

Modeling Toothpaste Brand Choice: An Empirical Comparison of Artificial Neural Networks and Multinomial Probit Model

Tolga Kaya*

*Management Engineering Department, Istanbul Technical University.
Macka, Besiktas, Istanbul, 34367, Turkey.*

Emel Aktaş

*Industrial Engineering Department, Istanbul Technical University.
Macka, Besiktas, Istanbul, 34367, Turkey.*

İlker Topçu

*Industrial Engineering Department, Istanbul Technical University.
Macka, Besiktas, Istanbul, 34367, Turkey.*

Burç Ülengin

*Management Engineering Department, Istanbul Technical University.
Macka, Besiktas, Istanbul, 34367, Turkey.*

Abstract

The purpose of this study is to compare the performances of Artificial Neural Networks (ANN) and Multinomial Probit (MNP) approaches in modeling the choice decision within fast moving consumer goods sector. To do this, based on 2597 toothpaste purchases of a panel sample of 404 households, choice models are built and their performances are compared on the 861 purchases of a test sample of 135 households. Results show that ANN's predictions are better while MNP is useful in providing marketing insight.

Keywords: Brand choice modeling, artificial neural networks, multinomial probit, toothpaste, household panel

1. Introduction

Due to the emergence of a strong trend towards the utilization of behavioral-based knowledge of consumer behavior, scanner panels which provide transactional data and consumer profile databases have recently gained more importance. The researchers which used to focus on the impacts of subjective aspects like cultural values, attitudes and psychological factors on the choice behavior turned their focus on measurable parameters like prices, purchase frequency, and average purchase size. Consequently, the effort of using behavioral data towards developing decision tools for planning marketing activities have resulted in numerous different modeling applications based on both statistical and non-statistical approaches.

It is critical for businesses to have successful estimations on the choices of their potential customers. Market share forecasts are vital for not only producers but also media planners and retailer companies. Modeling studies may be quite useful as brand choice decisions are usually associated with multiple variables at the same time. These variables may differ from relative prices, intensity of advertisements, levels of customer loyalty, and consumer characteristics to the usage and intensity of promotion activities (e.g. price cuts, couponing, display etc.) offered by the producers or retailers. There are compensatory linear models for determining consumer preferences, attitudes, judgments, and decision making process; namely regression models, analysis of variance, discriminant analysis, and structural equation modeling. The main issue with these models is the fact that preference structure of the

*Corresponding author: kayatolga@itu.edu.tr, Tel: +90 212 2931300 ext. 2789.

consumers is not linear and their judgments are not based on compensatory rules.¹⁻²

Multinomial logit model (MNL) is a non linear model which has been found to be a robust modeling tool in forecasting brand shares in terms of modeling the consumers' choice probabilities. However, as the number of brands analyzed increase MNL may have classification problems.³ More importantly, MNL model requires independence of irrelevant alternatives (IIA) principle to hold. According to IIA assumption, the probabilities of choosing the existing alternatives should equally be affected by the entrance of a new alternative to the choice set.⁴ In practice, within fast moving consumer goods (FMCG) industry, this principle rarely holds since new brand launches in a specific FMCG category seldom have the same effects on the existing brands. This drawback of MNL limits the usage of the model in many real life cases.

An alternative to MNL is the multinomial probit model (MNP), which assumes the errors are distributed multivariate normal with mean 0 and a covariance matrix, thus, does not require IIA to hold. Despite this advantage over MNL, MNP has also its own disadvantages in computational difficulties. Until the last decade, researchers had to deal with the multiple integrals of MNP in order to make estimations. As the number of alternatives increased, it became practically impossible to handle the calculations. In recent years, some statistical software packages like STATA and LIMDEP started providing MNP estimations. Although this increased the usage of the MNP modeling in practice, it should be noted that, estimation of a MNP model using econometric software may still take thousands of times longer than that of a MNL model.

In order to overcome the limitations of MNL and MNP, more general, non/semi-parametric, non-linear regression models capable of modeling nonlinear utility functions without a priori knowledge of relationships can be used. ANN is one such model that can be used to predict the consumer brand choice behavior. Despite having a relatively short history in consumer behavior, there are many studies on brand choice modeling using ANN as an alternative analysis tool.^{2,5} The advantage of ANN is that it does not have specification bias and it can be used to model highly complex relationships. However, the difficulty in interpreting the results combined with the fact that it does not provide an explanation on how it finds the outcomes are the

reasons why it is regarded as a black box. When studying consumer behavior, interpretability may often be as much important as the prediction performance.

The aim of this study is to compare the performances of ANN and MNP approaches in modeling the brand choice decision within Turkish fast moving consumer goods sector. In order to do this, initially, ANN and MNP models of brand choice are built based on 2597 real toothpaste purchases of a model sample of 404 households. In these models, variables, which were found to be significant in explaining the brand choice in Turkish toothpaste market, namely relative prices, socio-economical status, brand loyalty, and household size were used as inputs. After the models were built and the estimations were realized, the performances of these models were compared in terms of hit-rates (successful predictions of the actual choices) and market share prediction on the 861 purchases of a randomly selected test sample of 135 households. The transactional data was obtained from a diary based consumer panel company which keeps the tracks of shopping behavior on more than 100 FMCG product categories in Turkey since 1997.

Along with the theories of chaos, evidence and fuzzy sets, neural networks and discrete choice probabilistic computing are among the most widely used methodologies in establishing computational intelligence systems. This study makes use of two of these methodologies, ANN and MNP, in order to model the choice behavior of Turkish toothpaste consumers. As ANNs are able to handle the nonlinearities within the data structures, due to the nature of the sector under consideration, they may provide better predictions than probabilistic modeling. This gives birth to a necessity of sector specific modeling applications conducted in a comparative manner. Suggesting a solution to the missing price data in diary mode panels, to the authors' knowledge, this study is the first application of ANN modeling based on diary based household panel data.

The rest of the paper is organized as follows: In section two, a brief literature review on brand choice modeling using multinomial models and ANN is given. In the third section, theoretical backgrounds of MNP and ANN methodologies are summarized. Section four contains a comparative case study conducted in Turkish toothpaste market based on consumer panel data. Finally, in the fifth section concluding remarks are given.

