

**MODELING CONSIDERATION SETS AND BRAND CHOICE USING
ARTIFICIAL NEURAL NETWORKS**

BJÖRN VROOMEN, PHILIP HANS FRANSES, ERJEN VAN NIEROP

ERIM REPORT SERIES <i>RESEARCH IN MANAGEMENT</i>	
ERIM Report Series reference number	ERS-2001-10-MKT
Publication	March 2001
Number of pages	19
Email address first author	bvroomen@few.eur.nl
Address	Erasmus Research Institute of Management (ERIM) Erasmus Universiteit Rotterdam PoBox 1738 3000 DR Rotterdam, The Netherlands Phone: # 31-(0) 10-408 1182 Fax: # 31-(0) 10-408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:
www.erim.eur.nl

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES *RESEARCH IN MANAGEMENT*

BIBLIOGRAPHIC DATA AND CLASSIFICATIONS		
Abstract	<p>The concept of consideration sets makes brand choice a two-step process. House-holds first construct a consideration set which not necessarily includes all available brands and conditional on this set they make a final choice. In this paper we put forward a parametric econometric model for this two-step process, where we take into account that consideration sets usually are not observed. It turns out that our model is an artificial neural network, where the consideration set corresponds with the hidden layer. We discuss representation, parameter estimation and inference.</p> <p>We illustrate our model for the choice between six detergent brands and show that the model improves upon a one-step multinomial logit model, in terms of fit and out-of-sample forecasting.</p>	
Library of Congress Classification (LCC)	5001-6182	Business
	5410-5417.5	Marketing
	HF 6161.B4	Brand name: Advertising
Journal of Economic Literature (JEL)	M	Business Administration and Business Economics
	M 31	Marketing
	C 44	Statistical Decision Theory
	M 31 C 45	Marketing Neural Networks and Related Topics
European Business Schools Library Group (EBSLG)	85 A	Business General
	280 G 255 A	Managing the marketing function Decision theory (general)
	290 D	Branding
Gemeenschappelijke Onderwerpsontsluiting (GOO)		
Classification GOO	85.00	Bedrijfskunde, Organiseatiekunde: algemeen
	85.40	Marketing
	85.03	Methoden en technieken, operations research
	85.40	Marketing
	85.03	Methoden en technieken, operations research
Keywords GOO	Bedrijfskunde / Bedrijfseconomie	
	Marketing / Besliskunde	
	Merken, Econometrische modellen, Neurale netwerken	
Free keywords	consideration set, brand choice, artificial neural network.	
Other information		

Modeling consideration sets and brand choice using artificial neural networks*

Björn Vroomen[†]

*Erasmus Institute of Research in Management,
Erasmus University Rotterdam*

Philip Hans Franses

*Econometric Institute and Department of Marketing and Organization,
Erasmus University Rotterdam*

Erjen van Nierop

*Tinbergen Institute,
Erasmus University Rotterdam*

July 31, 2000

*This paper emerges from the M.Sc. thesis of the first author.

[†]Correspondence to Björn Vroomen, Erasmus Institute of Research in Management, Erasmus University Rotterdam, Office H16-12, P.O. Box 1738, NL-3000 DR Rotterdam, the Netherlands, email: bvroomen@few.eur.nl.

Modeling consideration sets and brand choice using artificial neural networks

Abstract

The concept of consideration sets makes brand choice a two-step process. Households first construct a consideration set which not necessarily includes all available brands and conditional on this set they make a final choice. In this paper we put forward a parametric econometric model for this two-step process, where we take into account that consideration sets usually are not observed. It turns out that our model is an artificial neural network, where the consideration set corresponds with the hidden layer. We discuss representation, parameter estimation and inference. We illustrate our model for the choice between six detergent brands and show that the model improves upon a one-step multinomial logit model, in terms of fit and out-of-sample forecasting.

Keywords: consideration set, brand choice, artificial neural network.

1 Introduction

Modeling brand choice is an important topic in marketing research. Due to the increasing amount of household-specific scanner data, marketing researchers should be able to make better predictions and generate improved explanations of consumer behavior. To describe the process of brand choice, various models have been developed and occasionally used in practice. Many of these brand choice models are based on a multinomial logit model [MNL] (McFadden, 1973, Guadagni and Little, 1983, Lattin and Bucklin, 1989, among others) or a multinomial probit model (Hausman and Wise, 1978 and Daganzo, 1979). A key property of these models is that they assume households to consider the full set of available brands.

The assumption that all brands are considered may not always be realistic. This naturally leads to the concept of consideration sets, that is, before making a choice households may reduce the number of brands to a smaller set called the consideration set. The consideration set thus contains only those brands that a household considers to buy. With the concept of a consideration set, it is implicitly assumed that the process of making a decision concerning brand choice is a two-step process. First, there is a reduction of the available set of brands. Second, from the resulting brands a final choice is to be made. In terms of a descriptive quantitative model, the resultant models for brand choice are therefore two-stage models. Note, that consideration sets typically have not been observed and hence have to be estimated from the data.

In this paper, we present a new parametric econometric model for the two-step process of brand choice, that is, we incorporate the formation of consideration sets. The basic premise in this paper is that we abstain from household-specific unobserved heterogeneity, and for the moment we consider an extension towards such additional flexibility as an interesting issue for further research. The model we propose for brand choice is shown to have the structure of an artificial neural network [ANN].

