

---

# A visual method for determining variable importance in an artificial neural network model: An empirical benchmark study

Received (in revised form): 28th June, 2002

## Kelly E. Fish

is Assistant Professor of Decision Sciences at Arkansas State University, Jonesboro, Arkansas, USA. His research interests involve applications of artificial neural networks involved in decision support systems including consumer choice modelling, data mining and database marketing.

## Jeffery G. Blodgett

is Associate Professor of Marketing at The University of Mississippi, Oxford, MS, USA. His research interests include consumer complaint behaviour, cultural values and consumer decision making in an Internet environment.

**Abstract** Although artificial neural networks (ANNs) are an established marketing support tool, they are criticised for their inability to explain their results. The purpose of this study is to demonstrate a visual approach to ANN variable interpretation. Scanner data are used to build a well-known choice model, first as a multinomial logit and secondly as an ANN. Response elasticity graphs are then built for each ANN model variable. These graphs are used to interpret the model variables and are benchmarked against the *t*-statistics of the logit. The results suggest that modellers using this visual ANN approach can obtain a result that aids in variable interpretation and possibly provides a richer understanding of model behaviour than a standard statistical methodology.

## INTRODUCTION

Marketers employ artificial neural networks (ANNs) in a number of operations including segmentation, choice modelling, brand share analysis and data mining. Today's use is due in part to the empirical evidence extolling the value of ANNs versus standard statistical techniques that began appearing in the marketing literature during the 1990s.<sup>1-4</sup> These studies provide support for the robustness and frequently greater predictive accuracy of ANNs *vis-à-vis* more traditional methodologies. Even though they have established themselves

as a prominent marketing support tool, ANNs are criticised for their failure to explain results. Central to this criticism is their inability to provide interpretation of the network connection weights. The purpose of this study is to demonstrate a methodology for ANN model variable interpretation that uses network connection weights. Empirical marketing data are used to optimise an ANN and a multinomial logit (MNL) model. Response elasticity graphs are built for each ANN model variable by plotting the derivative of the network output with respect to each variable, while

**Dr Kelly E. Fish**  
Department of Economics  
and Decision Sciences,  
Arkansas State University,  
Arkansas 72467, USA.  
Tel: +1 870 972 3416;  
e-mail: kfish@astate.edu

changing network input in equal increments across the range of inputs for each variable. These results are then benchmarked against the  $t$ -statistics of the MNL model. Since ANN variable interpretation is almost identical to the MNL model, it is concluded that this visual approach could be a valid method for ANN model variable interpretation but more empirical support is required. Additionally, it is demonstrated that it could provide a richer interpretation of variable importance than the standard statistical method.

In statistical modelling  $t$ -tests are often used as an indicator of model variable importance. In a simple regression model with one predictor variable ( $x$ ), the null hypothesis for the  $t$ -test is that a change in  $x$  yields no predicted change in model output ( $y$ ), and it follows that  $x$  has no value in predicting  $y$ . In other words, if the predicted  $y$ s were to be plotted against the various inputs of  $x$ , the slope of the line would be zero. In a model with multiple predictor variables the null hypothesis for each variable  $t$ -test is that a particular variable has no additional predictive value over and above that contributed by the other predictor variables; specifically if all other  $x$ s had already been used in the model and then an additional predictor variable ( $x_j$ ) was added last, no improvement in prediction would result. In other words, if predicted  $y$ s were to be plotted against changes in  $x_j$ , the partial slope of the line would be zero. While  $t$ -statistics have yet to be developed for ANNs, some of their logic can be extrapolated in determining variable importance by examining the actual partial slopes of lines based on predicted  $y$ s, given changes in respective  $x_j$ s.

Smith<sup>5</sup> suggests a method for ANN variable interpretation based on the elasticity of output from the derivative of the ANN equation and West *et al.*

demonstrate it using empirical data; however, they did not benchmark their results against a standard statistical model.<sup>6</sup> The contribution of this paper is that a similar, yet visual, approach to ANN variable interpretation is illustrated, and the first step towards validating it by benchmarking the results against the  $t$ -statistics of an MNL model is made.