2. Literature Review

The object of consumer choice models is to model the purchase behavior of consumers and more specifically, to model the procedure of purchase decision. A question of continuing interest to researchers and practitioners has always been how marketing mix variables affect different consumers' buying behavior. With the proliferation of scanner panel data usage in the middle 80s, an important number of statistical brand choice models have been developed to determine the effects of marketing tools such as pricing, promotions and advertisements on the brand sales, shares, and profits.⁶⁻⁸ One of the first attempts to build a multinomial logit model of brand choice based on household scanner panel data was the study of Guadagni and Little⁹ the success of which was attributed in part to the level of detail and completeness of the consumer panel data used, which had been gathered through scanning of the barcodes in retailers. Following Guadagni and Little, a number of researchers made important contributions to the brand choice models based on scanner data, by separating the purchase decision process into different levels. Targeting to decompose sales increases, Gupta⁶ proposed a method within which brand sales were considered the result of consumer decisions about when, what, and how much to buy. Leaning on the assumption that "a customer decides to purchase a product category first and, if so, buys a particular brand", Guadagni and Little¹⁰ rebuilt a nested logit model with the same ground coffee data they employed in their 1983 paper. Bucklin et al.¹¹ developed a joint approach to segment households on the basis of their response to price and promotion in brand choice, purchase incidence, and purchase quantity decisions. Most probably the biggest portion of the statistical brand choice models literature is devoted to the evaluation of the effectiveness of price cuts and other promotional activities. In shaping this story, Neslin et al.¹², were one of the first researchers in addressing the question of "borrowing from future sales" via promotions. Mela et al.¹³ examined the long term effects promotion and advertising on consumers' brand choice behavior. Another study extended the analysis by taking the consumer stockpiling behavior into consideration.¹⁴ In the model of Jedidi et al.¹⁵, instead of brand sales or shares, the analysis unit was profitability. Pauwels et al.¹⁶ calculated the long term equivalent of Gupta's breakdown of promotional effects and found a reversal of the importance of category

incidence and brand choice. While Klapper et al.¹⁷ was focusing on the loss aversion in brand choice data, Silva-Risso and Bucklin¹⁸ developed a logit modeling approach to assess the effects of coupon promotions on consumer brand choice. Leaning on scanner data, van Heerde et al.¹⁹ investigated the short-term and long-term effects of the price war between retailers. When studying the sensitivity of the consumers to the prices, some researchers took both the demand and the supply (manufacturers and retailers) sides into consideration.²⁰⁻²¹

Although ANN has a relatively short history in modeling brand choice and consumer behavior, it has been widely used in consumer decision making to predict shopping behavior. Agraval and Schorling³ compared the forecasting ability of ANN with MNL in the context of frequently purchased grocery products. West et al.²² explored the advantages and the disadvantages of ANN relative to statistical modeling procedures in predicting consumer choice. Bentz and Merunka⁵ developed a hybrid approach which combines ANN and MNL into a single framework in the brand choice modeling context. Hruschka et al.²³⁻²⁴ specified deterministic utility by means of a certain type of neural net for discovering nonlinear effects on brands' utilities and compared the performance of this model with different MNL models. Hu and Tsoukalass²⁵ used neural network models and the ensemble technique of stacked generalization to investigate the relative importance of situational and demographic factors on consumer choice. Fish et al.²⁶ introduced a new architectural approach to ANN choice modeling and used a feed-forward ANN trained with a genetic algorithm to model individual consumer choices and brand share in a retail coffee market. Vroomen et al.²⁷ proposed a two step ANN choice modeling framework in the first step of which they took consideration sets of the households into account. Hruschka²⁸ introduced a MNP model which combines heterogeneity across households with flexibility of the deterministic utility function which is approximated by a multilayer perceptron neural net.

3. Methodology

In this research, diary based household panel data is used to build MNP and ANN models of consumer choice. Initially, MNP model is established in order to determine the relevant and significant variables of consumer choice. Secondly an ANN based on the same

inputs (independent variables) is built to predict the consumer choice. Thirdly, performances of these two models are compared in terms of hit rates and market share estimations. Finally, a sensitivity analysis is conducted to see the change in choice probabilities with respect to different price and socio-economical status levels. A framework of the methodology is given in Fig. (1).

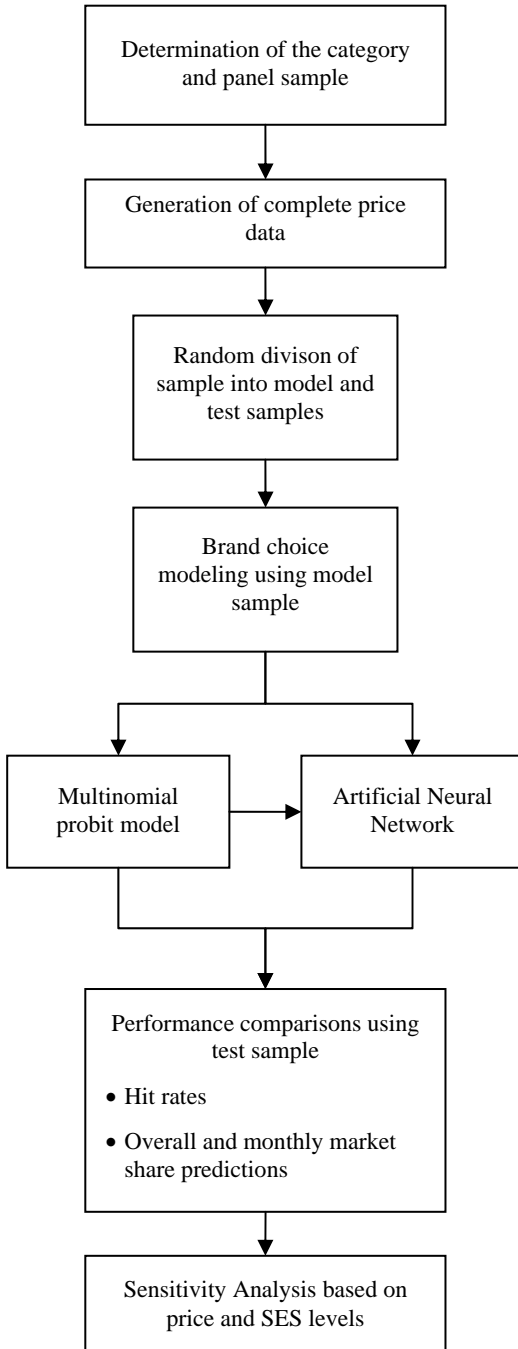


Figure 1 Framework of the proposed methodology

In the following subsections, a brief theoretical background on MNP and ANN will be provided.

