Journal of Forecasting
J. Forecast. **19**, 177–200 (2000)

Neural Networks and the Multinomial Logit for Brand Choice Modelling: a Hybrid Approach

YVES BENTZ[1]* and DWIGHT MERUNKA[2]
[1] *London Business School, UK*
[2] *IAE Aix-en-Provence, France*

ABSTRACT

The study of brand choice decisions with multiple alternatives has been successfully modelled for more than a decade using the Multinomial Logit model. Recently, neural network modelling has received increasing attention and has been applied to an array of marketing problems such as market response or segmentation. We show that a Feedforward Neural Network with Softmax output units and shared weights can be viewed as a generalization of the Multinomial Logit model. The main difference between the two approaches lies in the ability of neural networks to model non-linear preferences with few (if any) *a priori* assumptions about the nature of the underlying utility function, while the Multinomial Logit can suffer from a specification bias. Being complementary, these approaches are combined into a single framework. The neural network is used as a diagnostic and specification tool for the Logit model, which will provide interpretable coefficients and significance statistics. The method is illustrated on an artificial dataset where the market is heterogeneous. We then apply the approach to panel scanner data of purchase records, using the Logit to analyse the non-linearities detected by the neural network.

KEY WORDS neural networks; Multinomial Logit model; choice models; brand choice

INTRODUCTION

The study of the impact of marketing mix variables, in particular price and promotions, on consumer good sales is increasingly based on the understanding of how consumer behaviours are influenced by these variables, at an individual or household level of analysis. This is essentially due to the growing availability of scanner data of purchase histories both in the United States and in Europe. It is also due to the development and relative ease of use of analytical methods for studying and predicting consumer choice at a disaggregated level. Over the last fifteen years, many marketing researchers working with household-level scanner data have used the

* Correspondence to: Yves Bentz, London Business School, Sussex Place, Regent's Park, London NW1 4SA, UK.

Received February 1997
Accepted April 1998

Multinomial Logit model (MNL) (MacFadden, 1974) to study choice decisions involving multiple alternatives. This model has been very helpful for understanding and predicting brand choice behaviour (Guadagni and Little, 1983) and studying the effects of marketing mix and demographic variables on households' choice probabilities of brands, in particular the responsiveness to promotions (Lattin and Bucklin, 1989), variety-seeking behaviour (Lattin, 1987) and advertising (Tellis, 1988).

Logit models have the appeal of being stochastic and yet admitting decision variables. These decision variables, such as price, promotional price cut, advertising spending or brand loyalty, constitute the deterministic part of the utility function that is used to compute the probability of choosing a specific product among several alternatives. A major assumption of the model is the linearity of this utility function. The only way to take into account non-linearities is to add to the model variables that represent potential non-linear effects (quadratic terms, for instance). This requires, however, assumptions about the nature of the underlying utility function and therefore introduces a specification bias that may be unappealing in some marketing applications as the model assumptions can lead to misinterpretations (regarding for instance feature, display, price, etc.).

In order to overcome the limitations of the MNL model, it may be useful to use more general models that are capable of modelling non-linear utility functions without requiring an *a priori* knowledge of the type of relationships to be modelled. Such models are generally known as non- and semi-parametric, non-linear regression models. Models such as kernel smoothers (Härdle, 1990), projection pursuit (MARS models—Friedman, 1991), generalized additive models (Hastie and Tibshirani, 1990), penalized regressions and flexible discriminant analysis (Hastie *et al*., 1994, 1995) and artificial neural network models belong to this category. In this paper, we shall only address neural networks (1) because of their resemblance to prevailing models, (2) because they can be visually and graphically described and (3) because they can be easily used by experienced econometricians as well as non-specialists.

Artificial Neural Networks (ANNs) is a field of research that has enjoyed increasing popularity in the marketing research community. Neural networks have been applied to a large variety of problems going from consumer profiling (Kumar *et al*., 1995) and market segmentation (Hruschka and Natter, 1994) to market response analysis (Hruschka, 1993), database marketing (Burgess, 1995) and scoring (Desmet, 1995). It has been shown that ANNs can be viewed as non-linear generalizations of models such as linear or logistic regressions (Hertz *et al*., 1991; Ripley, 1994), discriminant and factor analysis (Baldi and Hornik, 1989) or cluster analysis (Kohonen, 1988).

In a similar way, the MNL has a neural representation. A perceptron with log-linear outputs (better known as 'Softmax' outputs (Bridle, 1990) is identical to a MNL, provided the network has no hidden neurones. However, the complexity of the ANN model can easily be increased by changing the architecture of the network, enabling more complex relationships resulting from non-linear consumer preferences (threshold and interaction effects).

Comparing the predictive performances of two models, one of which is a restriction of the other, makes sense only insofar as this comparison reveals the existence of relationships that are taken into account by one model (i.e. ANN) and not by the other (i.e. MNL). Choosing between the two models implicitly requires the user to decide whether the complexity of the data justifies the use of the more complex one. This is precisely the spirit of the paper, i.e. comparing the performances of the two approaches in order to evaluate the significance of non-linearities in the dataset.

Beyond this simple diagnostic, and from a more practical perspective, the two classes of model considered present methodological differences. Unlike parametric models, ANNs do not suffer from a specification bias and are able to model highly complex relationships. However, they are very difficult to interpret and have been perceived as 'black boxes' which can neither explain how they reach an outcome nor provide an explicit representation of the relationship that they estimate. On the other hand, MNL models provide easily interpretable coefficients and significance statistics. When studying consumer behaviours, interpretability is at least as important as predictive power. However, because the respective strengths and weaknesses of the models are complementary, it may be possible to combine them in a single framework in order to take into account potential non-linearities as well as to improve the model interpretability.

The remainder of the article consists of four sections. The next two sections describe the Multinomial Logit model as well as its neural counterpart and present a hybrid approach combining the two types of models. This approach is then illustrated with the simple example of a heterogeneous market, highlighting the interest and the stages of the proposed method. Finally, the approach is applied to Australian scanner data of household purchases on the instant coffee market.

THE MULTINOMIAL LOGIT MODEL (MNL)

Logit models are the natural complement of regression models when the regressand is not a continuous variable but a state which may or may not be obtained, or a category in a given classification. It is particularly helpful in behavioural sciences, for many of the behaviours of interest are nominal, or at least observed qualitatively. For instance, voting, buying or selling a good, entering a contract and getting married are all behaviours that can be measured in only a small number of categories. Logit modelling has become quite popular among the social sciences community because it provides much explanatory power, due to its multivariate nature, and also because it is easy to implement and to interpret. Logit modelling has been extensively described in the literature (Ben Akiva and Lerman, 1985; Gensch and Recker, 1979; Malhotra, 1984; Guadagni and Little, 1983), and we shall only describe briefly the characteristics of these models. However, in order to compare Logit models to artificial neural networks, it is important to mention their fundamental equations and properties.

Model description

We shall first describe the Binomial Logit model (or 'logistic regression'), for it is a particular case of the more general Multinomial Logit Model. In the case where the dependent variable is dichotomous, a binomial Logit (or logistic regression) model can be used. The binomial Logit model computes the probability of acceptance of the ith item, $P_i(Y = 1)$ as a function of several explanatory variables. The relationship between P_i and these variables is given by

$$\begin{aligned} P_i(Y = 1) &= f(\text{Utility}) \\ \text{Utility} &= a + \sum_{j \in T} b_j \times x_{ij} + \varepsilon \end{aligned} \tag{1}$$

where

a, b_j are the coefficients of the model, computed by Maximum Likelihood Estimation
x_{ji} the descriptors measured for the ith item

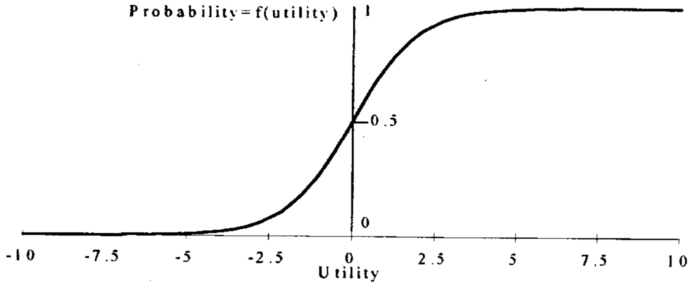

Figure 1. Graph of the logistic function

ε an additive noise that is partly the results of unobserved variables and f the logistic function, or sigmoid defined by

$$f(x) = \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-x}} \tag{2}$$

As we can see from equation (1), Logit models differ from multiple regression by the non-linear function (i.e. the sigmoid or logistic function) which maps utility into a choice probability. The shape of the sigmoid is given in Figure 1. As shown in Figure 1, the sigmoid constrains the probability to the limited range from 0 to 1, without also constraining the utility.