## BENCHMARK MODEL

The prevalent discrete choice modelling process today is the multinomial logit (MNL). When faced with a choice a consumer is presumed to view each option as a bundle of attributes. The consumer is then assumed to construct an overall evaluation of each option by combining perceptions of the option's attributes through a utility function. The chosen alternative is the option with the highest overall evaluation (ie maximum utility). With these models it is additionally assumed that the utilities are not fully observable, thus the utility of each option is a combination of the determinable components (which are observable) and random components (which are not observable) with the random components being defined by a distribution of possible values. Formally stated, the utility ( $U$ ) an individual assigns to option  $i$  is linked to its determinable component or strict utility ( $\beta x$ ), and its random component or random utility ( $\varepsilon$ ):

$$U_i = \beta x_i + \varepsilon_i \quad (1)$$

MNL gives estimates of the probability that each option will be chosen based on assumptions about the form of the random component ( $\varepsilon$ ) distributions. It assumes a logistic distribution and is thus termed logit, with the multinomial logit having three or more categories in the dependent variable which eventually

results in the ‘us over us plus them’ type formulation:<sup>7</sup>

$$Pr\{i | C\} = e^{\beta x_i} / \sum_{j \in C} e^{\beta x_j} \quad (2)$$

where the vector  $\beta x_i$  is the deterministic component of the utility that the customer derives from the selection of alternative  $i$ , given choice set  $C$ .

A well-known consumer choice model is chosen as the benchmark model.<sup>8</sup> The model is considered a classic and recently researchers found that it still outperforms many notable models developed since its introduction.<sup>9</sup> Since retailers promote coffee brands and sizes separately, brand sizes are modelled as the choice alternatives. For example, Folger’s has both a large size (three pound) and a small size (one pound) in the data set; each of these constitutes a different brand size. When modelling a coffee market such as in this study, the model assumes that the deterministic component of the utility that customer  $k$  gains from the purchase of a given choice alternative is a function of:

- the alternative’s regular price
- whether or not the brand was on promotion
- presence or absence of a promotional price cut
- whether or not customer  $k$ ’s previous and second previous purchases were on promotion
- customer’s brand loyalty
- customer’s size loyalty.

Mathematically, the deterministic component of utility ( $\beta x_{ik}(n)$ ) can be represented as:

$$\begin{aligned} \beta x_{ik}(n) = & \beta_{0i} + \beta_1 RP_{ik}(n) \\ & + \beta_2 PROM_{ik}(n) \\ & + \beta_3 PCUT_{ik}(n) + \beta_4 PRV_{ik}(n) \\ & + \beta_5 SPRV_{ik}(n) + \beta_6 BL_{ik}(n) \\ & + \beta_7 SL_{ik}(n) \end{aligned} \quad (3)$$

where

- $\beta_{0i}$  = brand-size specific constant for alternative  $i$
- $RP_{ik}(n)$  = regular (depromoted) price of alternative  $i$  at time of customer  $k$ ’s  $n$ th coffee purchase (\$/oz)
- $PROM_{ik}(n)$  = a 0 – 1 variable denoting the presence or absence of a promotion on alternative  $i$  at purchase occasion  $n$
- $PCUT_{ik}(n)$  = promotional price cut on alternative  $i$  at the  $k$ th customer’s purchase (\$/oz)
- $PRV_{ik}(n)$  = a 0 – 1 variable denoting whether or not customer  $k$ ’s previous purchase was a promotional purchase of an alternative with the same brand as alternative  $i$
- $SPRV_{ik}(n)$  = a 0 – 1 variable denoting whether or not customer  $k$ ’s second previous purchase was a promotional purchase of an alternative with the same brand as alternative  $i$
- $BL_{ik}(n)$  = customer  $k$ ’s loyalty toward the brand of alternative  $i$  at purchase occasion  $n$
- $SL_{ik}(n)$  = customer  $k$ ’s loyalty toward the size of alternative  $i$  at purchase occasion  $n$ .

The variable BL (brand loyalty) is more specifically defined as:

$$\begin{aligned} BL_{ik}(n) = & \alpha_b BL_{ik}(n-1) \\ & + (1 - \alpha_b) \begin{cases} 1 & \text{if customer } k \text{ bought brand of alternative } i \text{ at} \\ & \text{purchase occasion } (n-1), 0 \text{ otherwise} \end{cases} \end{aligned} \quad (4)$$

where the carry-over constant is  $\alpha_b$ , which represents the exponential decay rate of the impact of the previous ten purchases (the reader is referred to the original paper<sup>10</sup> for an explication of how the particular rate ( $\alpha$ ) is determined for both brand loyalty and size loyalty. Based on previous research involving coffee markets the authors use  $\alpha_b = 0.875$  and  $\alpha_s = 0.812$ ).