3.1. Multinomial Probit Model

Modeling the brand choice, researchers have to adopt the appropriate models of consumer decisions among multiple product alternatives. In many cases multinomial logit (MNL) and multinomial probit (MNP) statistical models meet this requirement as each may be derived from economic theories of utility maximization. In a multi-brand category, assume household i 's utility for brand j , U_{ij} ($i = 1, \dots, n; j = 1, \dots, p$); is a function of household attributes and a stochastic error. A typical representation is:²⁹

$$U_{ij} = \beta_j' X_i + \varepsilon_{ij}, \quad (1)$$

where X_i is a vector of household characteristics. The probability that a particular consumer will choose a particular alternative is given by the probability that the utility of that alternative to that consumer is greater than the utility of all other alternatives to that consumer.⁴ The probability that a household i will choose brand j is given by:²⁹

$$P(\text{choice} = j | \beta_j, X_i) = \frac{\exp(\beta_j' X_i)}{\sum_{k=1}^p \exp(\beta_k' X_i)} \quad (2)$$

A well known specification test for determining the validity of the IIA property is the Hausman test. The test statistics is asymptotically χ^2 distributed. The IIA assumption is rejected for large values of Hausman statistics.³⁰⁻³¹ In case of rejection, alternative models such as MNP or nested logit will be needed.³² On the other hand, the MNP assumes that the errors are multivariate normally distributed, with mean 0 and covariance matrix Σ . The probabilities are written:

$$P(\text{choice} = j | \beta_j, X_i, \Sigma^*) = \int_{-\infty}^{\beta_1^* X_i^*} \dots \int_{-\infty}^{\beta_{j-1}^* X_i^*} f(\varepsilon_{i1}^*, \dots, \varepsilon_{ij-1}^*) \partial \varepsilon_{i1}^*, \dots, \partial \varepsilon_{ij-1}^* \quad (3)$$

where $f(\cdot)$ is the probability density function of the multivariate normal distribution.²⁹

In choice models, accuracy can be measured either in terms of the fit between the calculated probabilities and observed frequencies or in terms of the model's performance of forecasting observed responses.³³ One

of the most widely used goodness of fit measures in brand choice models is the ρ^2 statistic suggested by McFadden. Given that the loglikelihoods of the restricted and unrestricted models are LL_0 and LL_F respectively, the ρ^2 statistic can be written as:

$$\rho^2 = 1 - \frac{LL_F}{LL_0} \quad (4)$$

As the ρ^2 statistic increases, the accuracy level of the model in question increases.³⁴

In probabilistic choice models, it is also useful to look at the proportion of successful predictions of the choices made. A table of success can be prepared for a case of m alternatives. Using this table, given that N_{ii} is the number of correct predictions for alternative i , a commonly used statistics can be calculated:

$$S_1 = \frac{1}{N_{..}} \left(\sum_{i=1}^m N_{ii} \right) \quad (5)$$

This statistics is simply the total number of choices that were predicted correctly divided by all choices.³³ Finally, keeping in mind that a choice model predicts a probability of purchase for each observation and any given brand, Guadagni and Little⁹ (letting s denote the predicted share and n the number of observations) suggests a calculation of standard error of the predicted share as below:

$$s = \sum_{i=1}^n p_i / n$$

$$SE(s) = \left[\sum_{i=1}^n p_i (1 - p_i) \right]^{1/2} / n \quad (6)$$

3.2. Artificial Neural Networks

A variety of problem areas are modeled using ANN³⁵⁻³⁷ and, in many instances, ANN has provided superior results compared to the conventional modeling techniques.³⁸ It is published by several researchers that ANN performs excellently on pattern recognition tasks and its potential advantages have been addressed in the literature.³⁹⁻⁴¹ ANN performs better in the presence of extreme values and its estimation process can be automated. However regression and ARIMA models must be re-estimated periodically as new data is obtained. ANN outperforms the traditional methods in problem domains with non-linear relationships⁴²; in fact,

it could be said that ANN is primarily used for complex non-linear mapping purposes⁴³.

The basic model of ANN consists of computational units, which as a whole mimic the human brain. ANN is regarded as a black box that takes a weighted sum of all inputs and computes an output value using a transformation or output function (Figure 2). The output value is then propagated to many other units via connections between units.

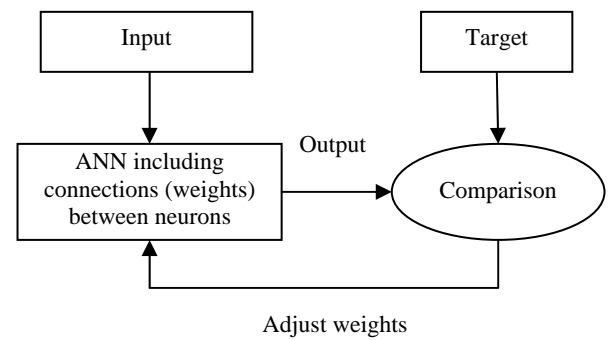


Figure 2 Conceptual operation of ANN models

In general, the output function is a linear function – a threshold function in which a unit becomes active only when its net input exceeds the threshold of the unit, or a sigmoid function which is a non-decreasing and differentiable function of the input. Computational units in an ANN model are hierarchically structured in layers and depending upon the layer in which a unit resides, the unit is called an input, a hidden or an output unit. An input (output) unit is similar to an independent (dependent) variable in a statistical model. A hidden unit is used to augment the input data in order to support any required function from input or output. In the ANN literature, the process of computing appropriate weights is known as “learning” or “training”. The learning process of ANN can be thought of as a reward and punishment mechanism⁴⁰, whereby when the system reacts appropriately to an input, the related weights are strengthened. As a result, it is possible to generate outputs, which are similar to those corresponding to the previously encountered inputs. Contrarily, when undesirable outputs are produced, the related weights are reduced. The model learns to give a different reaction when similar inputs occur, thus gearing the system towards producing desirable results, whilst the undesirable ones are “punished”.

In this study, a feedforward backpropagation network is used to model the consumer choice. The training algorithm was selected to be *trainscg*, which is a supervised learning algorithm based on a class of optimization techniques known as conjugate gradient methods⁴⁴. The *trainscg* may require more iterations to converge than the other conjugate gradient algorithms, but the number of computations in each iteration is significantly reduced because no line search is performed. This algorithm is too complex to explain in a few lines, see Ref. 44 for a detailed explanation of the algorithm.

