Report Series

An Artificial Neural Net Attraction Model (ANNAM) to Analyze Market Share Effects of Marketing Instruments

Harald Hruschka

Report No. 48
November 2000

Report Series

November 2000

SFB
'Adaptive Information Systems and Modelling in Economics and Management Science'

Vienna University of Economics
and Business Administration
Augasse 2–6, 1090 Wien, Austria

in cooperation with
University of Vienna
Vienna University of Technology

http://www.wu-wien.ac.at/am

This paper was accepted for publication in:
Schmalenbach Business Review

This piece of research was supported by the Austrian Science Foundation (FWF) under grant SFB#010 ('Adaptive Information Systems and Modelling in Economics and Management Science').

Abstract

Attraction models are very popular in marketing research for studying the effects of marketing instruments on market shares. However, so far the marketing literature only considers attraction models with certain functional forms that exclude threshold or saturation effects on attraction values. We can achieve greater flexibility by using the neural net based approach introduced here. This approach assesses brands' attraction values by means of a perceptron with one hidden layer. The approach uses log-ratio transformed market shares as dependent variables. Stochastic gradient descent followed by a quasi-Newton method estimates parameters. For store-level data, neural net models perform better and imply a price response that is qualitatively different from the well-known multinomial logit attraction model. Price elasticities of neural net attraction models also lead to specific managerial implications in terms of optimal prices.

1 Introduction

Marketing activities that change a brand's sales may affect both sales volume (i.e., total sales of all brands considered) and market shares of the brand's competitors. For example, a price decrease of a brand might increase sales volume and also decrease market shares of competing brands. Therefore, a complete understanding of the effects of marketing activities requires that the researcher distinguishes effects on sales volume and effects on market shares[1].

Attraction models are very popular in marketing research for studying the effects of marketing instruments on market shares. Attraction models are derived from the Market Share Theorem of *Bell et al.* (1975) which starts from the following assumptions:

- Each brand has an attraction.
- Attractions are non-negative and their sum is greater than zero.
- A brand with an attraction equal to zero has a market share equal to zero.
- Brands with equal attractions have equal market shares.
- The market share of a brand is affected in the same manner if the attraction of any other brand changes by a fixed amount.

  The last assumption means that if there is a change in the attraction level of any competitor the new market share of a brand does not depend on which competitor made this change.

The theorem says that the market share MS_{it} of brand i is the ratio of this brand's attraction A_{it} to the sum of attractions $A_{jt}, j = 1, J, t = 1, T$ of all J brands (including brand i) constituting a market (t denotes the observation period):

$$MS_{it} = f(A_{it}, sum) = \frac{A_{it}}{\sum_j A_{jt}} \tag{1}$$

[1] For a more detailed discussion see, for example, *Hanssens et al.* (1990).

Attraction models are logically consistent[2] in the sense that they satisfy the sum constraint $\sum_{j=1}^{J} MS_{jt} = 1$ and range constraints $0 \leq MS_{jt} \leq 1$ for all j and t.

As alternative to attraction models, the researcher could estimate sales by means of brand specific sales response functions and compute estimated market shares on the basis of these sales values. But this approach does not lead to a complete understanding of sales effects, because it confounds sales volume effects and market share effects.

This paper deals with differential effects attraction models which are characterized by two properties:

1. Coefficients for all predictors are brand-specific (i.e. not the same across brands).
2. Only a brand's own marketing instruments influence its attraction value. Marketing instruments of other brands have no effect on a brand's attraction value.

The marketing literature hitherto only considers attraction models with certain functional forms[3]. I introduce a more flexible neural net based approach that preserves logically consistency.

One can find some examples for estimating aggregate market share response functions by using artifical neural nets (i.e. multilayer perceptrons) and applying some variant of back-propagation[4]. The neural nets used in these contributions ignore the logical consistency issue mentioned above, especially the sum constraint. They are nonlinear nonparametric regression models that do without attraction values intervening between predictors and market shares. Among these papers only the one of *Gaul et al.* (1994) compares neural nets of this type to attraction models. In their study, Gaul et al. show that neural net models perform better in terms of relative absolute errors, but differences of model complexity are not considered[5].