To start up BL,  $BL_{ik(n)}$  is set to be  $\alpha_b$  if the brand of alternative  $i$  was the first purchase in the data history of consumer  $k$ , otherwise it is set at  $(1 - \alpha_b / (\text{number of brands} - 1))$ , thereby insuring that the sum of loyalties across brands always equals 1 for a consumer. To illustrate BL, suppose customer  $k$  bought alternative  $i$  on the first purchase occasion,  $BL_{ik(1)}$ , in such an instance  $\alpha_b = 0.875$ . On the second purchase occasion of customer  $k$ , the same brand is purchased resulting in an increase of brand loyalty:

$$BL_{ik(2)} = 0.890625 = (0.875(0.875) + (0.125)(1)).$$

On the next occasion suppose customer  $k$  selects another brand, resulting in a decrease in brand loyalty:

$$BL_{ik(3)} = 0.779296 = (0.875(0.890625) + (0.125)(0)).$$

Size loyalty (SL) is analogous:

$$SL_{ik}(n) = \alpha_s SL_{ik}(n-1) + (1 - \alpha_s) \left\{ \begin{array}{l} 1 \text{ if customer } k \text{ bought same size of alternative } i \text{ at} \\ \text{purchase occasion } (n-1), 0 \text{ otherwise} \end{array} \right\} \quad (5)$$

where  $\alpha_s$  is the carry-over constant for size. Initialisation methodology for SL is the same as BL.

Each of the first five brand sizes receives a brand-size specific constant ( $\beta_{0i}$ ) that is coded as one for the corresponding brand-size, zero otherwise. A brand-size constant of the sixth brand size is omitted in order to avoid singularity in the maximum likelihood estimation. These constants capture any uniqueness of an alternative that is not captured by the other explanatory variables. The ANN models will not use any brand-size specific constants.

## Data

This study uses the same scanner panel data that were used to develop an expected price model.<sup>11</sup> Selling Areas-Marketing, Inc. (SAMI) collected the data, which consist of individual records of ground coffee purchases from four Kansas City supermarkets over a 65-week period. As was done in previous studies, the data set was cleaned to remove panellists who joined or left during the 65-week period, who were extremely light users, or panellists who had gaps in their reported purchases. The top six brand-size combinations were retained and these account for 82.4 per cent of the total number of purchases and 87.1 per cent of the market share. Each of the rest of the brands that were excluded has less than 1 per cent market share. Twenty-five weeks of data are needed to initialise the brand loyalty and size loyalty variables leaving 40 weeks for the fitting/training and testing.

The data set does not contain any precise information on coupon, special point-of-purchase display or promotion. Thus the authors follow the previously accepted practice of assuming a sales promotion when any two of the following three criteria are met:

- a price reduction lasting one to four weeks
- an unusually high sales movement (ie store sales volume exceeding the mean level of unprompted sales by more than two standard deviations)
- the brand size is featured in a newspaper, flyer and/or bag stuffer during the week.

## MNL model calibration

Calibration of the multinomial logit models consisted of two separate fittings. In the replication fitting, the authors estimated a model using the entire

