114	1.00
var_{20}	Tourist Editions	0.0620	0.2421	1.00
var_{21}	Cost (Group 1)	8023.465	232823.81	193320.00
var_{22}	Cost (Group 2)	135665.40	25055.51	175000.00
var_{23}	Sales Volume (Store 1)	69.28	50.38	409.00
var_{23}	Sales Volume (Store 2)	57.34	37.18	317.00
var_{23}	Sales Volume (Store 3)	72.78	48.50	385.00
var_{23}	Sales Volume (Store 4)	144.37	80.36	654.00
var_{23}	Sales Volume (Store 5)	129.88	90.40	809.00
var_{23}	Sales Volume (Store 6)	149.57	107.21	967.00

efficient values. We used a t -test to test the significance of each individual coefficient. For the daily model, the variables with significant or nearly significant t -statistics are: var_2 (Weekend Indicator), var_4 (Ladies' Accessories), var_8 (Gift Ideas), var_9 (Cosmetics), var_{12} (Seasonal Fashion Statements), var_{13} (Fashion Shows) and var_{15} (Newspaper 2). For the weekly model, the variables with significant or nearly significant t -statistics are: var_4 (Toys/Stationery), var_7 (Gift Ideas), var_8 (Cosmetics), var_{11} (Fashion Shows), var_{13} (Newspaper 2) and var_{18} (Cost). These variables are generally consistent for both models except for Toys/Stationery and Statements. Pearson's correlation coefficient which is defined as

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{(N-1)S_x S_y} \quad (5)$$

(where S_x and S_y are the standard deviations of x and y respectively) is used to check on the collinearity of data set x and y . The results have shown that there are seven variables that are correlated with dependent variables for the weekly model. They are var_1 (Man Fashion, 0.335), var_4 (Toys/Stationery, 0.204), var_5 (Children's wear Products, 0.296), var_7 (Gift Ideas, 0.368), var_8 (Cosmetics, -0.216), var_{10} (Seasonal Fashion Statements, -0.202) and var_{13} (Newspaper 2, 0.267). More results on statistical models may be found in Jašić (1993) where the impact on seasonal indices is also discussed.

BUILDING THE NEURAL NETWORK MODELS

In this study there are 730 samples for the daily model and 104 for the weekly model in

the period between April 1989 and March 1991. The first 360 daily samples and 52 weekly samples data are used for training, while 92 daily samples (from January 1991 to March 1991) and 26 weekly samples (from October 1990 to March 1991) are used to observe how well the neural network performs. The remaining data can be used for validation purposes to avoid over fitting in backpropagation.

The data size requirement depends on the number of input and output units of the neural network. Data are normalized to values within the interval [0,1] for use by the continuous-valued units in the neural network. The normalizing factor can be taken as the maximum absolute value of each variable in the observed period of 2 years. An important issue here is the treatment of outliers and seasonal effects.

In order to enhance the predictive and explanatory power of the model, further information can be provided for the neural network. This can be achieved by adding auxiliary units to the neural network model. An additional indicator for the long weekend (Saturday and Sunday) is provided in the daily model because the sales pattern shows peaks at weekend.

Backpropagation Results for Daily and Weekly Data Models

The configuration of the neural network is one of the most important issues when using the backpropagation method. There are a number of theoretical observations which show that a backpropagation neural network is an efficient function approximator, given sufficiently many hidden nodes with sigmoid output functions (Funahashi, 1989; Hornik *et al.*, 1989; Lippmann, 1987). However, these theorems do not specify the optimal or minimal number of hidden units needed to reach a desired performance.

With regard to the number of units and weights, Baum and Haussler (1989) suggested that the number of weights for a satisfactory network performance should be less than one tenth of the number of training patterns. An obvious shortcoming of this rule is that it is very restrictive for limited data sets and large networks. For an oversized network (i.e. the number of weights is of the order of the num-

ber of training sets), a cross-validation procedure may be needed due to the large number of parameters provided which can cause the oversized network to fit a noise (Weigend *et al.*, 1990). Alternatively, certain pruning algorithms may be pursued.