4. Case Study

4.1. Data

Consumer panel data for toothpaste category is used in the MNP and ANN models. The raw data covers 7,681 toothpaste transactions in approximately 90% (6,943) of which three main brands were purchased by a panel of 1,955 households. Finally, 3,458 toothpaste purchases of 539 frequent category buyers are used for the study. Frequent category buyer is defined as a household who purchased toothpaste 5 times or more during the analysis year (2004).

Table 1 Demographic characteristics of the households in the sample

Socio-economical status		Primary shopper age	
	%		%
AB	30.1	25-	7.3
C1	33.9	26-35	21.5
C2	20.5	36-45	48.8
DE	15.4	46-55	17.6
		56+	4.8
Primary shopper education		Household size	
	%		%
Illiterate	2.3	2-	3.9
Literate	1.9	3	15.4
Primary school	41.2	4	40.7
Middle school	14.3	5	22.4
High school	34.8	6	7.7
University	5.5	7+	9.8

The set contains records of complete purchase information for each household in the panel (e.g.,

household id, brand purchased, price, quantity, place, time, etc.). In addition, data set includes household specific information such as socio-economical status, family size, age, education level, previous brands purchased, and total FMCG spending. The data does not have censored observations. In other words, panel members who either entered or left the panel during the study period are excluded from the data set. Table 1 gives a summary of the demographic profiles of the households used in the study:

According to 2004 panel records, three biggest brands represent more than 90% of the purchase occasions in toothpaste category. Among these three brands, market leader (Brand 1) has a share of 55.5% among all the purchases. Purchase shares of Brand 2 and Brand 3 are 22.2 % and 27.3 %, respectively. There are a number of small and private label brands competing in Turkish toothpaste sector, however these brands are not included in the analysis as they have a limited distribution and are not supported by similar marketing activities as of the three biggest brands. Another reason for the exclusion of the small brands is the difficulty of generating reliable price and loyalty information due to limited statistical base.

Table 2 Number of households and purchase observations before/after data reduction

	Number of households	Number of purchase observations
Toothpaste buyers	2030	7681
Buyers of the three main brands	1955	6943
Frequent buyers (households employed in the study)	539	3458
Model (training) sample	404	2597
Test (holdout) sample	135	861

The households in the sample have been randomly divided into two groups: Model and test samples (Table 2). MNP and ANN models of brand choice are built based on 2,597 purchase occasions of the model sample which includes 404 households. The performances of these models are tested on the 861 purchases of a test sample consisting of 135 households (25% of the total frequent buyers' sample).

4.2. Variables

Socio-economical status: Socio-economical status levels of the households are determined due to the results of a questionnaire filled and periodically updated by the households. The index takes the education level, occupation, ownership of certain items, and accommodation area of the household members into consider. The data set contains 2 different levels of socio-economical status: High SES and Low SES. If the SES level of the household is high, then the variable (SES High) takes the value of 1, otherwise 0.

Household Size: Household size (HHSize) represents the number of people living in the household according to the panel records during the study period.

Loyalty: Operationally, loyalty is defined as the weighted average of the last three purchases of the brand. The relative coefficient sizes of 0.5, 0.3, and 0.2 were used when weighting the first, second, and third prior purchases. As the sum of loyalties across brands equals 1 for a household and there are 3 alternatives (Brand 1, 2, and 3), two variables (Loyalty1, Loyalty2) are employed in the model.

Relative Prices: Price information for the brand purchased at a particular trip is simply generated by dividing the toothpaste spending made in Turkish liras (TL) by the quantity bought. On the other hand, as mentioned above, in diary based consumer panels, households do not record the prices of all the alternative brands displayed in the shelves of a store. Therefore, there is no direct price information available for the brands which are not purchased but present in the store during the shopping trip.

In order to generate unit price information for the alternative brands, in this study, a two stage procedure is implemented (Figure 3). Initially, the price information is generated according to the Stage 1. Based on 96% of the transactions (6,671 out of 6,943), unit prices of alternative brands are generated in this stage. When there is no transaction fulfilling the conditions suggested in stage 1, stage 2 is implemented. In stage 2, price for 272 observations are estimated.

After maintaining the purchase price and the prices for the alternatives that are not purchased, the relative prices are calculated. Finally, by computing the natural logarithms of these ratios, price variables employed in the model ($\log(\text{Price1}/\text{Price3})$) and $\log(\text{Price2}/\text{Price3})$) are obtained.

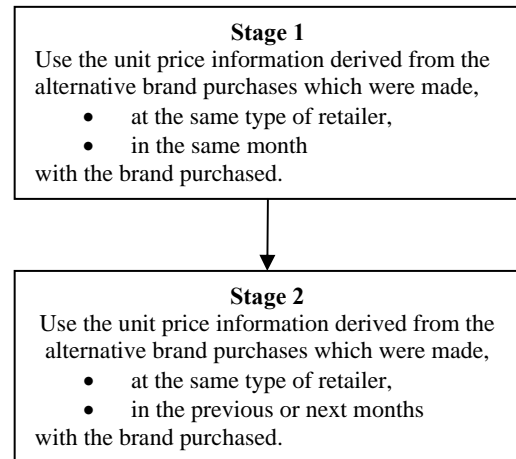


Figure 3 A two staged method of price data generation for the brands that are not purchased

4.3. The MNP Model

As it is seen in Table 3, estimation results show that price coefficients are significant and have expected signs. As the relative price of Brand 1 over Brand 3 increases, the probability of being chosen for Brand 1 over Brand 3 decreases which is in accordance with microeconomics theory. Similarly, as the relative price of Brand 2 over Brand 3 increases, the probability of being chosen for Brand 2 over Brand 3 decreases. Loyalty coefficients are positive and highly significant. As expected, if the loyalty of Brand 1 (Loyalty1) is higher, then it is more probable that Brand 1 is chosen instead of Brand 3. Similar findings are valid for other brands. Table 3 shows that there is an association between the SES levels and purchase decisions of the households between Brand 1 and Brand 3.

As the SES level increases, the probability of being purchased for Brand 1 against Brand 3 diminishes. Finally, as the household size increases choice probability of Brand 1 and Brand 2 over Brand 3 increases.

Wald and ρ^2 statistics are computed as 1,030 and 0.227, respectively. Both of the statistics are highly significant at 1 % level. Using Eq. (5) the hit rates (S_i) of the model are calculated as 66% and 63% for the model and test samples, respectively.