The use of ANNs is now quite common in economics and business and these models are used to describe and forecast many important variables. For example, they have been used to predict bankruptcy (Zhang *et al.*, 1999) and to highlight structural changes in time series data (Franses and Draisma, 1997; Franses and van Dijk, 2000). See Vellido *et al.* (1999), for a survey of applications of ANNs in business. A recent example of an ANN in marketing concerns forecasting market shares of grocery products, see Agrawal and Schorling (1996). Although ANNs are valuable tools for empirical modeling, a well-known drawback of ANNs concerns the difficulties when interpreting the parameters. There are ways to overcome this drawback and in this paper

we will focus on such an approach.

In marketing research, there is some evidence that an ANN yields more useful insights than a (logistic) regression-based model. For example, West *et al.* (1997) compare an ANN with discriminant analysis and logistic regression. They conclude that an ANN can outperform the two statistical techniques when the underlying choice rule is known and can give better out-of-sample forecasts when the choice rule is not known. Dasgupta *et al.* (1994) compare the same two statistical techniques with an ANN. They conclude that the superiority of the ANN is not statistically significant. Furthermore, they argue that the performance of an ANN is likely to be sensitive to the data used and perhaps even to the type of application being considered. Kumar *et al.* (1995) arrive at a similar conclusion. They conclude that it even depends on the purpose of the model, which technique to choose. Finally, Bentz and Merunka (2000) give a method to combine an ANN and a MNL model into a hybrid approach.

In sum, artificial neural networks are applied in marketing research, though with mixed results. To our knowledge, there have been some attempts so far to link ANNs with actual choice behavior, but none of these explicitly incorporate the intermediate step of consideration set formation. In this paper we address the interpretation aspects of an ANN by demonstrating that such a model naturally arises from a two-step brand choice model involving consideration sets, at least when the consideration set stage is assumed to correspond with the hidden layer. We illustrate our model for the choice between six liquid detergent brands and find that the model does not only fit well but also generates more accurate out-of-sample forecasts than a one-step model does.

The outline of this paper is as follows. In Section 2, we present a further but brief elaboration of the concept of consideration sets. In Section 3 we propose our model, discuss how it can be evaluated and how the parameters can be estimated. In Section 4, we illustrate our model on liquid detergent data. Finally, we conclude our paper with a discussion.

2 Consideration sets

The assumption that households can reduce the total number of alternative brands into a consideration set before making a final choice has recently raised much interest in theoretical and empirical marketing research. In this section we briefly review the relevant literature.

2.1 Theory

The concept of consideration sets can be useful for understanding the brand choice process. The theory of consideration sets first assumes a *universal set* containing all available brands. This set may contain unobtainable or irrelevant brands. Next, it is assumed that households select brands from the universal set that are appropriate for current goals and of which a household has some knowledge. This selection leads to an *awareness set*. A further reduction of the available brands is assumed to yield the *consideration set*, and this set only consists of those brands which a household evaluates prior to making the final choice.

Consideration sets are assumed to be constructed at each decision occasion and hence they may differ across occasions. Due to this dynamic property of the consideration set it is sometimes useful to define a closely related more static set, which is called the *choice set* (see Shocker *et al.*, 1991). The choice set consists of the brands considered immediately prior to the final choice. We refer to Roberts and Lattin (1991), Shocker *et al.* (1991), and Horowitz and Louviere (1995) for detailed discussions on the theory of consideration sets and for important surveys of the relevant literature.

2.2 Empirical results

The assumption that the brand choice process is a multi-level process is rather appealing, at least from a theoretical point of view. From a practical point of view however, the concept involving awareness sets and consideration sets is a little bit less appealing, as in many situations one cannot observe these sets. Only in cases of experimental surveys, the market researcher may obtain information on the various sets as constructed by households. A drawback of surveys is that they are expensive and not widely available. An additional disadvantage of survey data is that it is not straightforward to forecast the consideration sets.

On the other hand, if one only has scanner data, involving individual-specific and longitudinal information on purchases, one has no direct information on the composition of the consideration set. By treating the consideration set as a stochastic variable, which is somewhere in between all available brands and the selected brand, one can try to estimate this set. As the sets, discussed in the previous subsection, are unobserved most empirical work assumes that only one of these possible sets is to be included in the model. Chiang *et al.* (1999), for example, use scanner data to simulate probabilities that a certain consideration set is formed. Based on these probabilities, the final choice can be described.

Households can construct a consideration set in different ways. For example, the formation of the set can be memory-based or stimulus-based. Further, an assumption on the way households create consideration sets enlarges the number of different two-stage models. For example, the consideration set can be assumed constant or time varying over subsequent purchase occasions. Manrai and Andrews (1998) discuss these possibilities (among other things) and give an extensive review of the performance of two-stage models in comparison with a one-stage MNL model. They conclude that a model incorporating consideration set formation, tends to give better forecasts than a MNL model. Although the two-stage models may perform better, Manrai and Andrews (1998) argue that this does not automatically mean that the set formation process is described correctly. Moreover, Horowitz and Louviere (1995) indicate that in case the final choice model is not properly specified the use of a consideration set yields no extra information. Roberts and Lattin (1997) provide a recent review of consideration sets.

In sum, although consideration sets cannot be observed directly, empirical arguments can be found that support the concept. A brief review of the literature suggests that many two-stage brand choice models have been developed and that they can be more useful in practice than a one-stage MNL model.