In general, some experiments are carried out to find the configuration of the neural network for a particular problem. For the number of hidden-layer units, a rule of thumb is used that defines the number of hidden units as half of the sum of the input and output units. Thus, the network configuration used in this study for the daily model are 22-10-1 (i.e. 22 input units, one hidden layer with 10 units, and 1 output unit), and 18-9-1 for the weekly data model.

For a network with two hidden layers, there are no specific rules of thumb with respect to the number of units in the first and second layer or their ratio. In this study, the 22-15-5-1 and 22-10-3-1 configurations are chosen for the daily model, and 18-9-3-1 for the weekly model. The training and forecasting errors are measured in terms of a goodness of fit R^2 (where R^2 is calculated as the average forecasting error divided by the variance of the forecast data). The results are shown in Table 2.

Figure 5 shows the prediction performance of a 22-10-1 network, using a data set corresponding to a time period of 3 months. The network is trained with a learning rate of 0.05 and 3000 iterations. The actual output (solid line) is compared with the neural network output (dotted line).

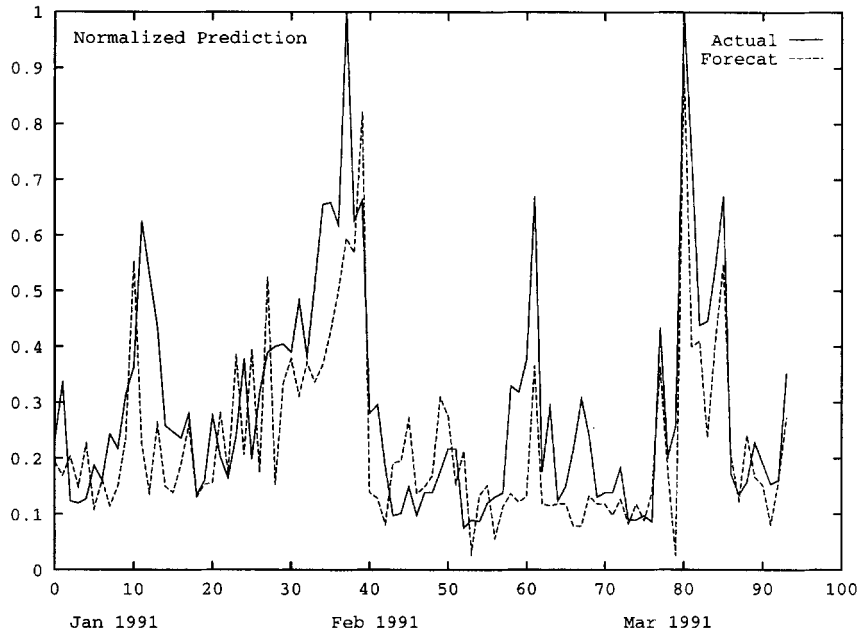
For the 18-9-1 network which is chosen for the weekly data, the choice of a cut-off point of training is not straightforward. In contrast to the daily data, the forecasting error for the weekly data is not greater than two times the training error for a relatively large number of training cycles (e.g. 1400 cycles for a 18-9-1 network). Table 3 shows the best generalization results for the 18-9-1 network which is trained up to 10,000 iterations at a learning rate of 0.05, using the data from different stores.

Counterpropagation Results for Daily and Weekly Data Models

For the counterpropagation network, the suitability of its use depends on the relationships

Table 2. The best forecasting results of a network (22-10-1) using daily data for different stores

Stores	1	2	3	4	5	6
Testing error	0.018188	0.015776	0.010636	0.016885	0.013949	0.013312
Training error	0.010754	0.004215	0.013811	0.004370	0.004502	0.004400
Iteration	200	575	150	350	500	475

**Figure 5** Three months' daily prediction for store 1 using a 22-10-1 network**Table 3.** The best forecasting results of a network (18-9-1) using weekly data for different stores

Stores	1	2	3	4	5	6
Testing error	0.009827	0.011415	0.007105	0.009750	0.012436	0.010038
Training error	0.008958	0.006300	0.009849	0.005192	0.005790	0.006409
Iteration	425	500	350	500	575	550

between the inputs and outputs. If the points in the input space that are close in Euclidean distance have widely varying outputs, the network will not perform well. In that case, a network may require a large number of Kohonen nodes in order to adequately represent the problem. On the other hand, if the input data are from a homogeneous source,

then the network will perform well, as the initial clustering in the Kohonen layer will be easier.