**Table 1:** Results of the MNL models (Coefficient estimates with the *t*-statistics in parentheses)

Variable	Kalwani <i>et al.</i> (1990) N = 3,330	Replication fitting N = 3,299	Second fitting N = 1,236
Brand size 1	-0.208 (-0.245)	-0.169 (-1.486)	-0.210 (-1.048)
Brand size 2	-0.239 (-1.873) <sup>b</sup>	-0.258 (-2.006) <sup>b</sup>	-0.268 (-1.233)
Brand size 3	0.488 (4.406) <sup>a</sup>	0.446 (3.980) <sup>a</sup>	0.962 (4.892) <sup>a</sup>
Brand size 4	0.249 (2.096) <sup>b</sup>	0.288 (2.400) <sup>b</sup>	0.420 (1.921)
Brand size 5	-0.105 (-0.135)	-0.142 (-1.296)	0.001 (0.007)
Brand loyalty	3.615 (36.087) <sup>a</sup>	3.791 (34.508) <sup>a</sup>	3.518 (20.653) <sup>a</sup>
Size loyalty	2.927 (26.111) <sup>a</sup>	2.908 (25.831) <sup>a</sup>	2.907 (14.855) <sup>a</sup>
Promotion	1.312 (13.284) <sup>a</sup>	1.437 (19.658) <sup>a</sup>	1.372 (11.616) <sup>a</sup>
Promotional price cut	0.278 (10.986) <sup>a</sup>	0.306 (15.854) <sup>a</sup>	0.296 (9.846) <sup>a</sup>
Regular price	-0.263 (-6.127) <sup>a</sup>	-0.278 (-6.430) <sup>a</sup>	-0.202 (-2.885) <sup>a</sup>
Previous promotional purchase	-0.228 (-3.561) <sup>a</sup>	-0.220 (-2.951) <sup>a</sup>	-0.041 (-0.336)
Second previous promotional purchase	-0.292 (-4.423) <sup>a</sup>	-0.249 (-3.274) <sup>a</sup>	-0.167 (-1.356)
$U^2$	0.5199	0.5297	0.5110
Adjusted $U^2$	0.5179	0.5277	0.5056
Log likelihood (at convergence)	-2864.3	-2780.0	-1082.9
Log likelihood (at zero)	-5966.6	-5911.0	-2214.6

<sup>a</sup>Significant at alpha = 0.01 in the one-tailed asymptotic *t*-test.

<sup>b</sup>Significant at alpha = 0.05 in the one-tailed asymptotic *t*-test.

40-week sample to compare results with previously published findings. The MNL calibration results for the replication phase of the study appear very close to the results reported by others,<sup>12</sup> who used the same data but with 31 additional observations. The small differences can be attributed to the difference in total purchase occasions contained in the respective data files. The previous study<sup>13</sup> had 3,330 purchases and the present study had 3,299 observed purchases for the 40-week period (actually the authors'  $U^2$  of 0.5297 was slightly higher and this better fit is also attributable to the difference in observations). The authors thus conclude that the MNL model used in the current study represents a model essentially identical to previous models.

In order to benchmark the MNL

model against an ANN a second fitting was used where 108 households are retained for fitting and 108 households are used as a holdout sample. This 50:50 split of the data was used because it is what the choice model's authors<sup>14</sup> used and it provides for a large holdout sample containing both interpolative and extrapolative results. Accordingly, the last 12 weeks were trimmed off the fitting sample so that the holdout sample has 28 weeks of interpolative results and 12 weeks of extrapolative results. This last phase results in data sets with 1,236 purchases in the fitting sample and 1,564 purchases in the holdout sample (1,124 interpolative, 440 extrapolative). All model estimations are accomplished using the discrete choice procedure in LIMDEP 6.0; calibration results are shown in Table 1.

**Table 2:** Holdout sample brand share estimation results (MAE)

	Entire holdout sample (240 weeks)	Interpolative subsample (168 weeks)	Extrapolative subsample (72 weeks)
MNL	5.68	4.83	7.66
ANN	4.65	4.06	6.03

### ANN model optimisation

In this study the authors use a feedforward network trained by the Genetic Adaptive Neural Network Training (GANNT) algorithm.<sup>15</sup> Although similar to previous genetic algorithms, GANNT has produced better results than prior efforts and has been shown to be superior to backpropagation training.<sup>16,17</sup> While marketing models are often optimised on individual choice, they are frequently used by managers to estimate brand share. Thus the aim is to know how well each model predicts brand share. Mean absolute error (MAE) is used as a measure of each model's ability to estimate brand share and it is calculated by finding the weekly differences (errors) between the actual market share and predicted market share for each brand size. The absolute values of those errors over the  $N$  week period are then averaged.