3 Consideration sets, brand choice and a neural network

As discussed in the previous section, a two-stage model can be more useful for modeling brand choice than a one-stage model. In this section we present a new parametric two-stage model, which is based on the decision making process of households, and which assumes the unobserved process of consideration set formation. In Sections 3.1 and 3.2 we indicate that this parametric model has strong similarities with an ANN. This is because we assume that the unobserved hidden layer in an ANN corresponds with such a consideration set. In Section 3.3 and 3.4 we elaborate on interpretation and on a method for parameter estimation.

3.1 Graphical representation

During the two stages of a decision making process a household needs to form an attitude towards a brand in order to determine whether that particular brand is going to be considered or even purchased. Such attitude formation

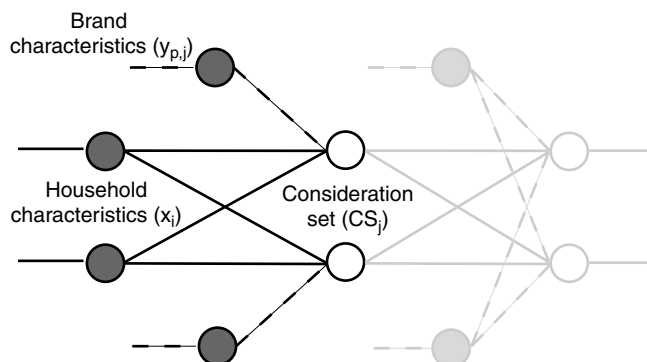


Figure 1: Formation of the consideration set for two brands based on household-specific (x_i) and brand-specific $(y_{p,j})$ characteristics.

depends on two kinds of characteristics. First, household-specific characteristics such as demographic and social factors, for example, the size of the household or the average inter-purchase time can influence attitudes towards brands, and thereby the potential consideration of a brand. Second, brand characteristics like price, promotion, and advertising are of importance.

Schematically the attitude formation, resulting in consideration set formation, is illustrated by the highlighted part in Figure 1. We denote by CS_j , a 1/0 node that brand j , $j = 1, 2, \dots, J$, is included in a consideration set of a household (1) or not (0). For each brand j , there are $p = 1, 2, \dots, P$ possible characteristics $y_{p,j}$, which can help to determine the value of CS_j . Also for every brand j there are $i = 1, 2, \dots, I$ household-specific characteristics x_i , which influences consideration set formation. We assume that the brand characteristics $y_{p,j}$, like price and promotion, only influence brand j , while the variables x_i can have an effect on each brand. In Figure 1 we graphically display these assumptions for two brands.

The second step which goes from consideration set to final choice, is illustrated by the highlighted part in Figure 2, again for two brands. The final choice for brand j , $j = 1, 2, \dots, J$ denoted as the 1/0 binary variable FC_j , is determined by the outcome of the consideration set formation stage (CS_j) and by the choice-specific characteristics z_q , $q = 1, 2, \dots, Q$. In our illustration below we have that Q equals 2. We assume that both CS_j and z_q have an effect on FC_j .

As depicted in Figures 1 and 2 the two stages together constitute a larger structure, that is, the whole decision making process. Interestingly, the structure of this process appears to have the familiar structure of an feed-forward artificial neural network. The resulting network consists of three layers. Be-

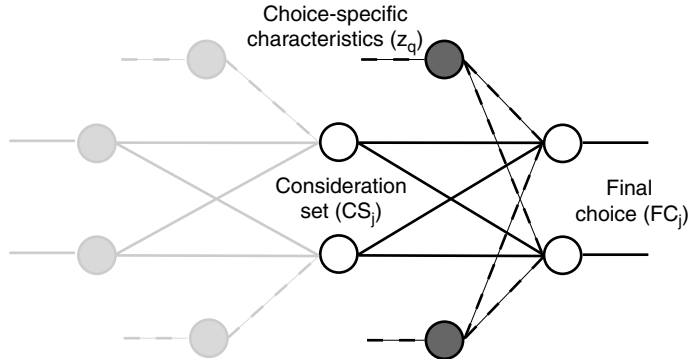


Figure 2: Determination of final choice between two brands, based on the consideration set and extra information

cause we aim to describe a brand choice process including a consideration set, the second (hidden) layer corresponds with this consideration set. Household-specific and brand-specific characteristics are used as input (the first layer) and the output (third) layer indicates which brand is chosen. A special property of our network is that the network is not fully connected (that is, some effects are restricted to zero). Additionally, the output layer is allowed to benefit from extra information.

To make the connection between the model in Figure 1 and 2 and an ANN more specific, we relax the assumption that the hidden layer contains nodes that give a value of 1 when the brand is considered and a value of 0 when not. Instead, the nodes in this layer indicate the probability that the brand is considered. This again results in a hidden layer with as many nodes as there are brands. Nodes in the output layer indicate the choice made from all available brands and hence this layer also contains as many nodes as there are brands. Additional to a flexible value of CS_j , we also assume that nodes in the output layer give the probability that the corresponding brand is chosen. Finally, we assume that the brand with the highest probability (FC_j) is chosen.