For both the daily and weekly data models, the counterpropagation network is trained with 100, 500, and 1000 iterations using the following parameters: 1 to 60 nodes in the Kohonen layer, the learning rate of the Kohonen layer is set to

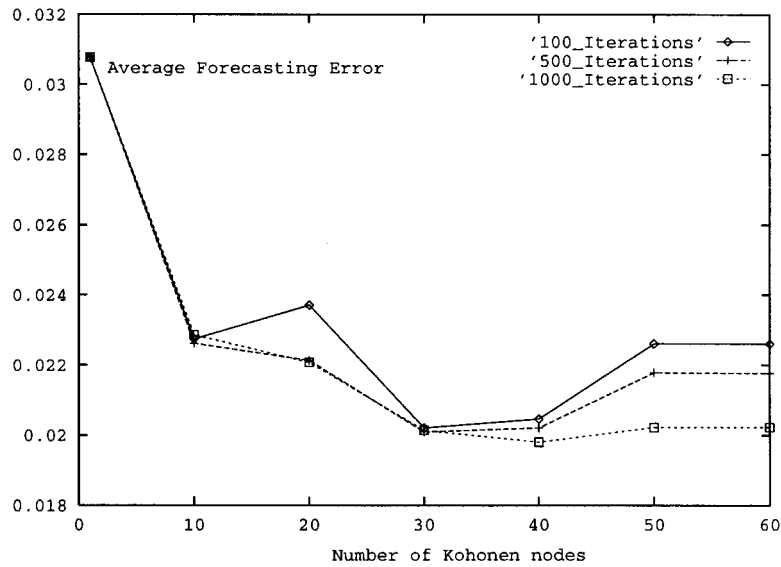


Figure 6 Forecasting error for the different counterpropagation daily models (data of store 1)

0.05, and the learning rate of the delta layer is set to 0.1. Figures 6 and 7 present the forecasting results in terms of the average error.

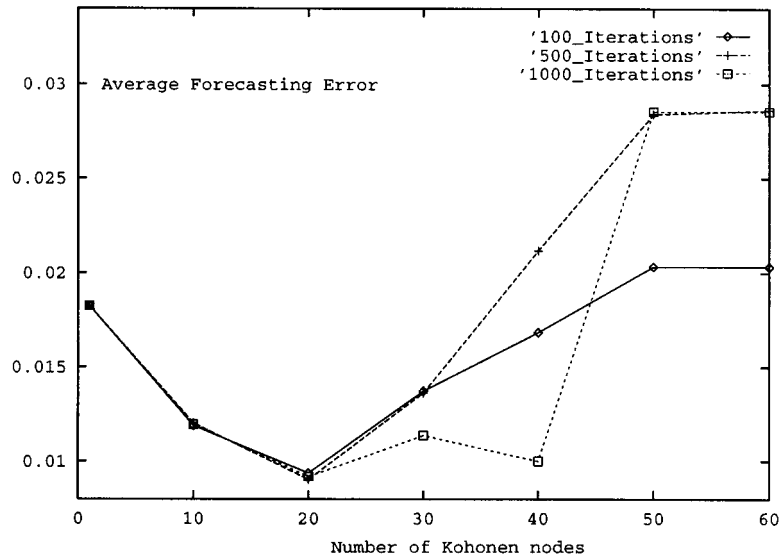


Figure 7 Forecasting error for the different counterpropagation daily models (data of store 1)

Results for the Daily Data Model

Given the results obtained for the counterpropagation network using the daily data sets, the major indications are that there is no significant difference in the forecasting results for a different number of iterations. In addition, the following observations are made.