Section 2 discusses the artificial neural net attraction model. Section 3 deals with estimation and model evaluation methods. Section 4 contains estimation results (i.e. model performance, price effects) of an empirical study using store-level data. The final section emphasizes managerial implications (i.e., price elasticities and optimal prices) which I obtained for the different models.

[2]e.g. *Naert/Bultez* (1973); *McGuire et al.* (1977).

[3]e.g. *Nakanishi/Cooper* (1974); *Naert/Bultez* (1973); *Bultez/Naert* (1975); *McGuire et al.* (1977); *Leeflang/Reuyl* (1984); *Cooper/Nakanishi* (1988); *Abeele et al.* (1990); *Cooper* (1993); *Chen et al.* (1994); *Houston et al.* (1994).

[4]e.g. *van Wezel/Baets* (1995); *Wierenga/Kluytmans* (1996); *Gaul et al.* (1994); *Natter/Hruschka* (1998).

[5]For one of the product groups analyzed the number of parameters of the neural net is 6 times the number of parameters of the attraction model.

2 Artificial Neural Net Attraction Model

According to the well-known differential effects multinomial logit attraction model[6] with x_{pit} as brand's i p-th predictor ($p = 1, P$) in period t and normally distributed errors ϵ_{it} having zero mean and constant variance a brand's attraction is:

$$A_{it} = \exp(\sum_p a_{pi} x_{pit} + \epsilon_{it}) \tag{2}$$

It seems obvious that a brand's attraction may be subject to threshold effects (e.g. the attraction changes only after a marketing instrument is above or below a certain value) or saturation effects (e.g. the attraction does not change if a marketing instrument is above a certain value). These effects are shown by some cognitive studies on price response[7]. In their experimental study, *Gupta/Cooper* (1992) found both threshold and saturation effects. Of course, these effects cannot be reproduced by the multinomial logit attraction model, because this model assumes that attraction is an exponential function of linearly combined predictors.

Therefore, we generalize the multinomial logit attraction model to an appropriate artificial neural net. This artificial neural net is guaranteed to approximate any continuous multivariate function with desired precision given a sufficient number of hidden units[8]. Thus, the artificial neural net attraction model can uncover threshold or saturation effects on attraction values. Algebraically, the first part of the neural net corresponds to the exponential attraction of a multinomial logit attraction model, the second part constitutes the flexible extension:

$$A_{it} = \exp(\sum_p a_{pi} x_{pit} + \sum_{k=1}^{K_i} b_{ki} h_{kit} + \epsilon_{it}) \tag{3}$$

The second part of an attraction equals a multilayer perceptron (which is the most widespread type of artificial neural net) with one layer of K_i hidden units having values h_{kit}. Hidden units are brand-specific, K_i symbolizes the number of hidden units of brand i. Values of hidden units are computed by plugging a linear combination of brand-specific predictors into the binomial logistic function $h()$:

$$h_{kit} = h(-\sum_p c_{pki} x_{pit}) = 1/(1 + \exp(-\sum_p c_{pki} x_{pit})) \tag{4}$$

Please note that the conventional multinomial logit attraction model is just a special case of the neural net which we obtain if no hidden units are specified (i.e. $K_i = 0$ for all brands). Therefore, this approach allows to decide on the usefuleness of the artificial neural net generalization compared to a conventional multinomial logit attraction model.

[6]e.g. *Cooper* (1993).
[7]e.g. *Monroe* (1973).
[8]e.g. *Cybenko* (1989); *Hornik et al.* (1989); *Ripley* (1993).

3 Model Estimation and Evaluation

Models are estimated by minimizing the error sum of squares E of log ratios[9] of market shares:

$$E = \frac{1}{2} \sum_t \sum_{i>1} (Y_{it} - \hat{Y}_{it})^2 \qquad (5)$$

$\hat{Y}_{it}$ for $i = 2, \cdots, I$ symbolizes the estimated log ratio of brand i in period t.