3.2 Parametric representation

In this section we transform the graphical model proposed in Figures 1 and 2 into an econometric model. When one is familiar with the underlying structure of an ANN, it is not difficult to see that the (overall) structure in Figures 1 and 2 already suggest such a model. Each of the J brands in the hidden layer gets represented by a sigmoidal node CS_j which gives the

probability that the corresponding brand is considered. This probability is determined according to

$$CS_j = F(\alpha_{0,j} + \sum_{i=1}^I \alpha_{i,j} x_i + \sum_{p=1}^P \gamma_{p,j} y_{p,j}) \quad j = 1, 2, \dots, J \quad (1)$$

We take a weighted sum for each of the J brands, of household-specific variables (x_i) and brand-specific variables ($y_{p,j}$). The weights are denoted by $\alpha_{i,j}$ and $\gamma_{p,j}$, respectively. Further, a constant $\alpha_{0,j}$ is added. This weighted sum is scaled to the $[0,1]$ -interval using the logistic function $F(v) = (1 + \exp(-v))^{-1}$.

The second stage, as illustrated by the highlighted part in Figure 2 can be interpreted as a classification problem where the classes are mutually exclusive. It is now desired that the nodes in the output layer give the choice probabilities for the J brands, and hence that the output of these nodes should lie in the $[0,1]$ -interval and they should sum up to unity. To achieve this, the probability of final choice (FC_j) is determined using

$$FC_j = \frac{\exp(net_j)}{\sum_{l=1}^J \exp(net_l)} \quad j = 1, 2, \dots, J \quad (2)$$

where net_j denotes a weighted sum of the inputs to node j . This equation is a generalization of the logistic sigmoid activation function, also known as the normalized exponential or softmax activation function. Upon using this softmax function, the nodes in the output layer can be interpreted as probabilities conditional on the output of the hidden nodes, see Bishop (1995).

The net input of node j in (2) is determined by a weighted sum of the outputs of the consideration layer (CS_j) and extra information (z_q), that is,

$$net_j = \beta_{0,j} + \sum_{k=1}^J \beta_{k,j} CS_k + \sum_{q=1}^Q \delta_{q,j} z_q \quad j = 1, 2, \dots, J \quad (3)$$

where the weights are denoted by $\beta_{k,j}$ and $\delta_{q,j}$, respectively. Again a constant $\beta_{0,j}$ is added. Note that the extra information in z_q may include information already used while determining the probabilities of consideration. In fact, no adverse consequences are expected from including the same variables for both the determination of consideration sets and the final choice, see also Andrews and Manrai (1998).

Combining (1) to (3) gives our two-stage model for describing brand choice. The model first determines for each brand the probability that it is considered, based on household-specific and brand-specific characteristics.

Subsequently, based on these probabilities and on extra information, the finally selected brand is determined. The model contains $J^2 + J \cdot (I + P + Q + 2)$ parameters. For example, if $J = 4$, $I = 2$, $P = 3$ and $Q = 1$, the model has 48 parameters. The number of parameters is quadratic in J , that is, in the number of brands. An increase of brands leads to a rapid increase of parameters.

3.3 Interpretation

It is well known that ANN parameters, which are the weights in (1) and (3), cannot be dealt with as in a linear regression model. See Bishop (1995) and Franses and Draisma (1997), among others. Hence, no t -statistics exist and determining the specific significance of a variable is tedious. The main reason for this drawback is that there are many more parameters than there are observable variables.

To determine the influence of variables, one usually performs a sensitivity analysis. The variables are fixed at their average value and only the value of one variable is allowed to vary. Changes in the outcome of the nodes in the hidden layer and the output layer can give an idea of the importance of that particular variable. Graphs of the outputs of these nodes are then typically used to interpret these changes.

3.4 Parameter estimation

To estimate the parameters in an ANN it is conventional to use the back-propagation algorithm. The principle of the back-propagation algorithm is to propagate a training example in forward direction through the network and to calculate for each node its output value. Arrived at the end of the network, the errors made at the nodes in the output layer can be determined. Next, these errors are propagated backwards through the network and accordingly, the parameters are adjusted. See Mitchell (1997) for a useful discussion on the back-propagation algorithm.

The structure of our model results in a number of nodes that is equal to the number of available brands for both the hidden and output layer. A node in the hidden layer, so to say, consults household-specific variables and the variables concerning the corresponding brand. This implies restricting some parameters to a value of zero. In network terminology, this means that the network is not fully connected. This property of the network has no effect on the back-propagation algorithm. The only adjustment made to the algorithm concerns incorporating the estimation of the parameters of the

extra information used to determine the final choice. These parameters are adjusted in a similar way as the parameters of the input layer.

Estimation of the parameters in an ANN can take thousands of iterations. Progress of this estimation process is usually measured by a sum-of-squares error function. Training an ANN, that is, estimating the parameters with a sum-of-squares error function assumes a priori that the target data are generated from a smooth deterministic function with added Gaussian noise. In our case, the Gaussian model does not provide a good description, as the target data are binary variables. For such models, that use binary variables, a cross-entropy function, in combination with (2), is more appropriate see Bishop (1995). The value of the cross-entropy function is determined according to

$$E = - \sum_{n=1}^N \sum_{j=1}^J t_j^n \ln FC_j^n \quad (4)$$

where N is the number of training examples and J denotes the number of brands in the output layer. The cross-entropy function differs from the sum-of-squares error function in that only the error of the node which should have been activated (FC_j^n) is considered. This is due to the target values (t_j^n). These are binary and only one of them has a nonzero value for the J alternatives. This ensures that the values of the non-desired nodes drop out.

Another point of concern is the problem of overfitting. To avoid this, two samples are used. The first sample is used to estimate the parameters. The second sample is used to determine the progress of the training process, see Mitchell (1997). The algorithm iterates until the cross-entropy value for the second sample reaches an approximately fixed value.