The best forecasting result is obtained using 1000 iterations and 40 nodes in the Kohonen layer (data of store 1), with a generalization error being 0.019806. When compared with the results obtained using backpropagation, the forecasting results are not any better. This may be attributed to the fact that changing a single bit in the weekend indicator (var_2), causes the output to change significantly, and this violates the assumption for using counterpropagation.

Results for the Weekly Data Model

For the counterpropagation network using the weekly data sets, the forecasting results for data of store 1 are shown in Figure 7. As in the daily data model, these results are not any better than the forecasting results obtained using backpropagation and there is no significant difference in the forecasting results obtained using different number of iterations. In addition, the following points should be emphasized.

The best result is obtained using 20 nodes in the Kohonen layer for data of store 1 after 500 iterations, with the generalization error being 0.009042. Increasing the number of the Kohonen nodes over 30 will generate very poor generalization results. This is contrary to the expected behavior as more Kohonen nodes should give better performance. The reason may be that some information is filtered out in the Kohonen layer due to the limited number of forecasting data and training data, while the properties of forecasting data are different from those in the training data. A counterpropagation network is more suitable for the weekly data model because the fluctuations of the dependent variables at the micro level (daily data model) are not reflected in the weekly data model.

Figure 8 shows the prediction performance of a counterpropagation network, using the weekly forecasting set corresponding to a time

period of 6 months. The network is trained with 100 iterations, using 20 nodes in the Kohonen layer. Compared with the actual target (solid line), it can be seen that the neural network output (dotted line) closely follows the actual target values.

SENSITIVITY ANALYSIS

A good and understandable data model can be provided by reducing the number of variables through a weight sensitivity analysis. The impact of each independent variables on the dependent variables can also be found from the sensitivity analysis. The analysis of weights can be accomplished using the following three methods.

The first method is the *Equation Method*. For a feedforward neural network with one hidden layer and one output unit, the influence of each input variable on the output can be calculated from the equation:

$$\sum_k O(1 - O)w_{k1}^2 v_k^2 (1 - v_k^2) w_{ik}^1 \forall_i \quad (6)$$

where O = the value of the output node;*

w_{k1}^2 = the outgoing weight of the k th node in the hidden (second) layer

v_k^2 = the output value of the k th node in the hidden (second) layer

w_{ik}^1 = the connection weights between the i th node of the input (first) layer and the k th node in the hidden layer.

Using the Equation Method, there will be n readings for n input variables for each input row into the network. If there are r input rows, there will be r readings for each of the n input variables. All the r readings for each input variable are subsequently plotted to obtain its mean influence, I_n , on the output. These I_n s indicate the relative influences each input variable could have on the output variable: the greater the value, the higher the influence.

The second method is the *Weight Magnitude Analysis Method*. The connecting weights

*The notation w_{bc}^a here can be read as the weight from the b th node in the a th layer to the c th node in the next layer.

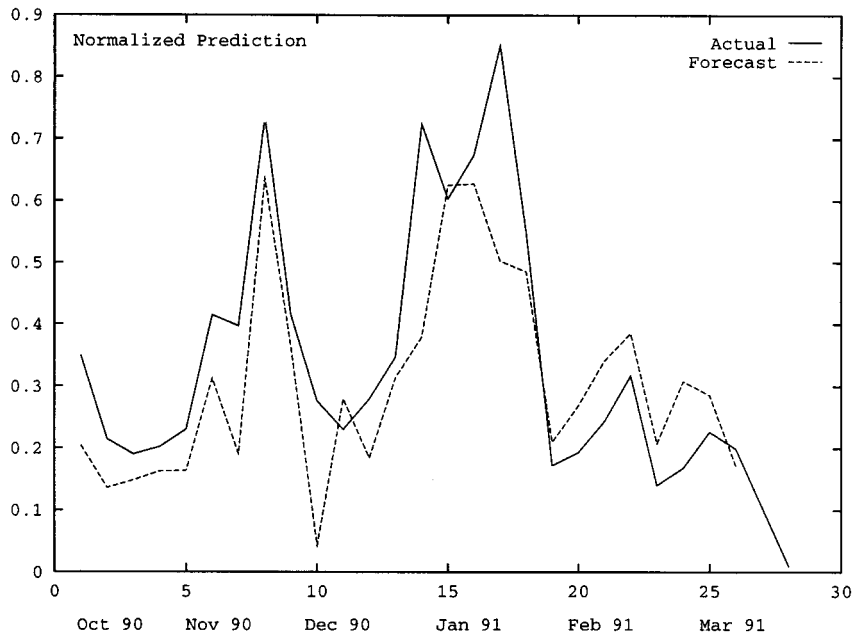


Figure 8 Six months' weekly prediction for store 1 using a counterpropagation network