Finding the optimum number of hidden nodes for an ANN has generally been considered a somewhat crude process. The method the authors use is to start with five hidden nodes and train the network, then build another network with five more hidden nodes and train it. If the ten hidden node network outperforms the five hidden node network on an in-sample validation test, then a 15 hidden node network is built and trained. This building-testing cycle is continued until the addition of five hidden nodes does not improve network performance. At that point it can be fine-tuned by adding or subtracting one hidden node at a time to find the optimal number of hidden nodes. With

the authors' ANN they are attempting to predict brand shares so lowest in-sample MAE is used as the decision criterion for hidden node optimisation. This hidden node optimisation process results in an ANN model that contains one output variable, four hidden nodes, seven input variables along a bias node connecting to both the hidden layer and output layer.

The MNL and ANN models are tested on the holdout sample and the results are given in Table 2. The ANN's MAEs appear to be lower than MNL's in the entire holdout sample, as well as the interpolative and extrapolative subsamples. The authors test for a statistically significant difference in MAEs using a two-tailed paired  $t$ -test and find the null hypothesis rejected at the 99+ per cent level ( $p < 0.001$ ) for the holdout sample. If just the interpolative part of the holdout sample is examined a difference between the MAEs at the 98+ per cent level ( $p < 0.013$ ) is found and when just the results of the extrapolative subsample are looked at a difference at the 96+ per cent level ( $p < 0.031$ ) is found. The authors thus conclude that they have an optimised ANN model that outperforms a very good MNL choice model.

### DETERMINING RESPONSE ELASTICITIES FROM THE ANN MODEL

The connection weights of the ANN model are shown in Table 3 and it is these weights that are used to calculate network output ( $y$ ). Recall there are seven model variables  $X_{1t}, X_{2t}, \dots, X_{7t}$ ,

**Table 3:** Connection weights for ANN choice model

Weight matrix from input layer to hidden layer				
From input	H-1	H-2	H-3	H-4
Bias node	0.0148	-0.0544	-0.0363	0.0752
Regular prices (RP)	-0.0616	0.0892	0.0387	0.0192
Promotion (PROM)	-0.0291	0.0434	0.0578	-0.0102
Price cut (PCUT)	-0.0208	-0.0328	0.0722	0.0359
Previous purchase (PRV)	0.0684	0.0195	-0.0881	0.0763
Second previous purchase (SPRV)	-0.0090	0.0543	0.0223	0.0482
Brand loyalty (BL)	-0.0932	0.0935	0.0492	0.1036
Size loyalty (SL)	-0.0588	0.0860	0.0199	0.0863
Weight matrix from hidden layer to output layer				
From hidden layer	Output (Y)			
Bias node	19.75448			
Hidden node 1 (H-1)	18.96961			
Hidden node 2 (H-2)	10.43411			
Hidden node 3 (H-3)	13.69286			
Hidden node 4 (H-4)	10.43411			

where  $t = 1$  to  $T$ , the total number of observations. There are four hidden nodes,  $h_{1t}$ ,  $h_{2t}$ ,  $h_{3t}$ ,  $h_{4t}$ , a bias node,  $\gamma_5$ , and there is one output node  $y_t$ . Therefore the ANN equation can be written as:

$$y_t = \sum_{i=1}^4 \gamma_i h_{it} + \gamma_5, \tag{6}$$

where

$$h_{it} = \frac{1}{1 + e^{-\sum_{k=1}^7 W_{ik} X_{kt} + W_{ik+1}}} \tag{7}$$

which gives the output of each  $i$ th hidden node at observation  $t$  resulting from the sigmoidal transformation of the summed hidden node inputs; which are seven input variables ( $X_{kt}$ ) each multiplied by a connection weight ( $W_{ik}$ ), with  $W_{ik+1}$  being the input of the biased node to that hidden node.

Additionally, the derivative of the ANN equation with respect to one of the variables  $X_m$  can be written as:

$$\frac{\delta y}{\delta x} = \sum_{i=1}^4 \gamma_i \frac{\delta h_i}{\delta X_m}, \tag{8}$$

where

$$\frac{\delta h_i}{\delta X_m} = \frac{W_{im} e^{-\sum_{j=1}^7 W_{ij} X_j + W_i^8}}{\left(1 + e^{-\sum_{j=1}^7 W_{ij} X_j + W_i^8}\right)^2} \tag{9}$$

and by taking the derivative of the network output, after training the full model and then varying each predictor variable across its range in the data file in equal increments while holding the other predictor variables fixed, the slopes of the functional relationships of the variables can be compared and used as indicators of response elasticities.