4 Illustration

To illustrate the empirical usefulness of our parametric model for a two-stage brand choice process, we demonstrate it on scanner data for six liquid detergent brands.

4.1 Data

Our data consists of scanner data with information on six liquid detergent brands, that is, Tide, Wisk, Eraplus, Surf, Solo and All¹. The data contain

¹This is the A.C. Nielsen household scanner panel data on purchases of liquid laundry detergents in the Sioux Falls, South Dakota, market.

3055 observations of liquid detergent purchases, concerning 400 households. Each observation consists of four household-specific variables. These include the volume the household purchased on the previous purchase occasion, the expenditures on non-detergents, the size of the household and the inter-purchase time. Furthermore, three variables are available for each brand per observation. These are the price of a brand (cents/oz.), a 1/0 dummy variable for feature promotion, and a 1/0 dummy for display. An average household consists of three members. The average inter-purchase time is 80 days and expenditures on non-detergents are \$34.69. Average prices for the six brands range from 3.9 to 6.1 cents/oz. See Chintagunta and Prasad (1998) for a further discussion of this data.

There are 398 incomplete observations, and these are eliminated from the sample. Further, we introduce an extra variable for each brand. For each observation, we determine whether that brand was purchased on the previous purchase occasion. As a result, the first recorded purchase of each household is lost. Hence, in total we have $3055 - 398 - 400 = 2257$ observations.

Due to the different scales of the variables, data transformations seem appropriate. All variables (say, w_t) are scaled to the [0,1]-interval using

$$w_t^* = \frac{w_t - \min(w_t)}{\max(w_t) - \min(w_t)} \quad (5)$$

Finally, for estimation and forecasting purposes, the data sample is divided into three parts. The first sample is used for estimation of the parameters. The second sample is used as a validation sample to monitor the progress of the estimation process and the third sample is used for out-of-sample forecast evaluation. Observations are distributed randomly over the three samples. The first sample contains 1000 observations, the second 500 and the third sample contains 757 observations.

4.2 Estimation results

As there are six brands to choose from, the hidden and output layer therefore both consist of six nodes. All variables mentioned above are to be used in this six brand choice model. Each household-specific and brand-specific variable is represented with a node in the input layer, which thus consists of $4 + 4 \cdot 6 = 28$ nodes. Due to the restriction that brand variables are only available to corresponding nodes in the hidden layer, parameters for other nodes are set equal to 0. The values of the output layer depend on the outputs of the hidden layer as well as on the prices of the brands, which serve as extra information for making the final choice.

Table 1: Results of the proposed ANN in comparison with an one-stage MNL model

Sample	Hitrate ¹		#observations
	ANN	MNL	
Estimation	78.5%	77.6%	1000
Validation	78.8%	78.4%	500
Forecast	76.0%	72.4%	757
#parameters	132	110	

¹ Percentage of correct predicted purchases.

Note that the parameters of the ANN are estimated for the estimation sample, using the validation sample for progress evaluation. The parameters of the MNL model are estimated for the estimation sample only.

The parameters of the model are estimated five times. Each estimation process consists of 1000 iterations of the back-propagation algorithm. After 1000 iterations, the cross-entropy value is approximately constant, thereby indicating that an optimum has been reached. The estimated parameters, which result in the lowest cross-entropy value for the validation sample, are selected to be used for the evaluation of the impact of the variables and forecasting.

The performance of the model is measured by the fit for each of the three samples. The hitrate, that is, the percentage of times the estimated selected brand matches the actually chosen brand, is determined. The results are shown in the second column of Table 1. Based on these results, the performance of our model appears to be rather good. To determine whether our two-stage model gives a better fit than a one-stage model, the parameters of a MNL model are estimated for comparison. The MNL model is estimated for the first (estimation) sample, using the same variables as the ANN. Results for this one-stage model are shown in the third column of Table 1. Apparently, the ANN performs slightly better, especially for the out-of-sample data.

4.3 Graphical inference

Due to the large number of explanatory variables, there are many ways to visualize the effects of variables. For example, one can investigate the competitive structure between brands or the influence of promotion activities such as featuring a brand. In this section, we will give a few examples.

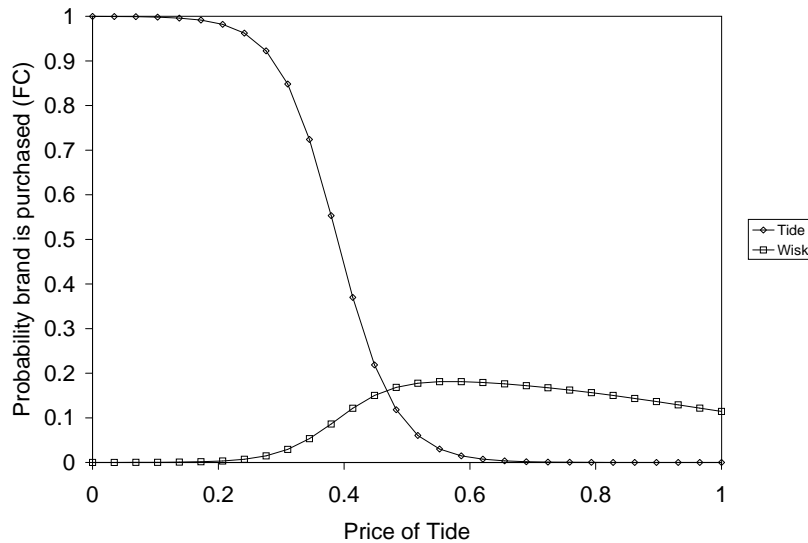


Figure 3: Change in the probability that Tide and Wisk are chosen due to price changes of Tide