between the input and the hidden nodes are observed. The rationale for this method is that variables with higher connecting weights between the input and output nodes will have a greater influence on the output node results. For each input node, the sum of its output weight magnitudes from each of the hidden layer nodes is the relative influence of that hidden node on the output. It is done for all input nodes. To find the sum of weight magnitudes from each input node, weight magnitudes of each of the input nodes are first divided by the largest connecting weight magnitude between the input and the hidden layer. This is called normalization. The normalization process is a necessary step whereby the weights are adjusted in terms of the largest weight magnitude. The weight magnitudes from each input node to the nodes in the hidden layer are subsequently summed and ranked in a descending order. The rank is an indication of the influence an input node has on the output node relative to the rest. The rank formula is as follows:

$$I_i = \sum_k \frac{w_{ik}^0}{\max_{\text{All } i,k}(w_{ik}^0)} \quad (7)$$

where the notation is as same as in equation (6).

The third method is the *Variable Perturbation Method* and tests the influence of the inputs on the output. The method adjusts the input values of one variable while keeping all the other variables untouched in this approach. These changes could take the form of $I_n \Rightarrow I_n + \delta$ or $I_n \Rightarrow I_n * \delta$, where I is the input variable to be adjusted and δ is the change introduced into I . The corresponding responses of the output against each change in the input variable are noted. The node whose changes affect the output most is the one considered most influential relative to the rest.

Sensitivity analysis of a suitably trained neural network can determine a set of input variables which have greater influence on the output variable. In the experiment, sensitivity analysis for different variables in the weekly model is performed. These variables are perturbed for different ranges of their values in

the forecasting set in order to obtain the corresponding changes in the output variable, i.e. var_{19} (Sales Volume). The changes of the output variable can then be observed for the direction of change and the degree of reaction, and then compared to changes obtained by the linear regression model. It should be noted that this is a weak sensitivity measure because of the possibility of nonlinear interactions. Changes to two or more inputs in pair can have a different effect from that of change to one input alone. However, sensitivity analysis can also be done for a combination of input variables to examine the interactions among the variables.

For this experiment, the 18-9-1 network is first trained with weekly data sets of store 1 using backpropagation, with a learning rate set to 0.05. The network with 900 training cycles is chosen.

Sensitivity analysis of variables var_8 (Number of Advertisements for Cosmetics Items), var_{12} (Newspaper 1), var_{13} (Newspaper 2), and var_{18} (Cost of Advertising) in the weekly model is conducted in our first experiment. The results of var_{13} are shown in Table 4. The values of var_{13} are discrete, and they are either 0.00, 0.25, 0.50, 0.75 or 1.00. These values are then perturbed by 0.25 or -0.25 except for the extreme values (0 and 1) which are perturbed only in one direction. In the linear model, the coefficient for the variable var_{13} is 0.172368 with a standard error 0.104700 (the *t*-test value is 1.646). Therefore, the sales volume change is 0.043092 if the change of the variable var_{13} is 0.25.

The results show that the neural network reacts similarly to the linear model with respect to the direction of change and the degree of reaction. In addition, the change of sales volume is sensitive to different values or ranges of the perturbed input variable. This is the

nonlinear characteristic of the problem which is captured by the neural network.

The results of var_8 are shown in Table 5. From here we can find that var_8 is negatively correlated to var_{19} (Sales Volume) and has a significant negative impact on dynamics of total sales. A sensitivity analysis of var_{12} finds that it also has a less significant impact on sales.