A similar method concerning model output based on Equation 6 could be used to obtain an indicator of response elasticity of variables, if all the variables' units and data ranges were equal. This is not, however, the case with the data here (GANNT does not normalise the data before submitting them to the network as backpropagation does), some variables use nominal data ranging from zero to one, while others (eg Regular Price, RP) use ratio data ranging from 15 to 21. Even though direct comparison

among variables is precluded, such analysis can give an indication of how the model behaves with respect to a certain variable. Response graphs for each model variable were developed and are presented in Figure 1.

In examining the loyalty variables, BL reflects an almost linear relationship — increases in BL have a strong positive effect on the consumer's propensity to purchase the brand. Brand loyal customers (greater than 0.75 on the BL input) cause network output to be above 0.6. In examining the derivative graph, the slope of the response function steadily increases slightly, while ranging from 0.475 at zero brand loyalty to 0.482 at perfect brand loyalty. This supports the notion that the relationship is not perfectly linear and that the rate in change increases as the customer becomes more loyal. SL shows a similar relationship, though not as profound. Network output ranges from 0.254 with no size loyalty to 0.533 for perfect size loyalty. The  $dy/dx$  function shows this relationship to be almost linear, ranging only from 0.278 to 0.279.

The response graph for promotional price cut (PCUT) shows a non-linear relationship between the amount of the price cut and network output as the graph begins to flatten at the points of large price cuts. The PCUT chart contains negative inputs which means that the promotional price is actually higher than the regular price. This can occur because of the previously mentioned defect in the database requiring that a promotion be assumed when two of three conditions are met. If there is an unusually high sales movement and the brand size is featured in a print promotion, then it is considered a promotional purchase even though the price may be higher than the regular price, thus a negative price cut. The downward sloping  $dy/dx$  graph also

shows non-linearity, ranging from 0.088 to 0.005.

Regular price (RP) results indicate that increases in the non-promoted price of a brand have a slightly negative effect on network output. Network output only varies from 0.426 to 0.35 across the range of inputs. The slope of the curve ranges from  $-0.004$  to  $-0.017$ , which is reflected in a negative and downward sloping curve of the derivative graph.

Lines are created for the dichotomous predictor variables by inputting zero then one into the network equation and its derivative. The promotion (PROM) variable shows the steepest slope, it is also positive. When there is no sale promotion network output is at 0.373, this increases to 0.506 when a promotion takes place. This supports the well-known fact that sale promotions have a positive effect on sales. This effect is substantial and depicted in the derivative output 0.13.

Both the previous purchase (PRV) and second previous purchase (SPRV) have downwardly sloping response functions and negative derivatives. If a customer's previous purchase of a brand size was on promotion then network output is reduced to 0.324 from 0.404. The effects of the SPRV variable are less profound with network output dropping only from 0.395 to 0.382 if the second previous purchase was made on promotion. An individual's previous promotional purchase history has more influence (derivative output of  $-0.08$ ) than that person's second previous promotional purchase history (derivative output of  $-0.013$ ). These results indicate that previous promotional purchases decrease the likelihood of a subsequent purchase of the brand, a notion well supported in the literature.<sup>18,19</sup>

With the MNL model containing qualitative and quantitative variables of differing units,  $t$ -statistics can be used as