Table 3 Estimation results for the MNP model

	Brand 1	Brand 2
Constant	-1.252** (.170)	-1.415** (.184)
log (Price1/Price3)	-.960** (.161)	-.175 (.157)
log (Price2/Price3)	.243 (.162)	-.579** (.154)
Loyalty 1	3.073** (.133)	1.162** (.143)
Loyalty 2	1.279** (.165)	2.221** (.160)
SES High	-.199** (.093)	-.073 (.098)
HHSize	.089** (.029)	.059* (.033)

**p<0.05, *p<0.1 Base category: Brand 3

4.4. The ANN Model

To compare MNP and ANN methods, the same model and test samples are employed. An ANN with 4 hidden neurons is built as shown in Figure 4. Training function is chosen to be TRAINSCG. The learning rate is assigned to be 0.01 and the momentum to be 0.9. Performance function is mean square error as is usual in many feedforward networks. The model is trained for 201 epochs until the validation error (0.093) stopped improving. Seventy five percent of the data is used for building the ANN model. Since ANN model needs a training, a validating and a testing set; 75% of the data reserved for building the ANN (n=1947) is used for training and the remaining 25% is used for validating (n=650) the model. The test set used for testing the performance of the ANN is the same as that of the MNP model (n=861).

Model performance of ANN on the training set as well as the test set is pleasantly high. Overall hit rates are around 81% for the training and test samples.

In artificial neural networks, weights do not indicate the relationship in effect, magnitude, or the rate of change with respect to the target variable and the input

variables. Assessing the importance of the inputs based on the size of the weights is not easy since the output depends on both the input-to-hidden and hidden-to-target weight estimates. However, if all input-to-hidden weights associated with an input variable are close to zero, then it will indicate that the input variable does not have an important effect.⁴⁵

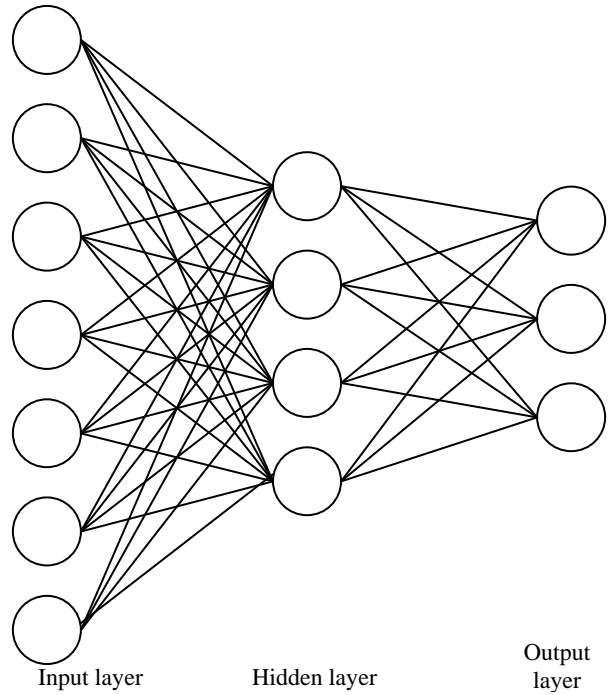


Figure 4 Structure of the ANN brand choice model

Due to the weights, relative price variables are more effective on brand choice than the rest of the variables. Similarly, loyalty indicators are among the most effective factors on choice decision (Table 4).

Table 4 Input Weights

	Hidden1	Hidden2	Hidden3	Hidden4
log (Price1/Price3)	0.333	-5.407	6.070	-0.628
log (Price2/Price3)	-2.827	0.392	-4.976	-2.123
Loyalty 1	0.578	0.593	0.525	0.506
Loyalty 2	0.358	0.655	0.351	0.788
Loyalty 3	0.123	0.795	0.328	0.659
SES High	-0.015	0.090	-0.027	0.034
HHSize	0.170	0.126	-0.129	0.095

Table 5 Hit rates for NN and MNP models on the test sample

Observed	ANN Model				MNP Model			
	Predicted				Predicted			
	Brand 1	Brand 2	Brand 3	Total	Brand 1	Brand 2	Brand 3	Total
Brand 1	374 (91%)	12	27	413	335 (81%)	32	46	413
Brand 2	43	154 (74%)	10	207	76	90 (43%)	41	207
Brand 3	52	16	173 (72%)	241	94	28	119 (49%)	241
Total	469	182	210	861	505	150	206	861

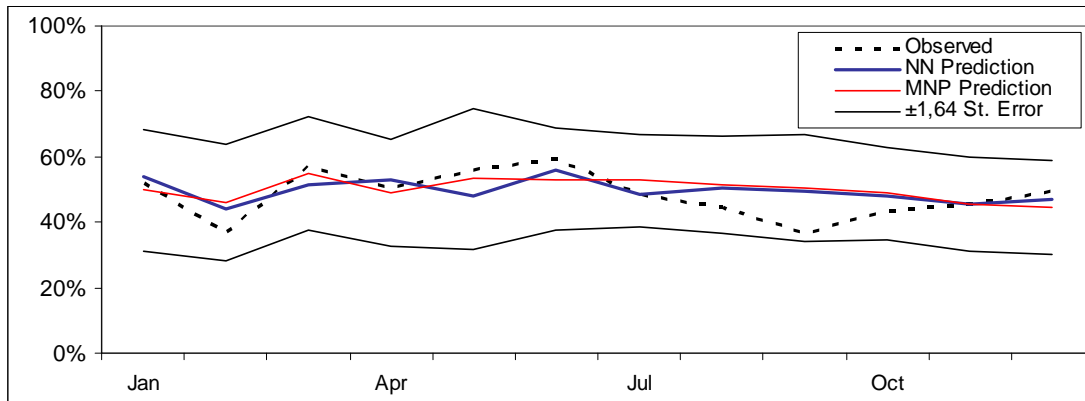


Figure 5 Predictions of Brand 1 monthly share of purchases within the holdout sample

4.5.A comparison of two models in terms of predictive performance

This section summarizes the performances of MNP and ANN models in terms of hit rates and market share estimation. Table 5 shows that the ANN model is successful at predicting the purchases of all three brands on the test sample (a hit rate of 81% on the average). On the other hand, MNP model is only good at predicting Brand 1 purchases. Table 5 shows that MNP can only correctly predict 43% and 49% of Brand 2 and Brand 3 purchases, respectively. Although these rates are still nearly two times higher than the purchase shares

(expected choice probability) of Brand 2 and Brand 3, the hit rate performance of the ANN model seems undoubtedly better on the test sample as well.

Monitoring the performances of the models throughout the whole study period is another essential issue since the typical toothpaste data may easily be affected by temporary marketing activities. Figures 5-7 show how ANN and MNP models predict the monthly shares of purchases for Brand 1, Brand 2, and Brand 3 respectively within the test sample throughout the twelve month period.