Because of its linear form, parameters of the multinomial logit attraction model are estimated by ordinary least squares. Estimation of neural net models consists of two stages[10], stochastic gradient descent and the quasi-Newton optimization procedure BFGS of *Broyden, Fletcher, Goldfarb, Shanno*[11].

Besides the conventional multinomial logit attraction model I estimate the following 46 neural net models:

- models with zero or one hidden unit per brand and at least one brand specific hidden unit;
- models with one or two hidden units per brand and at least two brand specific hidden units;
- models with two or three hidden units per brand and at least three brand specific hidden units;
- the model with four hidden units for each of the brands.

Instead of selecting a single model, I evaluate all estimated artificial neural net models by (approximate) posterior probabilities[12]. This way, I account for the uncertainty involved in model selection. I then average results over models according to their posterior probabilities[13].

4 Empirical Study

The empirical study analyzes store-level data of four brands (A, B, C, D) of a certain category of consumer non-durables. The data base consists of 104 weekly observations per brand on market shares, current retail prices and features (binary).

The neural net model with one hidden unit for each of the brands A and B, but no hidden unit for either Brand C or D attains an approximate probability of 0.99998. It dominates the remaining 45 neural net models, most of which have posterior probabilities

[9] Appendix A explains the log ratio transformation of market shares.
[10] Details are given in Appendix B.
[11] e.g. *Seber/Wild* (1989); *Bishop* (1995).
[12] See Appendix C.
[13] e.g. *Carlin/Louis* (1996).

less than 10^{-5} [*see tables 1 and 2*]. Please note that the number of parameters increases by a factor of 2/3 in relation to the conventional logit model. This is rather modest compared to the neural net models estimated by *Gaul et al.* (1994). Because of negligible posterior probabilities for models with at least two hidden units per brand, I do not estimate neural net models with at least three hidden units per brand except for the model having four hidden units for each of the brands.

==============================
put tables 1 and 2 about here
==============================

Figure 1 shows a path diagram that depicts the relations between predictors, hidden units, attraction values, and market shares for the dominant model. Both the condition number[14] of the Hessian[15], which amounts to 14.11, and the minimal absolute t-value of 12.10 over all parameters suggest that we can rule out multicollinearity or ill-conditioning for this model.

==============================
put figure 1 about here
==============================

Figure 2 contains plots of market shares for both the multinomial logit and the neural net models for each of the four brands versus its own price (given average prices of the other three brands, respectively). The plotted market shares for the artificial neural net models are weighted averages across all 46 models according to their posterior probabilities.

For brand A, the neural net models indicate a weaker marginal price response than the multinomial logit model except at very low prices. For brand B, neural net models imply a weaker marginal reponse, and only at very high prices marginal effects greater than those for the multinomial logit model. Obviously, the prices of brands A and B are subject to threshold effects (market share changes become more pronounced only if prices are below or above a certain level). This explains why neural net models perform much better than the multinomial logit model which is unable to reproduce such effects.

==============================
put figure 2 about here
==============================

For brands C and D, marginal effects do not differ between the multinomial logit and the neural net models. This result is not surprising, because the dominant neural net model possesses no hidden units for these two brands and in this respect is similar to the multinomial logit model.

[14] *e.g. Belsley et al. (1980).*

[15] Its computation based on a linear approximation in the neighborhood of the estimated parameter values can be found in *Seber/Wild* (1989).

5 Managerial Implications

I determine (cross-) elasticities for the models studied. From equation 1 I obtain as first derivative w.r.t. to predictor x_{pjt}:

$$\frac{\partial MS_{it}}{\partial x_{pjt}} = MS_{it}(\delta_{ij} - MS_{jt})\frac{\partial Y_{jt}}{\partial x_{pjt}} \tag{6}$$

δ_{ij} denotes Kronecker's delta which is equal to one for brand's i predictors ($j = i$), equal to zero for another brand's predictors ($j \neq i$).