Another property of this function is non-linearity. This transforms a utility that is a linear function of explanatory variables into a probability that is non-linear with respect to these variables. Hence the Logit model is non-linear in the sense that the sensitivity of an output to a particular input is a *function of the level of that input as well as of the level of all the other inputs of the model*. For instance, if we model the probability of a household possessing its own house, an increase in the household's wealth of £50,000 will not have the same influence on the probability whether the initial wealth was £1000 or £1 million. It is very important to note that this sensitivity is not conditional to each individual input variable but to all of them, as represented by the utility function. Therefore, an input variable can only interact with input variables as a whole rather than with a particular one. This constraint results from the linearity of the utility function.

The Multinomial Logit (MNL) model extends the Binomial Logit to more than two states and can therefore be used to model discrete choices. It computes the probability $P_i(Y = j)$ of choosing alternative j on choice occasion i as a function of all other alternatives. As for the Binomial Logit, it assumes that a linear combination of the explanatory variables is related to the choice probability $P_i(Y = j)$ as described by

$$P_i(Y = j) = \frac{e^{v_{ij}}}{\sum_{j=1}^{J} e^{v_{ij}}} \tag{3}$$

where

$P_i(Y = j)$ is the probability of choosing alternative j on choice occasion i

v_{ij} is the deterministic part of the utility of alternative j on occasion i,

$$v_{ij} = \sum_{k \in T} b_k \times x_{ijk}$$

with

b_k: kth coefficient in the utility function
x_{ijk}: value of attribute k for product j on purchase occasion i
J: number of alternatives considered
T: set of attributes.

The model coefficients are calculated by Maximum Likelihood Estimation (MLE). Assuming that the stochastic part ε of the utility function is distributed with a double exponential distribution, the likelihood of observing actual purchase choices, given input vector X and model parameter vector B, can be expressed by

$$L(Y|X, \ B) = \prod_{i=1}^{N} P_i^{Y_i}(1 - P_i)^{1-Y_i} = \prod_{i=1}^{N} \prod_{j=1}^{J} \left(\frac{e^{v_{ij}(x_{ij}, b)}}{\sum_{h=1}^{J} e^{v_{ih}(x_{ih}, b)}} \right)^{Y_{ij}} \tag{4}$$

where

$Y = (y_1 \dots y_n)$ is the dependent variable (categorical)
$X = (x_{11}, \dots, x_{1n}, \dots, x_{kn})$ represents the explanatory variables
$B = (b_1, \dots, b_k)$ represents the model parameters
P_i is the predicted probability for choice occasion i
v_{ij} is the value of the utility function for the jth alternative on the ith choice occasion
N is the number of choice occasions and J is the number of alternatives.

In practice, however, it is preferable to maximise the logarithm of the likelihood as expressed in

$$\log L = \sum_{i=1}^{N} \sum_{j=1}^{J} Y_{ij} \ln(P_i(Y = j/x, \ B)) \quad \text{with } P_i(Y = j/x, \ B) = \frac{e^{v_{ij}(x_{ij}, b)}}{\sum_{h=1}^{J} e^{v_{ih}(x_{ih}, b)}} \tag{5}$$

Three interesting properties of the MNL model arise from equation (3). First, because equation (3) can be rewritten

$$P_i(Y = j) = \frac{1}{\sum_{k \in T} e^{(v_{ik} - v_{ik})}}$$

the utility is undetermined to the extent of an additive constant. For example, if we add a constant to a variable for all the alternatives, $P_i(Y = j)$ will not be affected. Second, if there are only two alternatives, equation (3) can be rewritten as

$$P_i(Y = 1) = \frac{e^{(v_{i1})}}{\sum_{k \in S} e^{(v_{ik})}} = \frac{e^{(v_{i1})}}{e^0 + e^{(v_{i1})}} = \frac{e^{(v_{i1})}}{1 + e^{(v_{i1})}}$$

which shows clearly that the MNL is an extension of the logistic regression model. Finally, $P_i(Y = j)$ is S-shaped in v_{ij} when the other v_{ik} are held constant. Therefore, as for the binomial Logit, large or small values for v_{ij} make $P_i(Y = j)$ flat and insensitive to changes in v_{ij}.

Before describing neural network models and discussing their resemblance with Logit modelling, it is important to understand some of the assumptions made to derive Logit models.

Some limiting assumptions of Logit modelling

As for most models, MNL modelling rests on some simplifying assumptions. Making such assumptions is often suitable because it limits the model complexity and therefore limits the variance of the model, although it may introduce a bias, in particular when the assumptions made are far from reality. This is known as the bias/variance dilemma. Low-complexity models (with strong assumptions) are low-variance/high-bias models, whereas high-complexity models (with weak, if any assumptions) are high-variance/low-bias models. One of the aims of model specification and data pre-processing is therefore to find the optimal level of model complexity, i.e. a complexity level that is justified by the complexity of the data. When dealing with noisy and parsimonious data, low-complexity models should be preferred. When dealing with numerous and clean data, more complex models can be employed.

MNL modelling rests on various assumptions, three of which can have serious implications. The first assumption is known as the assumption of 'independence from irrelevant alternatives'. It supposes that the consumer ignores the positioning similarities among alternatives when he or she chooses a product. Hence, if we add to the alternatives a new one that is similar to an already existing one, the new alternative will reduce the probabilities of all alternatives rather than split the probability of the one it resembles (Meyer and Kahn, 1991). Violations of the IIA assumption can be formally tested (McFadden, Tye and Train, 1977). The limitation might also be overcome by the use of nested models (MacFadden, 1981), which cluster alternatives into a hierarchical tree structure according to their similarities and the underlying consumer choice process.

The second assumption is that all individuals (or households) choose among all available alternatives. This certainly does not hold in most cases as shown by the important research current concerning the role of consideration sets in consumer choice behaviour (Shocker *et al.*, 1991). Consumers choose from a restricted set of alternatives related to the purchase goals and occasion. The logit model has thus been criticized because the estimated consumer preferences reflect not only the true preferences of the consumers but also the characteristics of the consideration set. Recently, several modelling approaches have been developed to take into account the consideration phase of the choice process within a logit framework (Vanhonacker, 1993; Andrews and Srinivasan, 1995; Siddarth *et al.*, 1995).

Finally, interaction effects between the explanatory variables have been given very little attention in the literature (Bentz and Merunka, 1995). It is basically this issue which will be discussed and addressed in this paper. The linearity assumption about the utility function makes it difficult to model interactions between two variables and more generally non-linear effects in the utility function (threshold and interactions effects). Only a simple non-linear effect is modelled in the MNL model. This effect is taken into account by the output sigmoid, which formalizes the fact that the sensitivity of the utility function to inputs is conditional to its own level. In other words, a variable can only interact with all the other variables as a whole instead of interacting with a particular one. One way to take these non-linearities into account is by representing them explicitly in the utility function by additional variables, for instance quadratic terms (x^2, $x_1 \cdot x_2$, ...). However, in order to incorporate such non-linear effects into the utility function, the analyst needs either to know *a priori* where these potential effects lie or to try all possible combinations of input variables and term orders. This searching procedure can quickly become tedious, especially so when the number of variables is large and the market heterogeneous.

The three limitations that have been addressed are independent of each other and the first two have been overcome by the use of nested models and cautious pre-processing. As for the third limitation, non-linear non-parametric models such as neural networks can provide a solution.

FEEDFORWARD NEURAL NETWORKS AS AN EXTENSION OF LOGIT MODELS

In the recent years, neural networks have been increasingly applied to a wide range of business problems, going from modelling financial markets (Refenes, 1995; Trippi and Turban, 1992) and loan risk analysis (Burgess, 1995) to database marketing (Burgess, 1995) and market responses (Hruschka, 1993). Neural networks are particularly suitable for marketing applications such as brand choice models because the databases available usually contain a large number of examples, some variables such as perceived quality and price are known to interact and the consumer base can be composed of a number of different sub-markets.

The type of neural networks that are considered in this paper is the multi-layer feedforward neural network trained with the standard 'backpropagation' algorithm. Details of the algorithm can be found in any textbook on neural networks (Rumelhart and McClelland, 1986) and will not be addressed in this paper. However, it is interesting to draw a parallel between neural networks with sigmoidal outputs (respectively Softmax outputs) and logistic regression models (respectively Multinomial Logit model).

Analogy between feedforward neural networks and the binomial Logit model

Feedforward neural networks can be considered as non-linear regression models, the complexity of which can be changed. At their lowest complexity level, they consist only in an input layer and an output. The architecture of such a network is illustrated in Figure 2.