Tide and Wisk are the two largest brands in our data set. It is now interesting to see whether and how the probability that Wisk is chosen reacts on a price change of Tide. To determine the possible influence of the price of Tide on Wisk, all variables are fixed at their average value and the price of Tide is varied over the $[0,1]$ -interval. The effect of the price change is shown in Figure 3, which shows the probabilities for Tide and for Wisk that they are chosen for various price levels of Tide. Clearly, Wisk is not a key competitor to Tide. The probability that Wisk is purchased increases but not in an explosive way one would expect for a main competitor. It is also interesting to see that the probability that Tide is chosen stays long at a high level before rapidly dropping towards a value of zero. This suggests that there is some threshold for loyalty.

The effect of promotion activities, that is, feature and display for Tide can be determined by switching these variables on and off during a price increase of Tide. Figure 4 shows four relevant cases concerning the effect on the node for Tide in the hidden layer. The benchmark case concerns no promotion and no purchase of Tide at the previous occasion. Featuring has a clear positive effect on the hidden layer, that is, the probability of considering Tide starts at a higher level. A display has a similar though a little stronger effect on the probability of considering Tide. The effect of

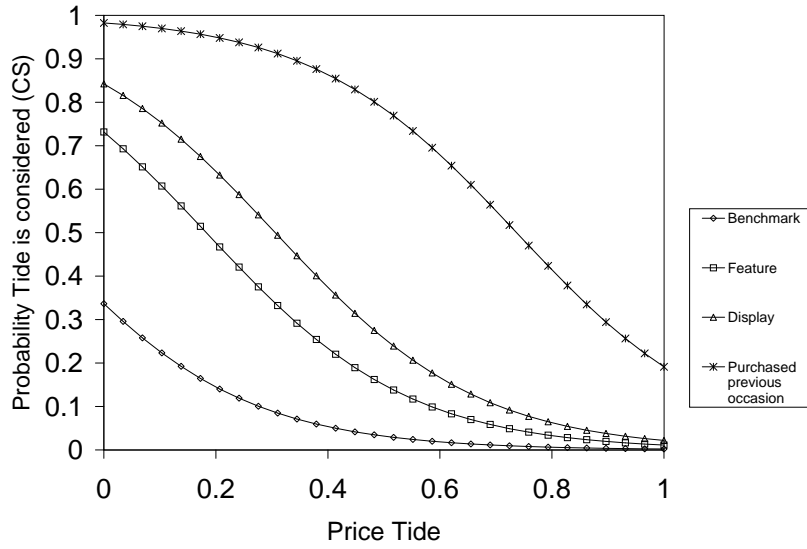


Figure 4: The estimated probability that Tide is considered due to changes in price of Tide, when combined with a feature, display and previous purchase

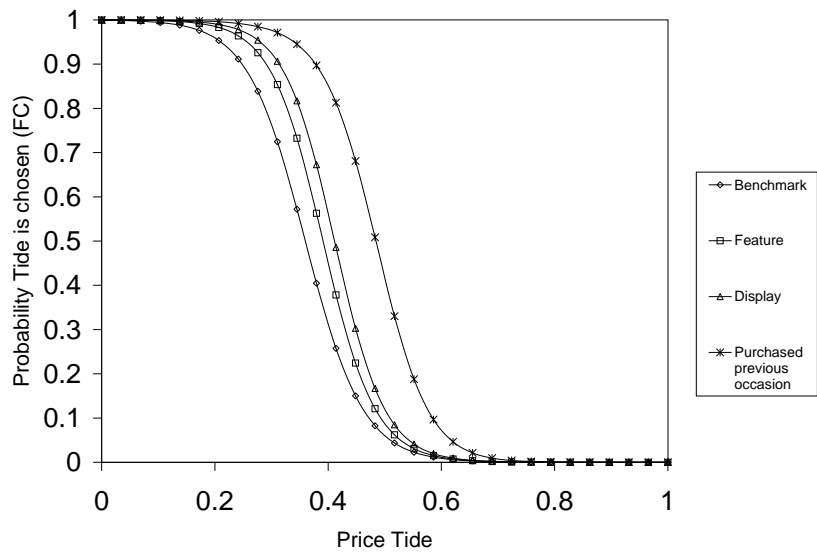


Figure 5: The estimated probability that Tide is purchased due to changes in price of Tide, when combined with a feature, display and previous purchase

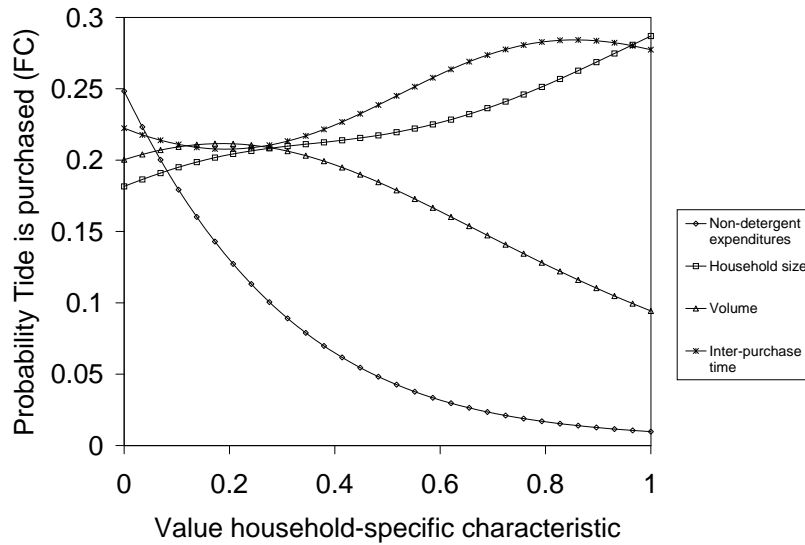


Figure 6: Changes in the probability that Tide is chosen due to changes in household-specific variables