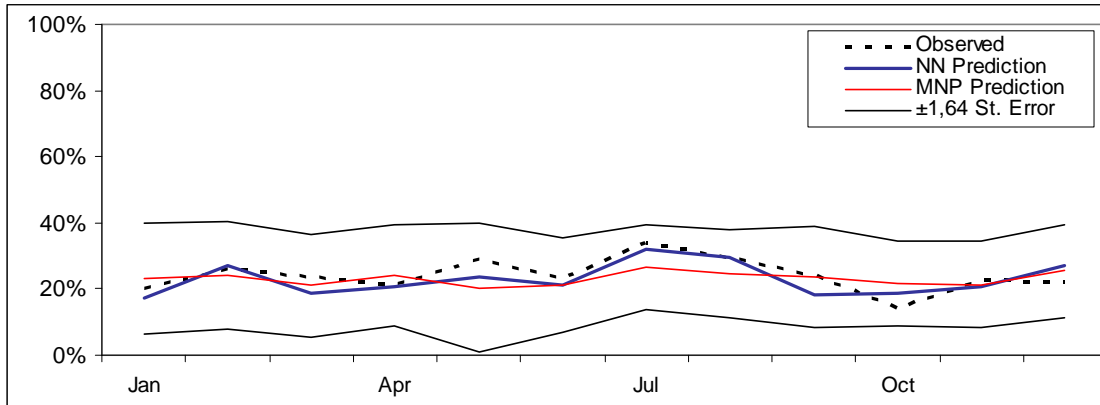


Figure 6 Predictions of Brand 2 monthly share of purchases within the holdout sample

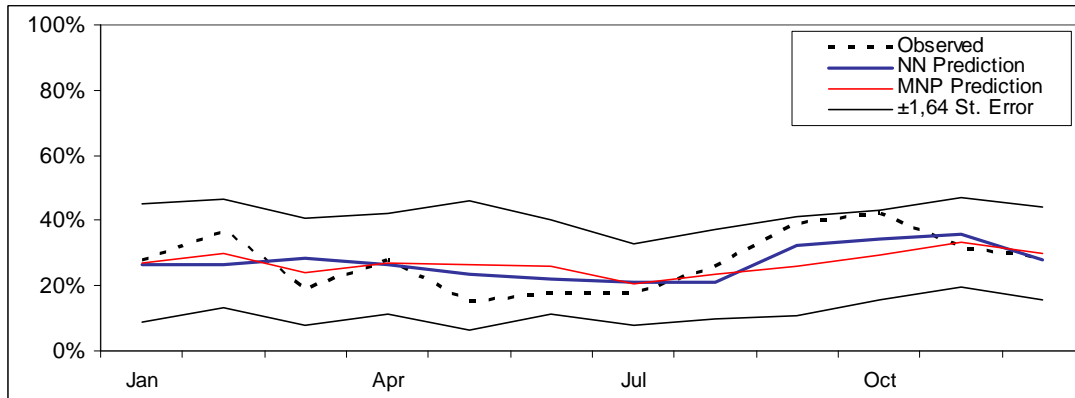


Figure 7 Predictions of Brand 3 monthly share of purchases within the holdout sample

Eq. (6) is used when calculating the predicted shares of purchases and the standard errors. The numbers of brand choices were divided by the total number of category purchases when generating the actual share of purchases. Each figure also includes a band of ± 1.64 SE confidence interval. Figures 5-7 reveal that, the predictions made by the models, showing a good performance, do not exceed a band of 90 % confidence within the test sample.

Finally, based on the monthly predictions of ANN and MNP models, mean absolute distances from the observed market shares are computed. Figure 8 gives a comparative summary of the mean absolute distance figures computed for the test sample.

As it is seen in Figure 8, mean absolute distances from the observed monthly market shares are smaller for neural network model for each of the three brands when compared to MNP.

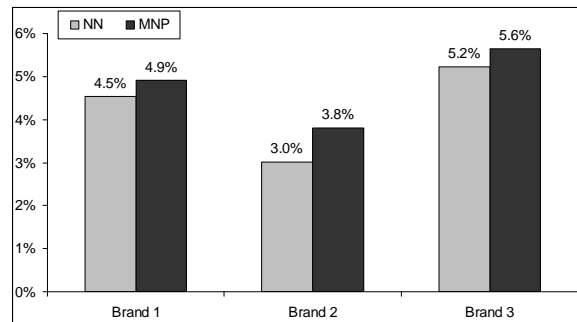


Figure 8 Mean absolute distances between observed and predicted shares

In other words, ANN model’s monthly prediction performance is better than MNP model for each of the brands within the test sample throughout the twelve month period.

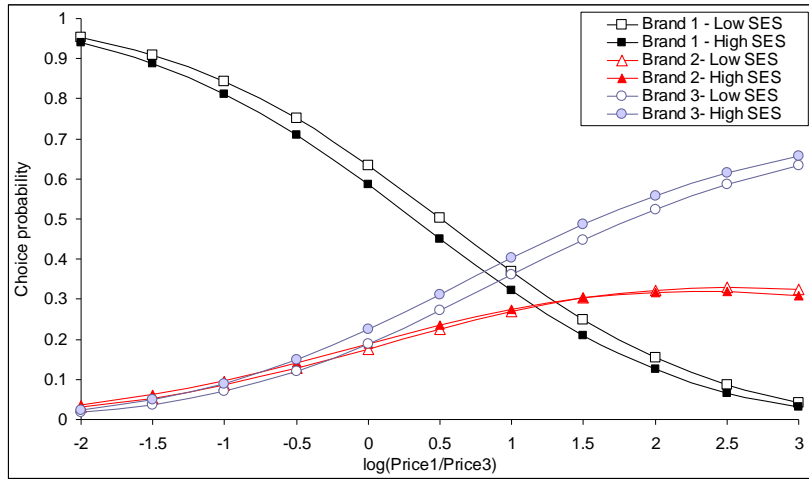


Figure 9 The impact of changes in the relative price of Brand 1 over 3 on the choice probabilities