These relationships between the inputs and the outputs are identical to the Binomial Logit (equations (1) and (2)). Parameters of feedforward neural networks are usually estimated by minimizing the mean square error produced by the model. This is equivalent to maximum likelihood estimation when the dependent variable is a continuous function of the inputs with additive Gaussian noise, as is the case for regression problems. For classification problems, however, the dependent variable is binary and a Gaussian distribution for the errors is inappropriate and therefore the error function is *a priori* not mean square error. The entropy

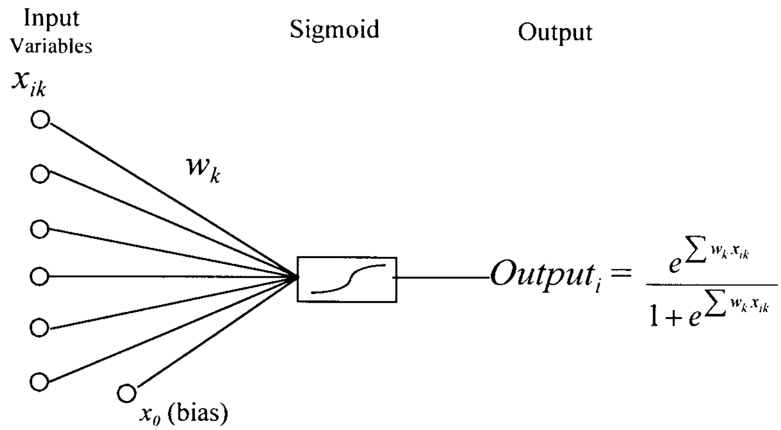

Figure 2. Architecture of a neural network without any hidden layer (perceptron). x_{jk} are the inputs (attributes of alternative k) of the model, and w_j are the coefficients to be estimated

function (Hopfield, 1987) is a more appropriate cost function. Equation (6) defines the entropy function E:

$$E = -\sum_{i=1}^{M}(Y_i \ln(S_i) + (1 - Y_i)\ln(1 - S_i)) \tag{6}$$

where S_i is the network output and Y_i the desired value.

Apart from its sign, this function is identical to the log-likelihood function used to estimate the coefficients of the Binomial Logit model (equation (5)). The feedforward network with a sigmoidal output is therefore exactly identical in form and estimation procedure to the binomial Logit model. It models the choice probability based on a linear utility function and therefore does not take into account potential interaction effects between product attributes in consumer preferences. However, it is possible to change the network complexity in order to take into account non-linearities, in particular interaction effects. This is done by adding one or more hidden layer(s) to the network. Each of these hidden layers contains neurones which act as the simple perceptron presented in Figure 2. The new architecture is presented in Figure 3.

Intuitively, we can think of each hidden node as the output of a linear regression that can be switched on or off according to the values of the variables. This process enables the network to map interactions and more generally non-linear relationships. Hence, a feedforward neural network with sigmoidal output can be expressed in the form of equation (7) and can therefore be viewed as a generalization of the Binomial model:

$$P_i(Y = 1) = \frac{1}{1 + e^{-g(\vec{x}_i)}} \tag{7}$$

with g being any function of the decision variables $\vec{x}_i$.

It can be shown (Hornik *et al.*, 1989) that neural networks with one hidden layer can approximate any function on compact sets, without any *a priori* assumption about the function. This feature makes the neural network a powerful tool to model consumer choices. On the other hand, because neural networks can map any kind of function, if they are used without due care

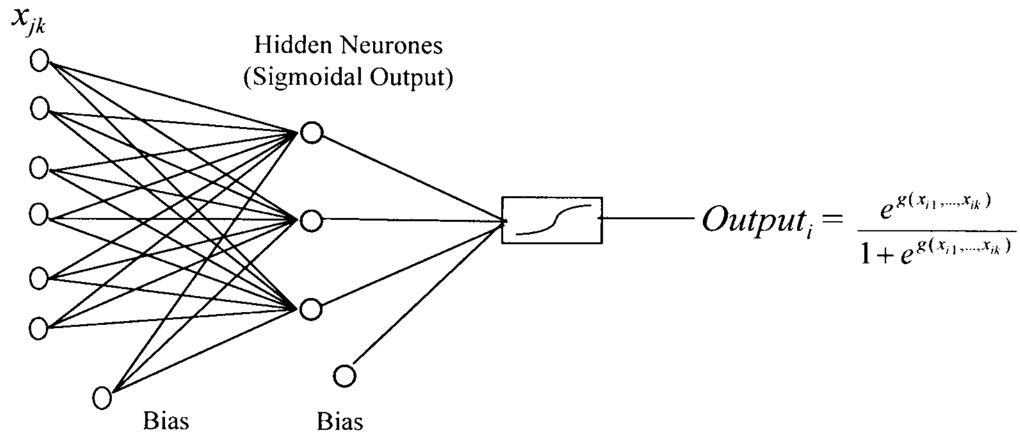

Figure 3. Neural network architecture with one hidden layer and sigmoidal output. This network is able to model interactions between individual variables

they may also pick up some of the noise in the dataset, especially if the number of examples is small with respect to the complexity of the network. As far as choice models are concerned, however, the amount of panel data available is usually large enough to avoid such problems. Therefore, these issues will not be addressed in this article although they represent a crucial aspect in neural network methodology.

Analogy between feedforward neural networks with Softmax outputs and the Multinomial Logit model

Just as the sigmoidal output network generalizes the Binomial Logit model, the Softmax Output network (Bridle, 1990) with shared weights generalizes the Multinomial Logit Model. The architecture of such a network is illustrated in Figure 4. The network comprises n output neurones, n being the number of considered alternatives. For each purchase occasion i, each alternative j is described by the same k attributes $(x_{ij1}, \dots, x_{ijk})$. The neurone activation functions are all linear except the ones for the output neurones, which are normalized exponentials

$$\frac{e^{\sum_k w_k x_{ijk}}}{\sum_{j=1}^{J} e^{\sum_k w_k x_{ijk}}}$$

This kind of output is called 'Softmax' because it is a continuous version of the 'all to the winner' activation function, for which the output of the neurone with the biggest input is 1, all the other outputs being 0. A Softmax output can therefore be viewed as a choice probability (Bishop, 1995).

In order to keep the coefficients of the utility function equal across alternatives during training it is necessary to keep equal the weights corresponding to the same attribute. This procedure is well known in shape recognition (Le Cun *et al.*, 1989) under the name 'Shared weights technique'. The technique consists of initializing all the weights corresponding to a same attribute with the

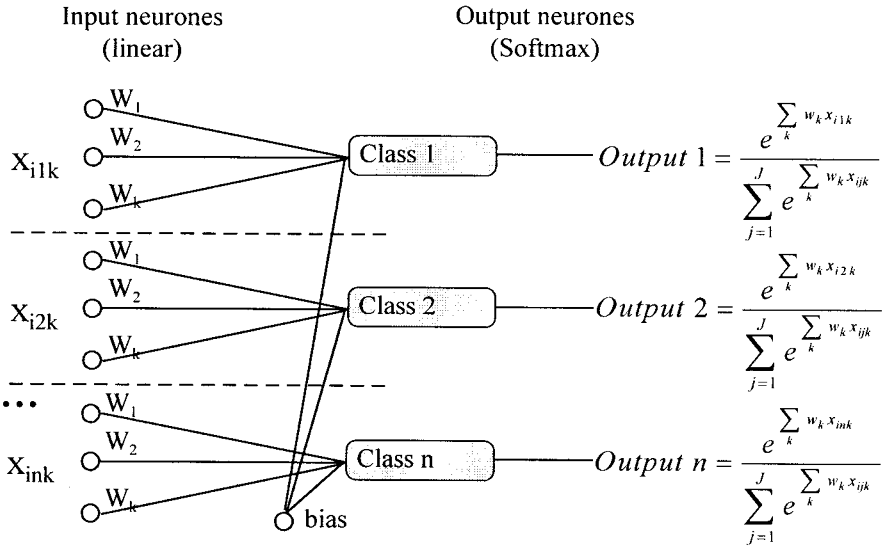

Figure 4. Neural network with Softmax outputs, partially connected and with shared weights. This model is identical to the Multinomial Logit model

same value. During training, these weights are updated with the average gradient of the error. The network is only partially connected so that the attributes of different alternatives can not interact. The formal relationships between inputs and outputs are identical for such a network and for the MNL model (equation (3)).