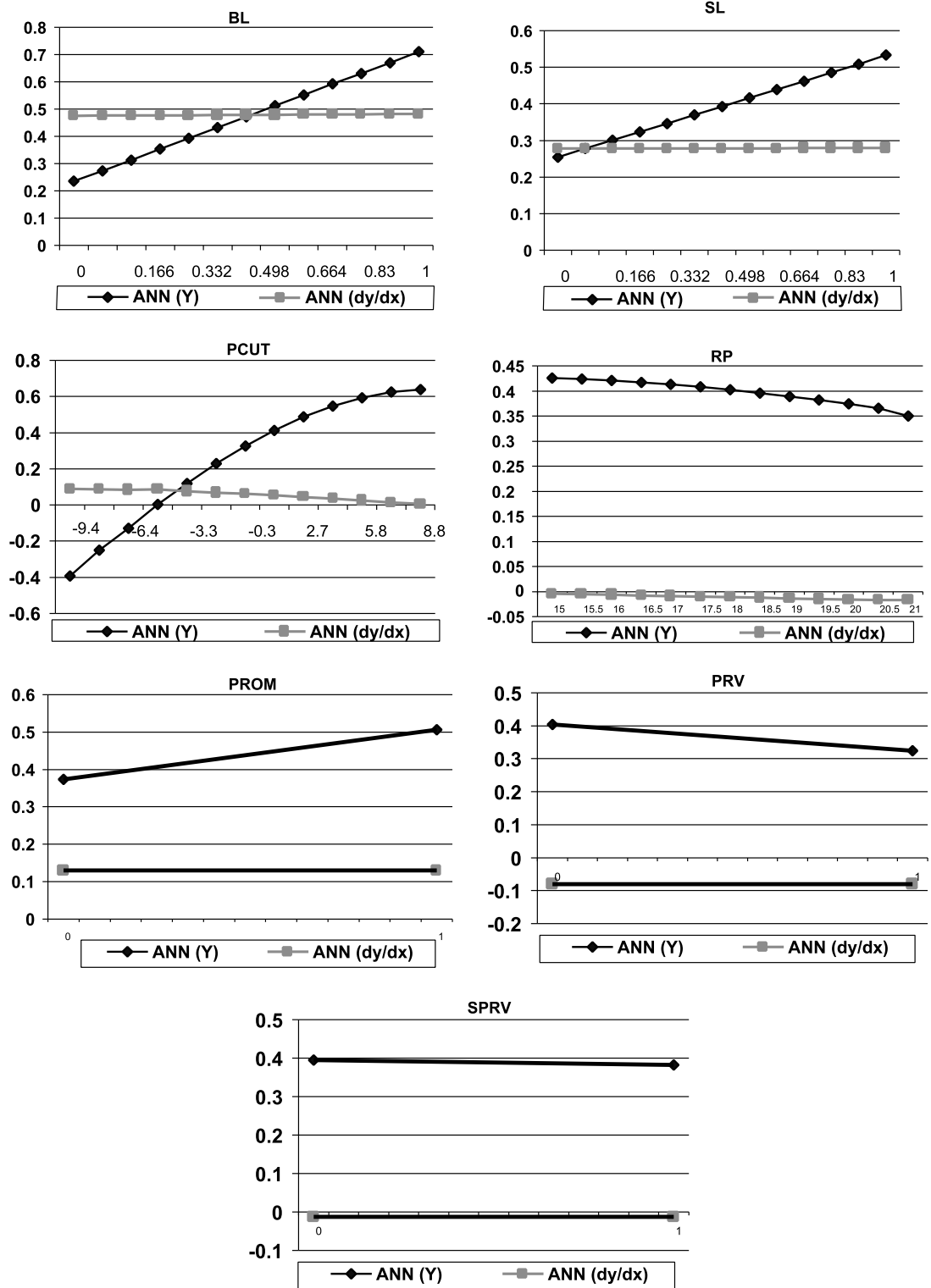


Figure 1 Response elasticities of the ANN model variables

an indication of explanatory importance. In the GANNT-trained ANN model the derivatives of the output can be compared to determine variable importance. The MNL  $t$ -values show BL, SL, PROM and PCUT to be the four most important variables in the model in descending order. The three other variables RP, PRV and SPRV are shown to be less important. The results are in general agreement with the MNL model; in descending order, the derivative graphs show BL, SL, PROM to be the three most important variables while the importance of PCUT varies. The efficacy of PCUT depends on the extent of reduction from the non-promoted price. The derivative approaches zero in the area of deep price cuts. Marketing managers should be aware that the effectiveness of promotional price cuts lessens as the cuts become deeper. In other words, the marketing manager gets more bang for the lost buck with cuts closer to the non-promoted price. Deeper price cuts will have little effect in spurring additional sales. The SPRV variable is shown to be of little explanatory importance based on its derivative being close to zero. It is also worth noting that the negative relationships of RP, PRV and SPRV are shown in the MNL model with negative coefficients and  $t$ -values, while in the ANN-GANNT model this type of relationship is captured with negative derivative outputs.