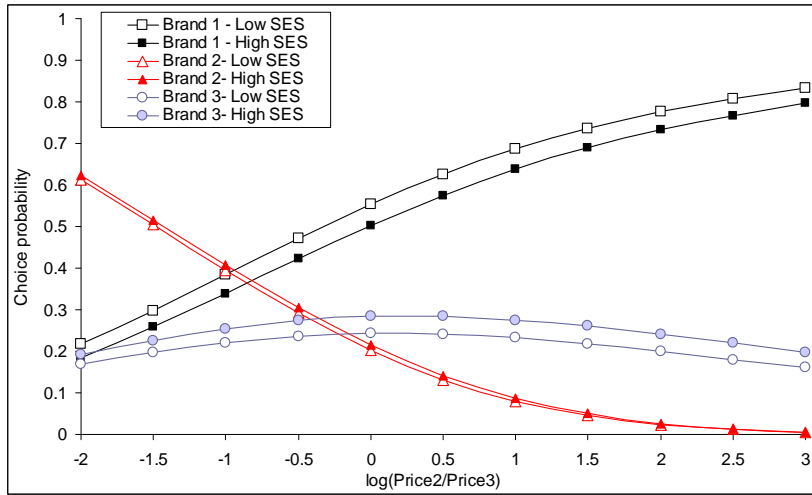


Figure 10 The impact of changes in the relative price of Brand 2 over 3 on the choice probabilities

4.6. Sensitivity Analysis

Providing answers to questions like ‘what if our brand’s price increases 5%?’ or ‘to which competitor and in which segments will we lose market share’, statistical brand choice models can be very useful for marketing managers in practical issues. Figure 9 shows how the probability of being chosen for Brand 1, 2, and 3 change within different SES levels, while the relative price of 1 over 3 ($\log(\text{Price } 1/\text{Price } 3)$) increases. While performing these marginal effect analyses, it is assumed that loyalty, household size, and $\log(\text{Price}2/\text{Price}3)$ variables take the average values.

Figure 9 shows that, as the relative price of Brand 1 over 3 increases, the probability of being chosen for Brand 1 decreases within both SES groups. Figure 9 also reveals that, as the relative price of Brand 1 over 3 increases, the probability of being chosen for Brand 2 slightly increases. Finally, as the relative price of Brand 1 over 3 increases the choice probability of Brand 3 increases within both SES groups. Similarly, Figure 10 shows how the choice probabilities of the brands change as the price of Brand 2 over Brand 3 increases. Although MNP is not very successful in terms of hit rates and monthly predictions, sensitivity analyses show

that it may still be very useful in providing insight about the competitive structure of a product category.

5. Concluding Remarks

Scanner panel data lets researchers study the effects of marketing variables on customer choice among product alternatives since the mid-1980s. Many dozens of papers are published on the brand and retailer choice issues in developed markets since then. The aim of this paper was to make an empirical comparison of ANN and MNP approaches in modeling the brand choice decision in Turkish fast moving consumer goods sector. In order to accomplish this goal, firstly, Turkish toothpaste market is selected for empirical study. Then to maintain price inputs, artificial unit prices are generated for the products which were present at the retailer during the shopping trips but not purchased. Thirdly, ANN and MNP models of brand choice are built based on 2597 toothpaste purchase occasions of a model sample of 404 households. Finally, the performances of the models are tested on the 861 purchases of a randomly selected holdout sample of 135 households.

Results show that ANN provides higher hit-rates and better market share predictions on monthly base. The prediction performance of the ANN model is significantly higher in the brands with smaller market shares. The main reason that ANN gives better results than MNP is the fact that ANNs are able to handle the nonlinearities in the data as explained by Hruschka⁴³. In most cases, this provides better predictions. Since we use household panel data which may contain significant nonlinearities, in our case, ANN must have acquired these nonlinearities better.

On the other hand, we should not conclude that ANN always provides higher hit-rates and better predictions. There are examples of case studies in the literature which show that discrete choice models may outperform ANNs.^{23,46-47} Moreover, statistical models are advantageous in providing insights about the nature of the markets studied. In particular, this study shows that MNP can provide useful insights about the nature of Turkish toothpaste market. Based on the coefficients of the estimated MNP model, simulation examples which evaluate the effects of price changes on brand choice within different socio-economical market segments are given.

For future research, findings of this study can be compared with the results of the studies which use MNL or discriminant analysis modeling methodologies. Moreover, using ANN and MNP, new studies can be conducted based on panel data collected for other FMCG categories like food, snacks, or soft drinks.

References

1. J. W. Payne, J. R. Bettman and E. J. Johnson, *The adaptive decision maker*. (Cambridge University Press Cambridge, 1993).
2. R. S. Moore, M. B. Beauchamp, C. Z. Barnez, and C. A. Stammerjohan, Artificial Neural Networks: An Introduction and Application in a Consumer Behavior Setting, *Marketing Management Journal*, Spring, (2007) 27-38.
3. D. Aggrawal and C. Schorling, Market Share Forecasting: An Empirical Comparison of Artificial Neural Networks and Multinomial Logit Model, *Journal of Retailing*, **72** (4) (1996) 383-407.
4. P. Kennedy, *A Guide to Econometrics*, 5th edn. (MIT Press, 2003).
5. Y. Bentz and D. Merunka, Neural Networks and the Multinomial Logit for Brand Choice Modelling: a Hybrid Approach, *Journal of Forecasting*, **19** (2000) 177-200.
6. S. Gupta, Impact of Sales Promotions on When, What and How Much to Buy, *Journal of Marketing Research*, November, (1988) 342-355.
7. R. E. Bucklin and S. Gupta (1999) Commercial Use of UPC Scanner Data: Industry and Academic Perspectives, *Marketing Science*, Vol. 18, No: 3, pp. 247-273.
8. N. F. Matsatsinis and N. M. More, A Knowledge Base for Brand Choice Model Selection, in *Proc. 2nd European Symposium on Intelligent Techniques*, (Crete, Greece, 1999).
9. P. M. Guadagni and J. D. C. Little, A Logit Model of Brand Choice Calibrated on Scanner Data, *Marketing Science*, **2** (3) (1983) 203-238.
10. P. M. Guadagni and J. D. C. Little, When and What to Buy: A Nested Logit Model of Coffee Purchase, *Journal of Forecasting*, **17** (1998) 303-326.
11. R. E. Bucklin, S. Gupta, and S. Siddharth, Determining Segmentation in Sales Response across Consumer Purchase Behaviors, *Journal of Marketing Research*, May **35** (1998) 189-197.
12. S. A. Neslin, C. Henderson, and J. Quelch, Consumer Promotions and the Acceleration of Product Purchases, *Marketing Science*, **4** (2) (1985) 147-165.
13. C. F. Mela, S. Gupta, and D. R. Lehmann, The Long Term Impact of Promotion and Advertising on Consumer Brand Choice, *Journal of Marketing Research*, May **34** (1997) 248-261.
14. C. F. Mela, K. Jedidi, and D. Bowman, The Long Term Impact of Promotions on Consumer Stockpiling