The used error function (relative entropy), given by equation (8), is identical (sign apart) to the log-likelihood function used to derive the Multinomial Logit (equation (5)):

$$E = -\sum_{i=1}^{M} \sum_{j=1}^{J} Y_{ij} \ln(S_{ij}) \tag{8}$$

As for the Binomial Logit, the Multinomial Logit can also be represented in the neural formalism. The MNL corresponds to a partially connected feedforward neural network with shared weights and Softmax outputs. By adding hidden neurones to the network, it is possible to model choices resulting from non-linear utility functions. Therefore, neural network models can be viewed as generalized MNL models.

Neural networks: black boxes or grey boxes?

Comparing the predictive performances of two models, one of which being a particular case of the other, tells more about the complexity of the modelled relationship than about the superiority of one model over the other. Because Artificial Neural Networks (ANNs) can be viewed as an extension of MNL modelling in cases where the utility function is non-linear, the comparison between the two models will enable us to investigate the existence of relationships taken into account by one model (ANN), but not by the other one (MNL).

Neural networks have often been criticized for their lack of interpretability. Unlike Logit modelling the analysis of network parameters does not reveal anything useful about the fitted function, except for very simple networks (e.g. perceptron with no hidden units). Indeed, information in a neural network is processed in a complete delocalized way. Furthermore, degrees of freedom are often large enough to allow the network to fit the same function with different combinations of parameters. This is probably the reason why neural networks have been called 'black boxes', capable of mimicking relationships between a set of variables but incapable of explaining the nature of these relationships. Although this critic is to a large extent justified, it is only the counter-part of the capacity of these models to approximate any function with few, if any, *a priori* assumptions.

Neural models may not provide interpretable coefficients nor do they provide significance measures (in particular, standard errors), but they can be used for scenario simulation and function visualization. If the fitted function is not too complex, it is possible to detect and identify some low order non-linear effects by projecting the function on sub-sets of the input space.

When studying consumer behaviour, there is little incentive in visualizing the relationship between choice probabilities and explanatory variables because these functions will always look more or less sigmoidal. However, it is possible to observe directly the utility function by plotting the relationship between the decision variables and the inputs of the output neurons (fitted function before passing through the Softmax function). If some non-linearities are identified by this analysis it becomes possible to respecify a Logit model in order to take these discovered effects into account. For instance, if the analysis reveals some quadratic terms in the utility function, it is possible to incorporate explicitly such effects by creating one or more additional variables representing the non-linear effects (e.g. x_1^2, x_2^2, $x_1 \cdot x_2$). The significance of these terms can then be easily tested, since their standard errors can be estimated.

Neural networks as diagnostics and specification tools can be used in combination with traditional Logit models in order to get both a better predictive performance and a greater understanding of the influences of the various product attributes. This approach is first illustrated by a synthetic example of modelling consumer behaviours in a heterogeneous market, then is applied to panel records on the instant coffee market collected on the Australian market in recent years.

HYBRID APPROACH FOR ANALYSING CONSUMER BEHAVIOUR: ILLUSTRATION OF SYNTHETIC DATA

In order to illustrate the proposed hybrid approach, let us consider the choice of chocolate brands when some attributes interact. This example rests on a synthetic dataset and may be over-simplified. However, it enables us to understand under which conditions a linear utility is inappropriate and shows clearly the consequences of a bad Logit model specification. By using synthetic data consumer preferences are perfectly known and can therefore be compared to the estimated ones computed by the various models.

Let us consider three brands of chocolate and two decision variables. The first variable, V_1, is a dummy variable indicating whether the consumers buy the product for themselves or for their children. The second variable, V_2, is the price per quantity. Let us now imagine that the consumer's sensitivity to price depends on whether the product is aimed at themselves (in which case he or she tends to buy the more expensive chocolate) or at the buyer's children (in which case

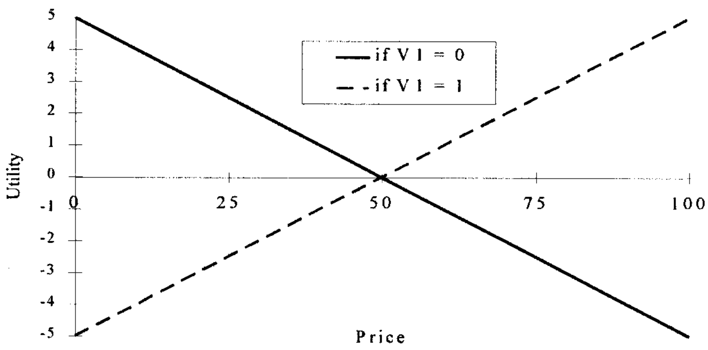

Figure 5. Graph of the utility function. This function is non-linear because of the interaction between variables V_1 and V_2

he or she buys the cheaper product, considering that children do not appreciate the quality of chocolate anyway). Hence, consumer preferences can be formalized by

$$\text{Utility} = \begin{cases} -0.1 \cdot V_2 + 5 & \text{if } V_1 = 0 \text{ (Chocolate for children)} \\ +0.1 \cdot V_2 + 5 & \text{if } V_1 = 1 \text{ (Chocolate for consumer)} \end{cases} \tag{9}$$

Clearly, V_1 and V_2 interact in this utility function. This relationship is illustrated in Figure 5. The function plotted in Figure 5 is clearly non-linear. The relationship between utility and price V_2 is conditional on the other variable V_1, although it is linear when V_1 is kept constant.

On each purchase occasion a random generator provides values for variables V_1 and V_2. The values for the utility are then computed for each alternative according to equation (9). Probabilities are calculated by equation (3). Finally, a random draw based on these probabilities is performed. On each choice occasion one alternative is selected. Both models are applied to that dataset and the results are analysed.

The modelling approach

Two models are applied to the dataset described above. The first model is a standard Multinomial Logit model and the second model is a three-layer perceptron with Softmax outputs and six neurones on the hidden layer (two neurones for each alternative). The network is partially connected and certain weights are shared. This network is represented in Figure 6.

Model performances and specification diagnostics

Several measures of quality of fit and parameter significance can be computed with artificial datasets. We shall use in this study R^2, U^2, and T-statistics for the estimated coefficients.

- **The model R^2.** R^2 compares the probability of choice with the predicted one. The R^2 measures the amount of variance (in choice probability) that is explained by the model. It is usually not available, because the actual probability of choice is never known. However, because the dataset utilized has been built rather than observed, we know the actual probability and we can compare it to the predicted one in terms of correlation or R^2. R^2 is calculated by

$$R^2 = 1 - \frac{\text{MSE}}{\text{Var}} \tag{10}$$

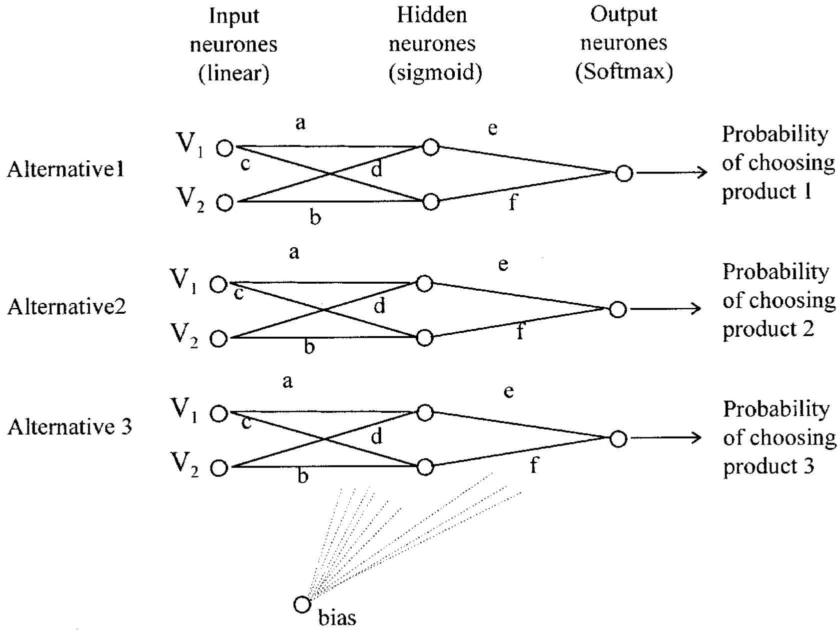

Figure 6. Perceptron with one hidden layer, partially connected. Weights having similar positions (aaa, bbb, ...) are kept equal during training. Outputs are log-linear 'Softmax' outputs

where MSE is the Mean Square Error of the model and Var is the variance of the probability to be modelled. If the model is perfect and explains accurately the probability of choice $R^2 = 1$. If the model does not map at all the relationship between dependent and independent variables $R^2 = 0$. Values of R^2 for both models are given in Table I.