the same promotion variables on the final choice is depicted in Figure 5. Activating one of the two promotion variables causes the curve to shift to the right, resulting in an increasing willingness to accept a higher price by the (average) household. Again, the effect of a display is larger than that of a feature. This can indicate that purchases are based more on stimuli, noticed at the time of purchase, than on recall of memory.

The variable that indicates whether Tide was purchased on the previous occasion can be interpreted as a kind of loyalty variable. When this variable is relevant, there should be a positive effect on consideration and final choice. The effect of this variable on the probability of consideration is indeed large in comparison with the benchmark situation as can be seen from Figure 4. The effect on the probability that Tide is finally selected, which is shown in Figure 5, is also positive as the curve shifts to the right. Hence, if one has purchased Tide on the previous occasion, one is more likely to choose for Tide in the current occasion.

The impact of the household variables can be determined in a similar way as in the previous two illustrative examples. Again all variables are fixed at their average value, though now each household-specific variable varies over the $[0,1]$ -interval. Figure 6 shows the influence of changes of these variables on the probability of choosing Tide. Clearly, their influence on

this probability is highly non-linear. The results indicate that Tide is more likely to be purchased in small amounts by larger households that do not frequently purchase liquid detergents and spend a limited amount of money on non-detergents when a detergent is purchased.

The above three illustrations give an indication of the possibilities of our artificial neural network model. Other illustrations can easily be generated in order to provide information that can be valuable to a marketing researcher. In fact, we only used the average values for our analysis. Another type of analysis could cluster households according to, for example, household size and analyze the influence of different scenarios on each cluster. In sum, the model turns out to be a good forecasting tool that can also be used to graphically analyze the effects of changes in important variables.

5 Discussion

The concept of consideration sets has often been used in marketing studies on brand choice. Considering the decision making process as a two-stage process seems to be more appropriate than imposing that a brand is selected in one step. Based on these two stages, we proposed a new parametric econometric model, which appeared to be an artificial neural network. The first stage, that is, the reduction of available brands, was transformed into a stage where for each brand the probability of consideration is determined. This resulted in the interpretation of the hidden layer as the consideration set. Based on the outcome of the first stage, the probability was determined for each brand that it would be chosen.

The forecasting abilities of the model were good in comparison with a one-stage multinomial logit model. Furthermore, an illustration for six liquid detergent brands showed that the model cannot only be used as a tool for forecasting but also for evaluating the effect of individual variables.

A limitation of our model concerns the estimation of the parameters. Due to the fact that the back-propagation algorithm is a heuristic algorithm, one cannot tell whether the parameters estimated are the optimal parameters. We tried to avoid this problem by estimating the model five times and selecting the best one, but still then there is no guarantee that this yields the best parameters. Furthermore, we studied only one data set and other illustrations should reveal whether our model indeed is more useful than a one-stage model. Finally, an issue we postpone for further research is the incorporation of household-specific unobserved heterogeneity.

References

- Agrawal, D. and C. Schorling (1996), Market share forecasting: An empirical comparison of artificial neural networks and multinomial logit model, *Journal of Retailing*, **72**, 383–407.
- Andrews, R. L. and A. K. Manrai (1998), Simulation experiments in choice simplification: The effects of task and context on forecasting performance, *Journal of Marketing Research*, **35**, 189–209.
- Bentz, Y. and D. Merunka (2000), Neural networks and the multinomial logit for brand choice modelling: a hybrid approach, *Journal of Forecasting*, **19**, 177–200.
- Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford.
- Chiang, J., S. Chib, and C. Narasimhan (1999), Markov chain monte carlo and models of consideration set and parameter heterogeneity, *Journal of Econometrics*, **89**, 23–248.
- Chintagunta, P. K. and A. R. Prasad (1998), An empirical investigation of the dynamic “McFadden model” of purchase timing and brand choice: Implications for market structure, *Journal of Business & Economic Statistics*, **16**, 2–11.
- Daganzo, C. (1979), *Multinomial Probit: The Theory and its Applications to Demand Forecasting*, Academic Press, New York.
- Dasgupta, C. G., G. S. Dispensa, and S. Ghose (1994), Comparing the predictive performance of a neural network model with some traditional market response models, *International Journal of Forecasting*, **10**, 235–244.
- Franses, P. H. and G. Draisma (1997), Recognizing changing seasonal patterns using artificial neural networks, *Journal of Econometrics*, **81**, 273–280.
- Franses, P. H. and D. van Dijk (2000), *Nonlinear Time Series Models in Empirical Finance*, Cambridge University Press, Cambridge.
- Guadagni, P. M. and J. D. C. Little (1983), A logit model of brand choice calibrated on scanner data, *Marketing Science*, **2**, 203–238.