- Behaviour, *Journal of Marketing Research*, May **35** (1998) 250-262.
15. K. Jedidi, C. F. Mela, and S. Gupta, Managing Advertising and Promotion for Long Run Profitability, *Marketing Science*, **18** (1) (1999) 1-22.
 16. K. Pauwells, D. M. Hanssens, and S. Siddarth, The Long Term Effects of Price Promotions on Category Incidence, Brand Choice, and Purchase Quantity, *Journal of Marketing Research*, Nov **39** (4) (2002) 421-439.
 17. D. Klapper, C. Ebling, and J. Temme, Another Look at Loss Aversion in Brand Choice Data: Can We Characterize the Loss Averse Consumer?, *International Journal of Research in Marketing*, **22** (3) (2005) 239-254.
 18. J. M. Silva-Risso and R. E. Bucklin, Capturing the Effects of Coupon Promotions in Scanner Panel Choice Models, *Journal of Product & Brand Management*, **13** (6) (2004) 442-452.
 19. H. J. Van Heerde, E. Gijsbrechts, and K. Pauwels, Winners and Losers in a Major Price War, *Journal of Marketing Research* October **45** (5)(2008) 499-518.
 20. M. Villas-Boas and M. Zhao, Retailer, Manufacturers, and Individual Consumers: Modeling the Supply Side in the Ketchup Marketplace, *Journal of Marketing Research* Feb **42** (1) (2005) 83-95.
 21. M. Villas-Boas and S. Berto, Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data, *Review of Economic Studies* Apr **74** (2007) 625-652.
 22. P. M. West, P. L. Brockett, and L. L. Golden A Comparative Analysis of Neural Networks and Statistical Methods for Predicting Consumer Choice, *Marketing Science*, **16** (4) (1997) 370-391.
 23. H. Hruschka, W. Fettes, M. Probst, and C. Mies, A flexible brand choice model based on neural net methodology: A comparison to the linear utility multinomial logit model and its latent class extension, *OR Spectrum*, **24** (2002) 127-143.
 24. H. Hruschka, W. Fettes, and M. Probst, An empirical comparison of the validity of a neural net based multinomial logit choice model to alternative model specifications, *European Journal of Operational Research* **159** (2004) 166-180.
 25. M. Y. Hu, and C. Tsoukalas, Explaining consumer choice through neural networks: The stacked generalization approach, *European Journal of Operational Research* **146** (2003) 650-660.
 26. K. E. Fish, J. D. Johnson, R. E. Dorsey, and J. G. Blodgett, Using an artificial neural network trained with a genetic algorithm to model brand share, *Journal of Business Research* **57** (2004) 79- 85.
 27. B. Vroomen, P. H. Franses, and E. van Nierop Modeling Consideration Sets and Brand Choice using Artificial Neural Networks, *European Journal of Operational Research* **154** (2004) 206-217.
 28. H. Hruschka, Using a Heterogeneous Multinomial Probit Model with a Neural Net Extension to Model Brand Choice, *Journal of Forecasting*, **26** (2007) 113-127.
 29. J. K. Dow, and J. W. Endersby, Multinomial Probit and Multinomial Logit: A Comparison of Choice Models for Voting Research, *Electoral Studies*, **23** (2004) 107-122.
 30. J. Hausman and D. McFadden, Specification Tests for the Multinomial Logit Model, *Econometrica*, **52** (5) (1984) 1219-1240.
 31. P. H. Franses and R. Paap, *Quantitative Models in Marketing Research*, (Cambridge University Press, Cambridge, 2001).
 32. W. H. Greene, *Econometric Analysis*, 5th edn. (Prentice Hall, New Jersey, 2003).
 33. G. S. Maddala, *Limited-dependent and Qualitative Variables in Econometrics*, (Cambridge University Press, Cambridge, 1983)
 34. J. Lattin, J. D. Carroll, and P. E. Green, *Analysis of Multivariate Data*, (Duxbury Applied Series, Canada, 2003).
 35. E. G. Tsionas, P. G. Michaelides, and A. T. Vouldis, Global Approximations to Cost and Production Functions using Artificial Neural Networks, *International Journal of Computational Intelligence Systems*, **2** (2), (2009) 132 – 139.
 36. E. Zio, Neural Networks Simulation of the Transport of Contaminants in Groundwater, *International Journal of Computational Intelligence Systems*, **2** (3), (2009) 267 – 276.
 37. Gowrishankar, P.S. Satyanarayana, Neural Network Based Traffic Prediction for Wireless Data Networks, *International Journal of Computational Intelligence Systems*, **1** (4), (2008) 379 – 389.
 38. Y. Yoon, G. Swales, and T. M. Margavio, A comparison of discriminant analysis versus artificial neural networks. *Journal of Operational Research* **44** (1) (1993) 51-60.
 39. M. Boznar, M. Lesjak, and P. Mlakar, A neural network-based method for short-term predictions of ambient SO₂ concentrations in highly polluted industrial areas of complex terrain. *Atmospheric Environment, B Urban Atmosphere* **27B** (1993) 221-230.
 40. N.R. Swanson and H. White, Forecasting economic time series using flexible versus fixed specification and linear versus nonlineareconometric models. *International Journal of Forecasting* **13** (4) (1997) 439-461.
 41. H.B. Hwang and H.T. Ang, A simple neural network for ARMA (p,q) time series. *Omega* **29** (2001) 319-333.
 42. E. Gately, *Neural Networks for Financial Forecasting*, (John Wiley & Sons, USA, 1996).
 43. H. Hruschka, Determining market response functions by neural network modeling: a comparison to econometric techniques, *European Journal of Operational Research*, **66** (1) (1993) 27-35.
 44. M. F. Moller, A scaled conjugate gradient algorithm for fast supervised learning, *Neural Networks*, **6** (1993) 525-533.

45. R. Matignon, *Neural Network Modeling using SAS Enterprise Miner*. (SAS Institute Inc., 2005).
46. D. A. Henscher and T. T. Ton, A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice, *Transportation Research Part E*, **36** (2000) 155-172.
47. T. B. Bell, Neural Nets or the Logit Model? A Comparison of Each Model's Ability to Predict Commercial Bank Failures, *Intell. Sys. Ace. Fin. Mgmt.*, **6** (1997) 249-264.