Besides, it is possible to represent on a scattergraph the predicted probability versus the actual one (Figure 7). In the case of a perfect fit all the points on the graph would lie on the first diagonal. It is clear, according to Figure 7, that the MNL model has failed estimating the true probabilities. Instead the model has only been able to account for unconditional probabilities, i.e. the average market share, which is 1/3 (the three products have an equal chance of being chosen).

- **The model U^2.** U^2 measures the amount of uncertainty (i.e. the entropy) about the product choice that is explained by the model. It is expressed as the ratio of explained entropy

Table I. Comparison of performance measures for the MNL model and ANN model

Performance measures	MNL	ANN
U^2	0.049	0.52
R^2	0.002	0.845
t-stat. for variable 1 (P-value)	−0.67 (0.50)	
t-stat. for variable 2 (P-value)	0.072 (0.94)	

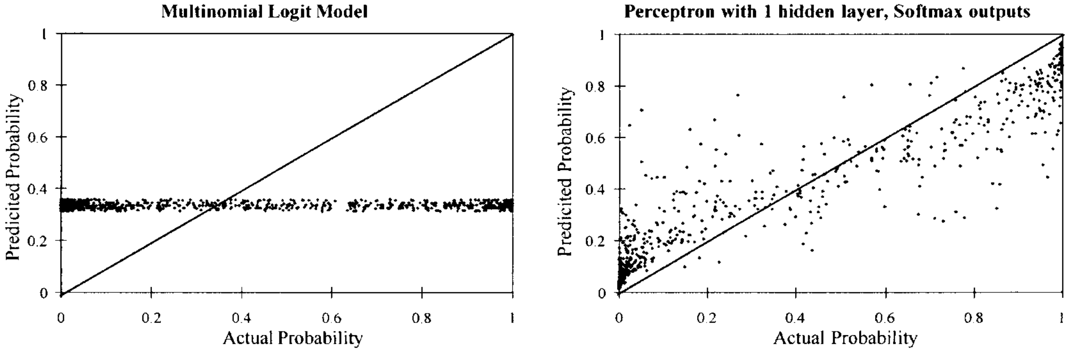

Figure 7. Comparison between actual and predicted probabilities for the two models. The first diagonal represents ideal predictions (predicted value = actual value)

to unconditional choice probability (i.e. product market share in estimation set) and is calculated by

$$U^2 = 1 - \frac{L(X)}{L_0} \tag{11}$$

where $L(X)$ is the log-likelihood of the calibrated model and L_0 is the log-likelihood of the null model. If $L(X)$ does not improve on L_0, $U^2 = 0$. If the model is perfect, $U^2 = 1$. Values of U^2 for both models are given in Table I.

- **The *T*-statistics for coefficient significance.** A coefficient T-statistic is the ratio between the value of the coefficient and the value of its standard error. In other terms, the t value measures how many standard errors the coefficient is away from 0. Large values of t (greater than 2) correspond to coefficients that are too large to be the result of chance (with a 95% confidence level). Because t-values follow a Student distribution it is possible to calculate their significance level (P-value). These values are only available for Logit models because neural models do not provide interpretable coefficients nor reliable significance measures for estimated parameters. T-values (and their significance levels) are also given in Table I.

Table I and Figure 7 suggest that the MNL model could not capture the relationship between decision variables and choice probabilities. A more general model, such as a two hidden unit neural network (designed to model simple non-linear functions), has been able to model consumer behaviour. This difference in predictive performance reveals relationships that are taken into account by the ANN but not by the MNL, i.e. non-linear relationships. This diagnostic calls for further investigations into the nature of the detected non-linearities.

Towards a better MNL model specification

Once the neural network has learned the relationship between decision variables and actual choice, it is possible to extract the estimated utility function from the model. To do so, the utility function (i.e. the input of an output neurone) is plotted against one or two decision variable(s), conditionally to the remaining variables. In the case of our artificial dataset with two inputs, it is easy to plot the utility as a function of price conditionally to V_1. Figure 8 illustrates this analysis.

Figure 8 clearly suggests that V_1 and V_2 strongly interact. The network enables us to discover the nature of the utility function (Figure 5). It is now possible to better specify the MNL model in

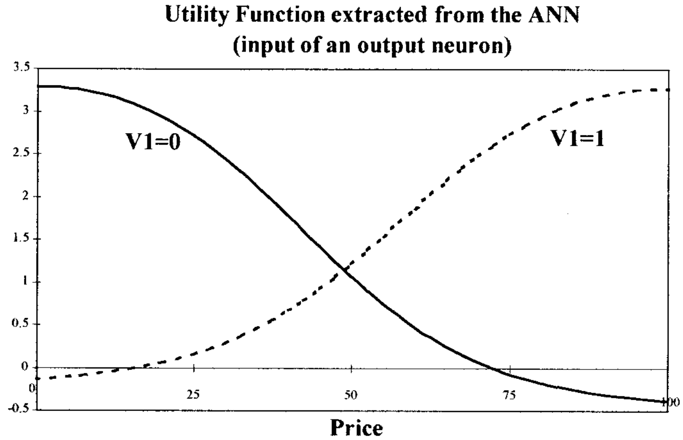

Figure 8. Graph of the utility function extracted from a partially connected ANN with two hidden units per alternative, shared weight and Softmax outputs

Table II. Performances of the respecified MNL model

Performance measures	MNL
U^2	0.059
R^2	0.99
t-stat. for variable 1 (P-value)	-10.7 (10^{-26})
t-stat. for variable 2 (P-value)	-9.9 (10^{-22})
t-stat. for variable 3 (P-value)	11.23 (10^{-28})

order to take into account the discovered effect. A new variable $V_3 = V_1 \times V_2$ is added to the dataset. This variable explicitly accounts for the discovered interaction. The predictive performance of the respecified MNL model is presented in Table II.

The respecified MNL model can estimate the interaction between V_1 and V_2. Not only are the overall model performances excellent but also the high individual t-values confirm the existence of a strong interaction between V_1 and V_2.

This example illustrates the relevance of a neural approach for the study of consumer behaviours. When the analysed market is not homogeneous but composed of segments for which behaviours differ, MNL models can underestimate or even ignore the influence of decision variables. In those circumstances it is always difficult to know whether the modelling problem arises from the model (misspecified or unsuitable model) or from the data (poor information content of the estimation dataset). A neural network model is then very useful because it enables us to know if there remains some exploitable information in the data.

To summarize, the proposed approach uses a neural network to detect the presence of potential non-linearities in the utility function. If it appears that a significant amount of non-linearity is contained in the dataset, the neural network can be used to determine the nature of the utility function (order and complexity of the function) by plotting the utility as a function of decision variables. If the non-linearities are of low order it is possible to take them explicitly into account in a MNL model. The respecified Logit model will then provide a better predictive performance, as well as useful and interpretable coefficients and significant statistics.

Now that we have illustrated the approach on a synthetic example we shall use it on some market data. The relevance of such an analysis with respect to the understanding of consumer behaviour will be analysed.

THE HYBRID APPROACH APPLIED TO SCANNER DATA

In this section the proposed hybrid approach is applied to forecasting individual purchases of instant coffee and to analysing the effects of marketing variables on consumer choices. We first describe the data used, then we study the performances of the various models and discuss the results obtained.

The data

The database consists of instant coffee store and panel records from 89 stores and five Australian towns. The data have been collected by AGB McNair Brandscan and kindly made available for this research. They cover 25,000 purchases made by about 1500 households between January 1993 and December 1994. Heterogeneity among stores and households was dealt with by data filtering in order to control for differences in product assortment across stores and to eliminate households with small purchase volumes or little store loyalty. The computation of loyalty variables and the detection of promotional price cuts imposed further constraints on the lengths of the purchase sequences. The filters used to clean the data are now described.

- **Stores:** Only stores accounting for at least 5% of the total purchase volume are kept. Six stores meet this criterion and they account for 76% of the purchases.
- **Alternatives:** The database contains information about 46 brands of instant coffee in various sizes. For the main three brands (63% market share), there are, on average, four formats that can be aggregated into two sizes. Five alternatives (brand-sizes) are eventually selected: A (Nescafé Blend 43, large), B (I/National Roast, large), C (Nescafé Blend 43, small), D (I/National Roast, small) and E (Hbrand, large).
- **Households:** Because the quality of loyalty variables computation depends on the amount of observations available per consumer only those households are kept, with at least 12 purchases in the two-year period and at least one purchase in each of the four semesters. Hence, 937 households (70% of purchases) have been selected.