We get the same derivatives for predictors of the reference brand $\partial Y_{jt}/\partial x_{p1t} = -\partial \log(A_{1t})/\partial x_{p1t}$ for $j = 2, \cdots, J$.

Using expression 6 market share (cross-) elasticities el_{it} of brand i in period t w.r.t. predictor x_{pjt} can be written:

$$el_{ijt} = (\delta_{ij} - MS_{jt})x_{pjt}\frac{\partial Y_{jt}}{\partial x_{pjt}} \tag{7}$$

As they are in the multinomial logit attraction model, the cross-elasticities for all brands $j \neq i$ are equal.

Substituting for $\partial log(A_{jt})/\partial x_{pjt}$ the expression for the neural net's (cross-) elasticities is:

$$el_{ijt} \quad = \quad (\delta_{ij} - MS_j)x_{pjt}(a_{pj} + \sum_{k=1}^{K_j} b_{kj}c_{pkj}h_{kjt}(1 - h_{kjt})) \tag{8}$$

It subsumes the well-known equation for the multinomial logit attraction model[16] as a special case:

$$el_{ijt} \quad = \quad (\delta_{ij} - MS_j)x_{pjt}a_{pj} \tag{9}$$

Table 3 contains elasticities and cross-elasticities for the average price of each brand. The prices of the other brands are also set to average values (e.g., the price elasticity for brand A is computed at the average price of 42.18 and the prices of the other brands B, C and D are set to 36.22, 35.14 and 40.10, respectively). The elasticities and cross-elasticities for artificial neural nets are model averages (i.e. elasticities of the individual 46 models weighted by their posterior probabilities). Contrary to the multinomial logit attraction model, the neural net models indicate lower absolute values for both elasticities and cross-elasticities[17].

============================

put table 3 about here

============================

[16] e.g. *Cooper/Nakanishi* (1988); *Cooper* (1993).

[17] These results are not due to systematic changes of prices over time. Time-dependent regression models with different functional forms (linear, quadratic, exponential, double log and semi log) explain maximally 14.76 %, 2.08 %, 7.04 % and 5.92 % of the variance in prices for each of the four brands, respectively.

For our data set the multinomial logit model misleads a brand manager into overestimating the effect of her/his own price changes as well as those of competitors, on market share. To give more insight into managerial implications, I determine optimal prices. To this end, I assume constant marginal costs that are equal for all brands. Because of estimation results for several parametric and non-parametric models, sales volume is determined by a multiplicative function of the average price of the product category (across the four brands studied) in each week. As solution concept that deals with competitive behavior, I use fictitious play, which assumes that competitors set prices following the frequency distributions observed in the past[18].

==========================
put table 4 about here
==========================

Table 4 contains optimal prices and profits based both on the MNL model and the neural net models (optimal prices for neural net models maximize average profits weighted by models' posterior probabilities), model averaged profits if optimal prices of the MNL model are chosen and profit increases by setting prices in accordance with neural net models. Prices recommended on the basis of neural net models are higher. This is to be expected, since the estimated neural net models imply lower elasticities, as shown above. These differences have important practical implications for brands A and B, for which the dominant neural net model has one hidden unit and therefore differs from a conventional multinomial logit model, but negligible for brands C and D, the brands without hidden units. On the basis of neural net models the profit increases we can expect for brands A and B amount to 10.70 % and 15.61 %, respectively.

Conclusions

The empirical study based on store-level data demonstrates that the proposed neural net models perform better in terms of posterior probability. Neural net models imply a price response qualitatively different from the well-known multinomial logit attraction model. Price elasticities also differ from those for the multinomial logit model. Moreover, neural net models lead to specific managerial implications in terms of optimal prices and profits.

Like the multinomial logit model, the neural net models studied in this paper are subject to the IIA property, i.e. they assume that from the viewpoint of any brand all its competitors are equally substitutable (in terms of cross-elasticities). Therefore, an interesting topic of future work might be extensions of these models allowing marketing instruments of other brands to influence any brand's attraction value.

[18]e.g. *Brown* (1951).