- Hausman, J. and D. Wise (1978), A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogenous preferences, *Econometrica*, **45**, 319–339.
- Horowitz, J. L. and J. J. Louviere (1995), What is the role of consideration sets in choice modelling?, *International Journal of Research in Marketing*, **12**, 39–54.
- Kumar, A., V. R. Rao, and H. Soni (1995), An empirical comparison of neural network and logistic regression models, *Marketing Letters*, **6**, 251–263.
- Lattin, J. M. and E. Bucklin (1989), Reference effects of price and promotion on brand choice behavior, *Journal of Marketing Research*, **26**, 299–310.
- Manrai, A. K. and R. L. Andrews (1998), Two-stage discrete choice models for scanner panel data: An assessment of process and assumptions, *European Journal of Operational Research*, **111**, 193–215.
- McFadden, D. (1973), Conditional logit analysis of qualitative choice behavior, in P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, pp. 105–142.
- Mitchell, T. M. (1997), *Machine Learning*, McGraw-Hill, Singapore.
- Roberts, J. H. and J. M. Lattin (1991), Development and testing of a model of consideration set composition, *Journal of Marketing Research*, **28**, 429–440.
- Roberts, J. H. and J. M. Lattin (1997), Consideration: review of research and prospects for future insights, *Journal of Marketing Research*, **34**, 406–410.
- Shocker, A. D., M. Ben-Akiva, B. Boccara, and P. Nedungadi (1991), Consideration set influences on consumer decision-making and choice: issues, models and suggestions, *Marketing Letters*, **2**, 181–197.
- Vellido, A., P. J. G. Lisboa, and J. Vaughan (1999), Neural networks in business: a survey of applications (1992-1998), *Expert Systems with Applications*, **17**, 51–70.
- West, P. M., P. L. Brockett, and L. L. Golden (1997), A Comparative analysis of neural networks and statistical methods for predicting consumer choice, *Marketing Science*, **16**, 370–391.
- Zhang, G., M. Y. Hu, B. E. Patuwo, and D. C. Indro (1999), Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis, *European Journal of Operational Research*, **116**, 16–32.

Publications in the Report Series Research* in Management

ERIM Research Program: "Decision Making in Marketing Management"

2001

Predicting Customer Potential Value. An application in the insurance industry

Peter C. Verhoef & Bas Donkers

ERS-2001-01-MKT

Modeling Potentially Time-Varying Effects of Promotions on Sales

Philip Hans Franses, Richard Paap & Philip A. Sijthoff

ERS-2001-05-MKT

Modeling Consideration Sets and Brand Choice Using Artificial Neural Networks

Björn Vroomen, Philip Hans Franses & Erjen van Nierop

ERS-2001-10-MKT

Firm Size and Export Intensity: A Transaction Costs and Resource-Based Perspective

Ernst Verwaal & Bas Donkers

ERS-2001-12-MKT

Customs-Related Transaction Costs, Firm Size and International Trade Intensity

Ernst Verwaal & Bas Donkers

ERS-2001-13-MKT

2000

Impact of the Employee Communication and Perceived External Prestige on Organizational Identification

Ale Smidts, Cees B.M. van Riel & Ad Th.H. Pruyn

ERS-2000-01-MKT

Forecasting Market Shares from Models for Sales

Dennis Fok & Philip Hans Franses

ERS-2000-03-MKT

The Effect of Relational Constructs on Relationship Performance: Does Duration Matter?

Peter C. Verhoef, Philip Hans Franses & Janny C. Hoekstra

ERS-2000-08-MKT

Informants in Organizational Marketing Research: How Many, Who, and How to Aggregate Response?

Gerrit H. van Bruggen, Gary L. Lilien & Manish Kacker

ERS-2000-32-MKT

The Powerful Triangle of Marketing Data, Managerial Judgment, and Marketing Management Support Systems

Gerrit H. van Bruggen, Ale Smidts & Berend Wierenga

ERS-2000-33-MKT

Consumer Perception and Evaluation of Waiting Time: A Field Experiment

Gerrit Antonides, Peter C. Verhoef & Marcel van Aalst

* A complete overview of the ERIM Report Series Research in Management:

<http://www.ers.erim.nl>

ERIM Research Programs:

LIS Business Processes, Logistics and Information Systems

ORG Organizing for Performance

MKT Decision Making in Marketing Management

F&A Financial Decision Making and Accounting

STR Strategic Renewal and the Dynamics of Firms, Networks and Industries

ERS-2000-35-MKT

Broker Positions in Task-Specific Knowledge Networks: Effects on Perceived Performance and Role Stressors in an Account Management System

David Dekker, Frans Stokman & Philip Hans Franses

ERS-2000-37-MKT

Modeling Unobserved Consideration Sets for Household Panel Data

Erjen van Nierop, Richard Paap, Bart Bronnenberg, Philip Hans Franses & Michel Wedel

ERS-2000-42-MKT

A Managerial Perspective on the Logic of Increasing Returns

Erik den Hartigh, Fred Langerak & Harry Commandeur

ERS-2000-48-MKT

The Mediating Effect of NPD-Activities and NPD-Performance on the Relationship between Market Orientation and Organizational Performance

Fred Langerak, Erik Jan Hultink & Henry S.J. Robben

ERS-2000-50-MKT

Sensemaking from actions: Deriving organization members' means and ends from their day-to-day behavior

Johan van Rekom, Cees B.M. van Riel & Berend Wierenga

ERS-2000-52-MKT