The decision variables used in the model are marketing-mix variables (price per quantity, promotional price cut as a percentage of normal price), product characteristics (dummy variables specific to each alternatives) and household-specific variables (brand and size loyalties). The loyalty variables are computed by exponential smoothing of previous consumer choices (Guadagni and Little, 1983). The aim of the paper being a comparison between models, optimization of loyalty variables is not an issue here (see Fader *et al.*, 1992, for a discussion). Smoothing parameters α_1 and α_2 are only roughly estimated by optimizing the model's log-likelihood in successive trials. Values for the smoothing parameters are 0.85 for brand loyalty and 0.80 for size loyalty. Initial values for the loyalty variables are set to α_i for the first brand purchased and $(1 - \alpha_i)/(N - 1)$ for the remaining brands (Guadagni and Little, 1983). Duration of the initialisation period is arbitrarily set to 16 weeks.

After pre-processing, the database eventually is composed of 4952 purchases, covering the period from May 1993 to December 1994. Data are split into two periods. There is one calibration period for estimating model parameters (3221 purchases, May 1993 to June 1994) and there is one test period for evaluating model predictive performances on unseen data (1731 purchases, June 1994 to December 1994).

The models and performance measures

A Multinomial Logit model and a feedforward multi-layer neural network with Softmax outputs are both applied to the dataset and their performances are compared and discussed. For the neural network model the calibration period is split into a training set (90% of the data) and a cross-validation set (10% randomly chosen data) used for model selection. Model entropy (equation (8)) is the criterion used to select the optimal model complexity. Various architectures, initial connection values, training parameters and cross-validation sets have been tried. Results are remarkably stable in terms of prediction errors and fitted function. The results that are presented in the paper have been obtained with a network of four hidden neurones per alternative.

In order to measure model performance we use U^2 for in-sample performance and R^2 (percentage of variance in market shares predicted) for out-of-sample performance. Unlike where the choice probability is known, it is here impossible to compare actual and predicted choice probabilities because only actual choices (binary variable) are known. However, market shares can be viewed as aggregated individual choices and actual and predicted market shares can be compared. Hence, choice probabilities are aggregated over one week periods. In order to compare the models we use a weighted R^2 over all the alternatives:

$$R^2 = \sum_{i=1}^{5} \mathrm{MS}_i \times R_i^2 \text{ with } R_i^2 = 1 - \frac{\mathrm{MSE}_i}{\mathrm{VAR}_i} \tag{12}$$

with

MSE_i the mean square error in the market share prediction for alternative i
VAR_i the variance in actual market share for alternative i
MS_i the actual market share for alternative i.

In addition to these performance metrics, Logit models with different specifications are compared with a *likelihood ratio test*. $2 \times [\log\mathrm{L}(\mathrm{model}_2) - \log\mathrm{L}(\mathrm{model}_1)]$ follows a χ^2 distribution with n degrees of freedom, n being the difference between the models degrees of freedom. The significance of the differences in model performance can therefore be evaluated. This test can unfortunately not be used to compare MNL and ANNs because the effective degree of freedom for an ANN is unknown. However, this simple test is useful because it helps to determine whether a more complex specification is justified, in particular whether an additional variable improves significantly model performance.

Finally, variable significance is measured by the variable's t-value provided by the Logit model software package. This statistic is useful in discussing the importance of non-linearities in the utility function.

Table III. Comparison of model performances on calibration and test periods

Models	U^2 calibration	Average R^2 test	R^2 over the test period for each alternative				
			A	B	C	D	E
MNL	0.870	0.903	0.939	0.927	0.713	0.822	0.715
ANN	0.883	0.916	0.955	0.938	0.740	0.811	0.705

Comparison of predictive performances

Model performances are presented in Table III. For both models the figures are very good. The neural network, however, outperforms slightly the Multinomial Logit model. This suggests that the utility function has a weak non-linear component. It is not surprising to obtain such good results since we know for each purchase occasion that one of the alternatives is going to be bought. We do not forecast a purchase but only the choice between the five alternatives (A to E).

The performances are relatively homogeneous across alternatives. Not surprisingly, products with higher market shares seem to be better modelled. Indeed, such products have a greater influence on parameter estimation since because they are more represented in the database. Performance constancy over time is also very good. Figure 9 represents observed and predicted market shares for alternative A over the calibration and test periods.

The relationships between decision variables and choice probability seem to be very stationary. Consumer preferences do not change quickly over time. Model performance seems even to improve over time. This may be attributed to the shortness of the initialization period used for loyalty variables computation.

Utility function analysis and Multinomial Logit Model respecification

Now that we have diagnosed that the utility function is slightly non-linear, it would be interesting to identify the detected non-linear components. This is possible because the utility function can easily be extracted from the neural network by taking the signal before it enters a Softmax output

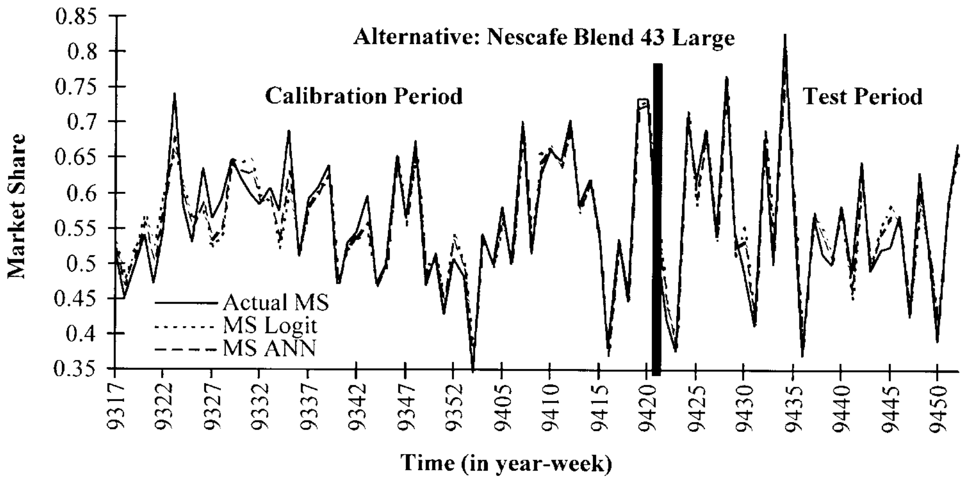

Figure 9. Actual and predicted market shares for alternative A

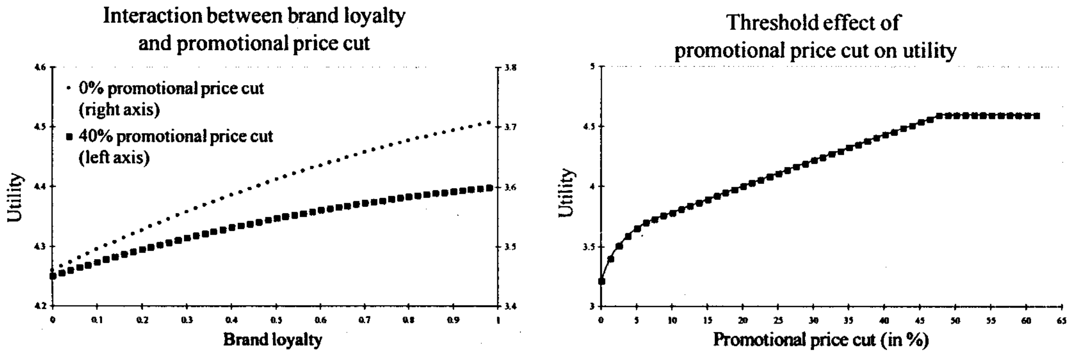

Figure 10. Non-linear components of the estimated utility function as extracted from the neural network model

unit. The relationships between decision variables and utility can then be plotted and analysed. As suspected, the utility function is mainly linear. However, three non-linear components are clearly visible. Figures 10(a) and 10(b) illustrate these non-linear effects.

The utility's non-linear components are now described and additional explanatory variables are created in order to account for these non-linear effects.