References

Abeele, Piet Vanden/Gijsbrechts, Else/Vanhuele, Marc (1990), Specification and Empirical Evaluation of a Cluster-Asymmetry Market Share Model, in: International Journal of Research in Marketing, Vol. 7, pp. 223–247.

Bell, David E./Keeney, Ralph L./Little, John D.C. (1975), A Market Share Theorem, in: Journal of Marketing Research, Vol. 12, pp. 136–141.

Belsley, David A./Kuh, Edwin/Welsch, Roy E. (1980), Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.

Bishop, Christopher M. (1995), Neural Networks for Pattern Recognition.

Brown, George W. (1951): Iterative Solutions of Games by Fictitious Play, in: Koopmans, Tjalling C. (ed.), Activity Analysis of Production and Allocation, pp. 374–376.

Bultez, Alain V./Naert, Philippe A. (1975), Consistent Sum-Constrained Models, in: Journal of the American Statistical Association, Vol. 70 , pp. 529–535.

Carlin, Bradley P./Louis, Thomas A. (1996), Bayes and Empirical Bayes Methods for Data Analysis.

Chen, Youhua/Kanetkar, Vinay/Weiss, Doyle L. (1994), Forecasting Market Shares with Disaggregate or Pooled Data: A Comparison of Attraction Models, in: International Journal of Forecasting, Vol. 10, pp. 263–276.

Cooper, Lee G. (1993), Market-Share Models, in: Eliashberg, Jehoshua/Lilien, Gary L. (eds.), Marketing, pp. 259–314.

Cooper, Lee G./Nakanishi, Masao (1988), Market-Share Analysis.

Cybenko, George (1989), Continuous Value Neural Networks with Two Hidden Layers are Sufficient, in: Mathematics of Control, Signal and Systems, Vol. 2, pp. 303–314.

Gaul, Wolfgang/Decker, Reinhold/Wartenberg, Frank (1994), Analyse von Panel- und POS-Scanner-Daten mit Neuronalen Netzen in: Jahrbuch der Absatz- und Verbrauchsforschung, Vol. 40, pp. 281–306.

Gupta, Sunil/Cooper, Lee G. (1992), The Discounting of Discounts and Promotion Thresholds, in: Journal of Consumer Research, Vol. 19, pp. 401–411.

Hanssens, Dominique M./Parsons, Leonard J./Schultz, Randall L. (1990), Market Response Models. Econometric and Time Series Analysis.

Hertz, John/Krogh, Anders/Palmer, Richard G. (1991), Introduction to the Theory of Neural Computation.

Hornik, Kurt/Stinchcombe, Max/White, Halbert (1989), Multilayer Feedforward Networks are Universal Approximators, in: Neural Networks, Vol. 3, pp. 359–366.

Houston, Franklin S./Kanetkar, Vinay/Weiss, Doyle L. (1994), Cross Effects Share Model: Identification and Optimality. Working Paper, University of Toronto.

Leeflang, Peter S.H./Reuyl, Jan C. (1984), On the Predictive Power of Market Share Attraction Models, in: Journal of Marketing Research, Vol. 21, pp. 211–215.

McGuire, Timothy W./Weiss, Doyle L./Houston, Franklin S. (1977), Consistent Multiplicative Market Share Models, in: Greenwood, B.A./Bellinger, Danny N. (eds.), Contemporary Marketing Thought, pp. 129–134.

Monroe, Kent B. (1973), Buyers' Subjective Perceptions of Price, in: Journal of Marketing Research, Vol. 10, pp. 70–80.

Naert, Philippe A./Bultez, Alain V. (1973), Logically Consistent Market Share Models, in: Journal of Marketing Research, Vol. 10, pp. 334–40.

Nakanishi, Masao (1972), Measurement of Sales Promotion Effect at the Retail Level. AMA Proceedings, Spring and Full Conference, pp. 338–343.

Nakanishi, Masao/Cooper, Lee G. (1974), Parameter Estimation for a Multiplicative Competitive Interaction Model – Least Squares Approach, in: Journal of Marketing Research, Vol. 11, pp. 303–311.