In the second experiment, two pairs of variables (var_8 and var_{12} , and var_8 and var_{13}) in the weekly model are perturbed simultaneously to observe the sales volume change. In other words, the interaction among these variables is tested. For both pairs of variables, the direction of change is the same as in the case of a sensitivity analysis of variable var_8 , which dominates the change in the output variable.

FINDINGS AND DISCUSSION

Technical Observation

In this study, the forecasting problem is limited to an explicative approach because the autoregressive method is not observed. However, the primary concern is the use of a neural network to capture the mapping between the dynamics of input variables and the fluctuations of the output variable. Such a model is able to yield a good out-of-sample performance given the short-time prediction horizon.

The backpropagation network is suitable for prediction in both the daily and the weekly data models. Training for the daily model is a more difficult task but the prediction performance in a short-time period is acceptable. A counterpropagation network is not suitable for the daily model as explained above but for a weekly model it gives good predictions (see Figure 8). This may point to the existence of

Table 4 Sensitivity analysis of var_{13} at values 0.00, 0.25, and 0.50 for a perturbation value of ± 0.25

var_{13}	0.00		0.25		0.50	
Δ Input	+0.25	-0.25	+0.25	-0.25	+0.25	-0.25
Δ Sales	+0.0425	N.A.	0.0464	-0.0436	0.0485	-0.045

Table 5 Sensitivity analysis of var_g at values 0.00, 0.25, and 0.50 for a perturbation value of ± 0.25

var_g	0.00		0.25		0.50	
Δ Input	+0.25	-0.25	+0.25	-0.25	+0.25	-0.25
Δ Sales	-0.0664	N.A.	-0.0577	+0.0628	-0.0474	+0.0552

higher degrees of correlations of input variables in the weekly model which is captured by the counterpropagation network. However, this type of neural network cannot be used for sensitivity analysis as opposed to the backpropagation network.

Another point which should be noted is that the validity of neural network prediction performance assumes that the dynamics of input variables is similar to that of the recent past, especially if insufficient data sets are available for training.

Marketing Implications

By conducting a sensitivity analysis on a suitably trained neural network model, the following findings of this study can be translated into practical applications in marketing management. First, the results indicate that continuous and heavy promotion of certain items (e.g. cosmetics) is either not correlated or may have negative effects on the overall sales. Second, special events (shows, statements) and the intermittent advertising and promotion of some product categories are positively correlated and have a significant positive impact on the overall sales. Third, the results indicate that advertising in the leading local language (Chinese) newspaper has a more significant effect on the total sales than through other channels. On the other hand, featuring advertisements in the English newspaper that has the largest circulation, which carried the majority of advertisements, is found to have no significant effect on the overall sales.

These results have been verified by the estimation of different regression methods. However, it is the intention of this study to generalize the results. It is quite possible that different advertising retention exists for differ-

ent product categories and different periods of time.

CONCLUSION AND FUTURE RESEARCH

The results from this study indicate that a neural network is able to capture the nonlinear relationships in a causal model even though there is no explicit structures of the domain field. The model developed thus can be used in assisting in the short-term forecasting of a variable of interest. In addition, it can be used to conduct a sensitivity analysis, which is a useful procedure for analyzing possible strategies with respect to the observed variables.

The results indicate that backpropagation is an efficient method for studying the relationship between input and output variables. It could be a useful tool for planning and allocating advertising expenses. However, the forecasting accuracy of the dependent variables depends on different data models and size of data sets. In the case of the daily model, a suitable network architecture is not easy to identify. For the weekly data, the choice of a suitable network architecture is easier and can be determined by looking at the best forecasting result.

Speed is an advantage when using a counterpropagation network as compared to backpropagation. However, for the daily data model, counterpropagation does not yield better results than backpropagation. A possible explanation is that some points in the input space that are close in Euclidean distance have widely varying outputs, causing neural networks to perform poorly. In particular, changing one bit in the input corresponding to a weekend indicator in the daily model induces significant change in the output. This will cause unsatisfactory per-

formance of Kohonen layer, or require a large number of Kohonen nodes, which is impractical. For the weekly model, the counterpropagation network produces better results, but again, not better than the corresponding results of the backpropagation method.