- Variable $x_9 = \textit{brand loyalty} \times \textit{promotion}$ accounts for the interaction between brand loyalty and promotional price cut. This effect is visible in Figure 10(a). The combined influence of promotional price cut and brand loyalty on utility is weaker than the sum of influences taken separately. This interaction can be viewed from two different perspectives depending on the variable considered. Viewed from a promotion perspective, it seems that a promotional price cut has less impact on the consumer when the product is his or her favourite brand than when it is not. The distance between the two curves of Figure 10(a) (representing the conditional effect of promotion) is more negative when brand loyalty is strong than when it is weak. Viewed from a brand loyalty perspective, the sensitivity (slope) of utility to brand loyalty is smaller when there is a promotion on the considered brand. According to this analysis the coefficient for variable x_9 should be negative in a Logit model.
- Variable $x_{10} = (\textit{brand loyalty})^2$ accounts for the concave relationship between utility and brand loyalty. This effect is visible in Figure 10(a). Utility increases strongly with brand loyalty when loyalty levels are low or medium (typically < 0.7) and increases more slowly for larger levels of loyalty. Here the sensitivity of utility to brand loyalty is conditional to the level of brand loyalty itself. This effect can be explained by variety seeking behaviours in the product range. A household purchases on average 1.23 brands (there are three brands in our database) over the calibration period (May 1993 to June 1994). To account for this variety-seeking behaviour, the coefficient for variable x_{10} should be negative in a Logit model.
- Variable

$$x_{11} = \begin{cases} 1 & \text{if promotion} \neq 0\% \\ 0 & \text{if promotion} = 0\% \end{cases}$$

accounts for the threshold effect of promotional price cut. This effect is visible in Figure 10(b). In addition to the linear dependence of utility to promotional price cut, the mere existence of a promotion seems to have an effect on utility. This is not surprising since a promotional price cut is not just about price. Consumers seem to be sensitive not only to price cuts but also to the promotional activity itself which involves also higher feature and display activity such as end-aisle display, small shelf-talker or advertising. Our database does not contain information about feature or display activity for the stores and therefore this non-linear effect accounts for attributes that are not explicitly represented in the data. It is also interesting to analyse the saturation effect on the right-hand side of the graph in Figure 10(b). This effect illustrates the influence of leverage points in the dataset. Promotional price cuts over 50% represent only 0.25% of the purchases. This region of the input space is very sparse compared to other regions, so that the training in this region becomes hazardous and the estimated function unreliable.

Five new specifications for the Logit model are now attempted. These new specifications are designed to take into account the identified non-linear components of the utility function.

Specification a: Original model + variable x_9
Specification b: Original model + variable x_{10}
Specification c: Original model + variable x_{11}
Specification d: Original model + variables x_9 and x_{10}
Specification e: Original model + variables x_9, x_{10} and x_{10}

Results from respecified MNL models

The modelling results for the respecified Logit models are presented in Table IV. All coefficients for the new variables have the sign expected, i.e. x_9 and x_{10} are negative and x_{11} is positive. Unfortunately, models have become more difficult to interpret because the influence of an attribute is now represented by several coefficients. For instance, for the '*b*' specification the influence of brand loyalty is represented by two variables, i.e. *x*- and x_9, which are strongly collinear. The significance (or *t*-value) of brand loyalty is also shared between two coefficients and therefore cannot be compared to the significance of an attribute represented by only one variable (for instance, depromoted price).

Table IV. Performances for respecified Logit models (non-linear utility function)

Model specifications		O	a	b	c	d	e
Performance measures	U^2 (calibration)	0.8701	0.8708	0.8709	0.8719	0.8717	0.8737
	χ^2	0	7.66	8.34	19.16	16.30	36.58
	(significance level)		(5.6 10^{-3})	(3.9 10^{-3})	(1.2 10^{-5})	(2.9 10^{-4})	(5.6 10^{-8})
	R^2 over test set	0.9027	0.9005	0.9003	0.9065	0.9025	0.9101
t-statistics for model variables	Depromoted price	−6.4	−6.4	−6.4	−6.2	−6.4	−6.3
	Price cut	6.2	6.6	6.1	0.1	6.6	1.8
	Brand loyalty	30.3	27.4	6.5	30.0	6.6	6.6
	Size loyalty	12.7	12.7	12.7	12.7	12.7	12.7
	x_9		−2.9			−2.9	−3.3
	x_{10}			−2.7		−2.8	−2.7
	x_{11}				4.3		4.5
t-statistics for model constants	A	6.8	6.4	6.4	6.4	6.1	5.7
	B	5.0	4.6	4.8	4.5	4.5	3.9
	C	−7.0	−7.1	−7.1	−7.2	−7.2	−7.4
	D	−8.2	−8.3	−8.3	−8.4	−8.4	−8.5

Critical values for *t*-statistics are: 1.645 (sig. level = 10%), 1.960 (s.l. = 5%), 2.576 (s.l. = 1%), 3.290 (s.l. = 0.1%).

Model performances are marginally improved by the explicit representation of non-linear components in the utility function. Although this improvement is small (U^2 increases from 0.8701 to 0.8737), it is significant for all respecifications (χ^2 are significant at a 0.01 significance level), in particular for the model taking into account all three detected non-linearities (χ^2 is significant at a 10^{-7} level). This improvement can also be observed on the test period (R^2 increases from 90.27% to 91.01%). However, although the new specifications account for some non-linearities, there are still some deterministic components that are not taken into account by the Logit model as the difference in performances (U^2, R^2) between the best Logit model and the neural network model would suggest (0.883 versus 0.8737 for U^2, 0.916 versus 0.910 for R^2). The remaining non-linearities are of higher order and their detection is tricky since they probably involve several decision variables.

The *t*-values for variables x_9, x_{10} and x_{11} confirm the conclusions about the nature of the non-linear components that were drawn from the neural analysis. All the *t*-values are significant at a 0.01 level. The significance of variable x_{11} is stronger than the neural network suggests. This binary variable seems to concentrate all the influences of promotions, leaving virtually no variance left for x_6 (promotional price cut) to explain. This does not necessarily mean that a 30% promotional price cut has the same effect as a 10% price cut because it is probable that a binary variable better accounts for other promotional activities (such as feature and display activities) that accompany the promotion but for which no information is available in the database.

T-statistics for alternative specific constants are remarkably stable across specifications. This suggests that the influence of these constants (accounting for unconditional market shares) could not be attributed to the new additional variables. It is not surprising that the 'non-linear' variables are independent of the alternative constants because there is no reason why a non-linear component would be specific to a particular alternative. Likewise, there seems to be no major co-linearity problems between the new variables, depromoted price and brand loyalty.

Loyalty variables seem very important as determinants of the utility function. With *t*-values over 30, 12 past purchases are excellent predictors for future choices. This high autocorrelation in purchases explains mainly the very good model performances and is not uncommon in studies on consumer choices (Guadagni and Little, 1983). Three comments can be made about this feature. First, instant coffee seems to be a market where brand and size loyalties are important. Hence, in our database households purchase on average 1.2 brands (there are three brands in our database). Second, the importance of the loyalty variables can be partly attributed to pre-processing biases, in particular, filters used to select alternatives and households. The sample used for modelling is not strictly random since data constraints were imposed for homogeneity reasons and loyalty variable computation. These constraints are related to purchasing behaviour (for practical reasons only households and alternatives representing a large share of the purchases were selected). Households who often change brands and stores are therefore under-represented. Third, loyalty variables do not explain heterogeneity among households but merely measure it. It would therefore be interesting in further stages of the study to use other variables (for instance, demographic attributes) to explain households' heterogeneity.

CONCLUSION

Logistic regression, as well as a Multinomial Logit model (MNL), can be represented in a neural formalism. The neural equivalent of a MNL is a perceptron without hidden units, partially connected, with Softmax outputs and shared weights and entropy as cost function. By adding hidden units to this perceptron, it is possible to generalize the MNL model for cases where the utility function is non-linear (threshold and interaction effects between the decision variables). It only makes sense to use such a model when it is justified by the complexity in the data. It is therefore impossible to conclude that one model is superior to the other, because this superiority depends on the complexity of the relationships between the various variables in the database.

The main difference between the two models is that the neural network does not require any *a priori* knowledge about the relationships to be modelled, whereas the Multinomial Logit suffers from a specification bias. An inappropriate model specification leads to under-estimating, when not completely ignoring, the influence of the various decision variables. Therefore, it is often difficult to know whether a poor predictive performance should be blamed on the data or on the model. Because neural networks do not suffer from such a specification bias they can be used as diagnostics. The bad news is that, unlike coefficients of parametric models, neural network parameters cannot be interpreted. The two families of models are therefore complementary and would both gain by being combined into a hybrid approach.

In such an approach a neural network model would be used first, to diagnose non-linear utility functions and second, to determine the nature of the non-linear components of the function. A Multinomial Logit model can then be specified so that it takes explicitly into account the detected non-linearities. Statistical tests can be used to measure the significance of the various effects. This approach was illustrated in a synthetic example of a heterogeneous market and was applied to market data of instant coffee purchases.

On our database, the neural network model slightly outperforms the Multinomial Logit model in terms of explanatory and predictive power. This suggests that the modelled consumer utility function has some non-linear components and that the specification for the MNL model is sub-optimal. It has been shown how the fitted utility function can be extracted from the neural model, how it can then be analysed to discover the nature of the non-linearities and how these non-linear effects can explicitly be taken into account by a respecified MNL model. In our dataset, three non-linear components have been detected and analysed and three additional variables have been created to account for these effects. Once respecified, the MNL model performance was significantly improved not only on the calibration set but also on the test set. In addition, the significance of the non-linear effects has been measured and discussed.