Natter, Martin/Hruschka, Harald (1998), Using Artificial Neural Nets to Specify and Estimate Aggregate Reference Price Models, in: Aurifeille, Jacques-Marie/Deissenberg, Christophe (eds.), Bio-Mimetic Approaches in Management Science, pp. 101–118.

Ripley, Brian D. (1993), Statistical Aspects of Neural Networks, in: Barndorff-Nielsen, Ole E./Jensen, Johan L./Kendall, Wilfried S. (eds.), Networks and Chaos – Statistical and Probabilistic Aspects, pp. 40–123.

Ripley, Brian D. (1996), Pattern Recognition and Neural Networks.

Saito, Kazumi/Nakano, Ryohei (1997), Partial BFGS Update and Efficient Step-Length Calculation for Three-Layer Neural Networks, in: Neural Computation, Vol. 9, pp. 123–141.

Schwarz, Gideon (1979), Estimating the Dimension of a Model, in: Annals of Statistics, Vol. 6, pp. 461–464.

Seber, George A.F./Wild, C.J. (1989), Nonlinear Regression.

van Wezel, Michiel C./Baets, Walter R.J. (1995), Predicting Market Responses with a Neural Network, in: Marketing Intelligence & Planning, Vol. 13, Nr. 7, pp. 23–30.

Wierenga, Berend/Kluytmans, Jack(1996), Prediction with Neural Nets in Marketing Time Series Data. Working Paper, School of Management, Erasmus Universiteit Rotterdam.

Appendix A: Log Ratio Transformation of Market Shares

Estimation of attraction models is simplified by applying the so-called log ratio transformation[19] which is equivalent to the well-known log-centering transformation developed by *Nakanishi* (1972) as well as *Cooper/Nakanishi* (1974).

Taking the log of equation 1 gives:

$$\log(MS_{it}) = \log(A_{it}) - \log(\sum_{j=1}^{J} A_{jt}) \tag{10}$$

Without loss of generality I take brand 1 as reference and subtract $\log(MS_{1t})$ from equation 10. This leads to:

$$Y_{it} \equiv \log(MS_{it}) - \log(MS_{1t}) = \log(A_{it}) - \log(A_{1t}) \tag{11}$$

Y_{it}, the log ratio of market share of brand i in period t, serves as dependent variable in our regression models. Forming the antilog of Y_{it} results in:

$$\exp(Y_{it}) = \frac{A_{it}}{A_{1t}} \tag{12}$$

Dividing both numerator and denominator of equation 1 by A_{1t} and substituting shows how to compute market shares on the basis of log ratios Y_{it}:

$$MS_{it} = \frac{A_{it}/A_{1t}}{1 + \sum_{j>1} A_{jt}/A_{1t}} = \frac{\exp(Y_{it})}{1 + \sum_{j>1} \exp(Y_{jt})} \tag{13}$$

For the reference brand this expression simplifies to:

$$MS_{1t} = \frac{1}{1 + \sum_{j>1} \exp(Y_{jt})} \tag{14}$$

The estimated log ratio for the conventional multinomial logit attraction is:

$$\hat{Y}_{it} = \log(A_{it}) - \log(A_{1t}) = \sum_{p} a_{pi} x_{pit} - \sum_{p} a_{p1} x_{p1t} \tag{15}$$

Because expression 15 is linear, I estimate parameters of the multinomial logit attraction model by ordinary least squares.

The estimated log ratio for the neural net model can be written as:

$$\begin{aligned} \hat{Y}_{it} &= \log(A_{it}) - \log(A_{1t}) \\ &= \sum_{p} a_{pi} x_{pit} - \sum_{p} a_{p1} x_{p1t} + \sum_{k=1}^{K_i} b_{ki} h_{kit} - \sum_{l=1}^{K_1} b_{l1} h_{l1t} \end{aligned} \tag{16}$$

[19] e.g. *McGuire et al.* (1977), *Houston et al.* (1994).