Practitioners may deal with similar problems in retail markets using market findings in this study. A decision support system could be built using a neural network as a black box to assist in human decision making. Combining with statistic models and linking to a variety of data sources, an expert system can also be built.

Finally, further research should lay the ground for the incorporation of the neural network models into a broader framework of a marketing decision support system (MDSS). The neural network model in such a system should be able to deal with different problems (e.g. sales analysis, market share evaluation, assessment of the impact of different marketing strategies etc.) using extrapolative (time series) as well as explicative (causal) analyses, or a combination of both.

Acknowledgments

The authors are grateful to the anonymous referees whose insightful comments enabled us to make significant improvements. We thank Professor Chew Lim Tan for his invaluable advice and helpful comments. Thanks also go to Robyn E. Wilson for her precious comments on the technical writing.

References

- Aaker, D.A. and Myers, J.G., *Advertising Management*, Prentice Hall, Englewood Cliffs, NJ, 1982.
- Baum, E.B. and Haussler, D., 'What size net gives valid generalization?' *Neural Computation*, **1**, 1989, 151.
- Dutta, S., Shekhar, S. and Wong, W.Y., 'Decision support in non-conservative domains: generalisation with neural networks', *Decision Support Systems*, **11**, 1994.
- Doyle, P. and Saunders, J., 'Multiproduct advertising budgeting', *Marketing Science*, **9**, 2, 1990, 97–113.
- Funahashi, K., 'On the approximate realization of continuous mappings by neural networks', *Neural Networks*, **2**, 1989, 183–192.
- Hecht-Nielsen, R., 'Counterpropagation networks', *IEEE Proc. of the International Conference on Neural Networks*, Vol. 2, 19–32, 1987.
- Hill, T. and Ramus, W., 'Neural network models for intelligent support of managerial decision making', *Decision Support Systems*, **11**, 1994.
- Hornik, K., Stinchcombe, M. and White, H., 'Multilayer feedforward networks are universal approximators', *Neural Networks*, **2**, 1989, 359–366.
- Jašić, T., *A Neural Network Approach to Retail Marketing Analysis and its Implications for Decision Support*, Masters thesis, National University of Singapore, 1993.
- Lippmann, R.P., 'An introduction to computing with neural nets', *IEEE ASSP Magazine*, April, 1987, 4–22.
- Poh, H.L., *A Neural Network Approach for Marketing Strategies Research and Decision Support*, PhD thesis, Stanford University, March 1991.
- Poh, H.L. and Jašić, T., 'Forecasting and analysis of marketing data using neural networks: a case of advertising and promotion impact', *Proceedings of the 11th Conference on Artificial Intelligence for Applications*, Los Angeles, 1995, pp. 224–230.
- Raju, J.S., 'The effect of price promotions on variability in product category sales', *Marketing Science*, **11**, 3, 1992, 207–220.
- Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 'Learning internal representations by error propagation', in Rumelhart, D.E. and McClelland, J.L. (eds), *Parallel Distributed Processing (PDP): Explorations in the microstructure of cognition*, Vol. 1, MIT Press, Cambridge, MA, 1986.
- Weigend, A.S., Rumelhart, D.E. and Huberman, B.A., 'Predicting the future: a connectionist approach', *International Journal of Neural Systems*, **1**, 3, 1990, 193–209.
- Werbos, P.J., *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, PhD Thesis, Harvard University, November, 1974.
- White, H., 'Connectionist nonparametric regression: multilayer feedforward neural networks can learn arbitrary mappings', *Neural Networks*, **3**, 1990.
- Wong, B.K, Bodnovich, T.A. and Selvi, Y., 'A bibliography of neural network business applications research: 1988-September 1994', *Expert Systems*, August 1995.
- Yao, J.T., Li, Y.L. and Tan, C.L., 'Forecasting the exchange rates of CHF vs USD using neural networks', *Journal of Computational Intelligence in Finance*, **5**, No. 2, 1997, 7–13.