The complexity of our dataset probably did not justify the use of a neural network model. Indeed, the marginal improvement in predictive performance suggested the presence of very weak non-linear components in the utility function. However, the study of this database has clearly shown that the proposed hybrid approach is relevant in cases where decision variables interact. Such interaction detection and modelling suggest an interesting research direction in the understanding of consumer choices at an individual level. The hybrid model can be particularly relevant when we go beyond the traditional representation of household heterogeneity by loyalty variables and product idiosyncrasies by dummy variables. One can imagine not merely representing such heterogeneities, but also explaining them with socio-demographic variables and understanding their interactions with product attributes. In such perspectives, products are no longer described by a single dummy variable but rather by a series of variables describing relevant products characteristics in the choice context (brand, package size and type, flavour, product form, formula, odour . . .) (Fader and Hardie, 1996).

ACKNOWLEDGEMENTS

The authors gratefully thank AGB McNair Brandscan for providing the scanner panel data on which this work is done. We also like to thank Bruce G. S. Hardie, from the London Business School, for his useful comments about this study.

REFERENCES

Andrews, R. L. and Srinivasan, T. C., 'Studying consideration effects in empirical choice models using scanner panel data', *Journal of Marketing Research*, **32** (1995), 30–41.

Baldi, P. and Hornik, K., 'Neural networks and principal component analysis: learning from examples without local minima', *Neural Networks*, **2**(1) (1989), 53–8.

Ben Akiva, M. and Lerman, S., *Discrete Choice Analysis*, Cambridge, MA: MIT Press, 1985.

Bentz, Y. and Merunka, M., 'Modelling brand choice with the Multinomial Logit model and neural networks: a comparison of methods and results', in *Actes des Deuxièmes Rencontres de la Recherche Neuronale en Sciences Economiques et de Gestion, Université de Poitiers*, France, 1995, pp. 105–24.

Bishop, C. M., *Neural Networks for Pattern Recognition*, Oxford: Oxford University Press, 1995.

Bridle, J., 'Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition', in *Neuro-Computing: Algorithms, Architectures and Applications*, New York: Springer, 1990.

Burgess, A. N., 'How neural networks can improve database marketing', *The Journal of Database Marketing*, **12** (1995), 312–27.

Burgess, A. N., 'Loan risk analysis using neural networks', *The Journal of Targeting, Measurement and Analysis for Marketing*, **4** (1995), 24–37.

Desmet, P., 'Apport des réseaux de neurônes pour le scoring en marketing direct', in *Actes des Deuxièmes Recontres de la Recherche Neuronale en Sciences Economiques et de Gestion, Université de Poiters*, France, 1995.

Fader, P. S. and Hardie, B., 'Modeling consumer choice among SKU's', Marketing Working paper No. 95-603, London Business School, 1995.

Fader, P. S., Lattin, J. M. and Little, J. D. C., 'Estimating nonlinear parameters in the Multinomial Logit model', *Marketing Science*, **11**(4) (1992), 372–85.

Friedman, J. H., 'Multivariate adaptative regression splines', *Ann. Statist.*, **19** (1991), 1–141.

Gensch, D. and Recker, W., 'The Multinomial Multiattribute Logit choice model', *Journal of Marketing Research*, **16**(1) (1979), 124–32.

Guadagni, P. M. and Little, J. D. C., 'A logit model of brand choice calibrated on scanner data', *Marketing Science*, **2**(3) (1983), 203–38.

Härdle, W., *Applied Nonparametric Regression*, Cambridge: Cambridge University Press, 1990.

Hastie, T., Buja, A. and Tibshirani, R., 'Penalized discriminant analysis', *Annals of Statistics 1995*, **23**(1) (1995), 73–102.

Hastie, T. and Tibshirani, R., *Generalized Additive Models*, London: Chapman & Hall, 1990.

Hastie, T., Tibshirani, R. and Buja, A., 'Flexible discriminant analysis by optimal scoring', *Journal of the American Statistical Association*, **89**(428) (1994), 1255–70.

Hauser, J. R., 'Testing the accuracy, usefulness, and significance of probabilistic choice models: an information theoretic approach', *Operations Research*, **26** (1978), 406–21.

Hertz, J., Krogh, A. and Palmer, R., *Introduction to the Theory of Neural Computation*, Reading, MA: Addison-Wesley, 1991.

Hopfield, J. J., 'Learning algorithms and probability distributions in feed-forward and feed-back networks', In *Proceedings of the National Academy of Sciences*, **84** (1987), 8429–33.

Hornik, K., Stinchcombe, M. and White, H., 'Multilayer feedforward networks are universal approximators', *Neural Networks*, **2** (1989), 359–66.

Hruschka, H., 'Determining market response functions by neural network modelling: A comparison to econometric techniques', *European Journal of Operational Research*, **66** (1993), 27–35.

Hruschka, H. and Natter, M., 'Clustering-based market segmentation using neural networks', Working Paper, revised version, March, 1994.
Kohonen, T., *Self-Organizing and Associative Memory*, 2nd edition, New York: Springer-Verlag, 1988.
Kumar, A., Rao, V. R. and Soni, H., 'An empirical comparison of neural networks and logistic regression models', *Marketing Letters*, **6**(4) (1995), 251–63.
Lattin, J. M., 'A model of balanced choice behavior', *Marketing Science*, **6**(1) (1987), 48–65.
Lattin, J. M., and Bucklin, E. 'Reference effects of price and promotion on brand choice behavior', *Journal of Marketing Research*, **26** (1989), 299–310.
Le Cun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. and Jackel, L. D., 'Backpropagation applied to handwritten zip code recognition', *Neural Computation*, **1**(4) (1989), 541–51.
MacFadden, D., 'Conditional logit analysis of qualitative choice behavior', in Zarembka, P. (ed.), *Frontiers of Econometrics*, New York: Academic Press, 1974.
MacFadden, D., Tye, W. and Train, D., 'An application of diagnostic tests for the independance of irrelevant alternatives property of the Multinomial Logit model', *Transportation Research Record*, **637** (1977), 39–45.
MacFadden, D., 'Econometric models of probabilistic choice', in Manski, C. F. and MacFadden, D. (eds), *Structural Analysis of Discrete Data with Economic Application*, MIT Press, 1981.
Malhotra, N., 'The use of linear Logit models in marketing research', *Journal of Marketing Research*, **23**(1) (1984), 20–31.
Meyer, R. and Kahn, B., 'Probabilistics models of consumer choice behavior', In *Handbook of Consumer Behavior*, Englewood Cliffs, NJ: Prentice Hall, 1991, pp. 85–123.
Refenes, A. N., *Neural Networks in the Capital Markets*, Chichester: Wiley Finance, 1995.
Ripley, B., 'Neural networks and related methods for classification', *Journal of the Royal Statistical Society*, **56**(3) (1994), 409–56.
Rumelhart, D. E. and McClelland, J. L., *Parallel Distributed Processing*, Cambridge, MA: MIT Press, 1986.
Shocker, A. D., Ben-Akiva, M., Boccara, B. and Nedungadi, P., 'Consideration set influences on consumer decision-making and choice: issues, models and suggestions', *Marketing Letters*, **2** (August 1991), 181–98.
Siddarth, S., Bucklin, R. E. and Morrison, D. G., 'Making the cut: modeling and analyzing choice set restriction in scanner panel data', *Journal of Marketing Research*, **32** (1995), 255–66.
Tellis, G. J., 'Advertising exposure, loyalty and brand purchase: a two-stage model of choice', *Journal of Marketing Research*, **25** (1988), 134–44.
Trippi, R. and Turban, E., *Neural Networks in Finance and Investing*, New York: Probus Publishing, 1992.
Vanhonacker, W., 'A new brand choice model incorporating a choice set formation process', Working Paper MKTG 94.001, School of Business and Management, The Hong Kong University of Science and Technology, 1993.

Authors' biographies:
Yves Bentz is a PhD student at London Business School. His research interests are in non-linear modeling, in particular stochastic parameter models and non-parametric statistics. Application domains are essentially Equity Investment Management and consumer choice.

Dwight Merunka holds a PhD in Marketing from Aix-en-Provence University. His research interests are in the fields of consumer choice and market modeling. After having been at Paris-Dauphine University he is now Professor and Associate Dean for Research at Aix Graduate School of Management, France.

Authors' addresses:
Yves Bentz, London Business School, Sussex Place, Regent's Park, London NW1 4SA, UK.

Dwight Merunka, IAE Aix-en-Provence, Le Clos Cuiot, 13540 Puyricard, France.