## DISCUSSION OF RESULTS AND CONCLUSIONS

Since the variable interpretation results are similar to the MNL model, the original discussion of marketing ramifications of the variables still suffices.<sup>20</sup> However, the ANN approach can provide managers with additional information. The varying degree of

efficacy of promotional price cuts (PCUT) is not apparent with the MNL model because the  $t$ -statistic is a constant elasticity measure. Additionally, the RP graph tells researchers that as regular price points increase across the market range, they have an increasingly negative effect on sales. Although the negative relationship is captured in the MNL model, the *changes* in the rate of effectiveness are not captured with statistical constants.

This benchmark study presents empirical evidence that takes a first step in validating a visual approach to ANN variable interpretation. Any sense of true validation can only come with further testing. Future variable interpretative research should focus not only on comparing results against MNL in other modelling situations but also against other standard statistical techniques (eg multiple linear regression). Although a feedforward network trained by a genetic algorithm is used, this method of variable interpretation is amenable to backpropagation trained networks as well. ANN modellers should consider such an approach after they develop marketing models.

## References

- 1 Kumar, A., Rao, V. R. and Soni, H. (1995) 'An empirical comparison of neural network and logistic regression models', *Marketing Letters*, Vol. 6, No. 4, pp. 251–263.
- 2 Fish, K. E., Barnes, J. H. and Aiken, M. (1995) 'Artificial neural networks: A new methodology for industrial market segmentation', *Industrial Marketing Management*, Vol. 24, pp. 431–438.
- 3 Agrawal, D. and Schorling, C. (1996) 'Market share forecasting: An empirical comparison of artificial neural networks and multinomial logit model', *Journal of Retailing*, Vol. 72, pp. 383–407.
- 4 West, P. M., Brockett, P. L. and Golden, L. L. (1997) 'A comparative analysis of neural networks and statistical methods for predicting consumer choice', *Marketing Science*, Vol. 16, No. 4, pp. 370–391.
- 5 Smith, M. (1996) 'Neural networks for statistical modeling', International Thomson Computer Press, Boston, MA.
- 6 West *et al.* (1997) *op cit.*

- 7 Luce, R. D. (1959) 'Individual choice behavior', Wiley and Sons, Inc., New York, NY.
- 8 Guadagni, P. M. and Little, J. D. C. (1983) 'A logit model of brand choice calibrated on scanner data', *Marketing Science*, Vol. 2, Summer, pp. 203–238.
- 9 Abramson, C., Andrews, R. L., Currim, I. S. and Jones, M. (2000) 'Parameter bias from unobserved effects in the multinomial logit model of consumer choice,' *Journal of Marketing Research*, Vol. 37, pp. 410–426.
- 10 Guadagni and Little (1983) *op cit.*
- 11 Kalwani, M. U., Yim, C. K., Rinne, H. J. and Sugita, Y. (1990) 'A price expectations model of customer brand choice', *Journal of Marketing Research*, Vol. 27, August, pp. 251–262.
- 12 *Ibid.*
- 13 *ibid.*
- 14 Guadagni and Little (1983) *op cit.*
- 15 Dorsey, R. E., Johnson, J. D. and Mayer, W. J. (1994) 'A genetic algorithm for the training of feedforward neural networks', *Advances in Artificial Intelligence in Economics, Finance and Management*, Vol. 1, J. D. Johnson and A. B. Whinston (eds), JAI Press, Greenwich, CT, pp. 93–111.
- 16 Gupta, J. N. D. and Sexton, R. S. (1999) 'Comparing backpropagation with a genetic algorithm for neural network training', *Omega*, Vol. 27, No. 6, pp. 679–684.
- 17 Sexton, R. S. and Dorsey, R. E. (2000) 'Reliable classification using neural networks: A genetic algorithm and backpropagation comparison', *Decision Support Systems*, Vol. 30, No. 1, pp. 11–22.
- 18 Shoemaker, R. W. and Shoaf, R. F. (1981) 'Repeat rates of deal purchases', *Journal of Advertising Research*, Vol. 17, April, pp. 47–53.
- 19 Dobson, J. A., Tybout, A. M. and Sternthal, B. (1978) 'Impact of deals and deal retraction on brand switching', *Journal of Marketing Research*, Vol. 15, February, pp. 72–81.
- 20 Guadagni and Little (1983) *op cit.*