Appendix B: Gradients and Stochastic Gradient Descent

The Gradients of the parameters needed both for stochastic gradient descent and BFGS are:

$$\frac{\partial \hat{Y}_{it}}{\partial a_{pj}} = \begin{cases} -x_{p1t} & : j = 1 \\ x_{pit} & : j = i \\ 0 & : \text{else} \end{cases} \tag{17}$$

$$\frac{\partial \hat{Y}_{it}}{\partial b_{kj}} = \begin{cases} -h_{k1t} & : j = 1 \\ h_{kit} & : j = i \\ 0 & : \text{else} \end{cases} \tag{18}$$

$$\frac{\partial \hat{Y}_{it}}{\partial c_{pkj}} = \begin{cases} -b_{k1} h_{k1t}(1 - h_{k1t}) x_{p1t} & : j = 1 \\ b_{ki} h_{kit}(1 - h_{kit}) x_{pit} & : j = i \\ 0 & : \text{else} \end{cases} \tag{19}$$

Stochasting gradient descent changes each parameter w by an amount proportional to the gradient of E for a randomly chosen observation[20]:

$$\Delta w \quad = \quad -\eta \frac{\partial E}{\partial w} = -\eta (Y_{it} - \hat{Y}_{it}) \frac{\partial \hat{Y}_{it}}{\partial w} \tag{20}$$

Random selection from observations Y_{it} allows wider exploration of the parameter space. Stochastic gradient descent stops if no percentual improvement of E greater than 0.01 is found for each of the last $T \times (I - 1)$ selected observed log ratios Y_{it}. I set the learning constant η to 0.5 and for each of the various models (distinguished by the number of hidden units) perform 100 runs of stochastic gradient descent, using different normally distributed initial parameter values with zero means and standard deviations equal to 0.3.

My implementation of BFGS calculates descent directions following a proposal of *Saito/ Nakano* (1997). We consider two alternative starting values of model parameters:

- the best parameter set among the stochastic gradient runs mentioned above in terms of E;
- the parameter values estimated for the best artificial neural net model in terms of E with fewer hidden units than the model considered together with zero values for parameters connected with the additional hidden units.

This way, BFGS as a rule provides two different parameter vectors from which I finally choose the vector associated with the smaller E value.

[20] e.g. *Hertz et al.* (1991); *Ripley* (1996).

Appendix C: Approximate Posterior Probabilities of Models

The change of the Bayesian Information Criterion[21] $\Delta BIC(m)$ caused by the m-th neural net model replacing the conventional multinomial logit attraction model is:

$$\Delta BIC(m) \quad = \quad N\left(\ln(E_m) - \ln(E_0)\right) - \ln(N)\left(p_0 - p_m\right) \tag{21}$$

N denotes the number of observations E_0, E_m are the error sum of squares, p_0, p_m the number of parameters for the multinomial logit model and the m-th artificial neural net model, respectively.

The Bayes factor $B(m)$, i.e. the ratio of the posterior odds of the m-th artificial neural net model to the multinomial logit model, can be approximated by means of $\Delta BIC(m)$ as follows:

$$B(m) \approx \exp(-\frac{1}{2}\Delta BIC(m)) \tag{22}$$

Because each Bayes factor $B(m^{'})$ is proportional to the posterior probability $p(m^{'})$ of the respective model $m^{'}$, posterior probabilities of models can be computed by:

$$p(m^{'}) = \frac{B(m^{'})}{\sum_m B(m)} \qquad \text{for } m^{'} = 1, m \tag{23}$$

[21] e.g. *Schwarz* (1978).

Table 1: *Model Evaluation*

Number of Hidden Units per Brand				Error Sum of Squares E	Posterior Probability $p(m)$
A	B	C	D		
0	0	0	0	3.396	
0 or 1 hidden unit per brand at least one brand with 1 hidden unit					
1	0	0	0	3.167	< 0.00001
0	1	0	0	2.861	< 0.00001
0	0	1	0	3.396	< 0.00001
0	0	0	1	3.396	< 0.00001
1	1	0	0	2.430	0.99998
1	0	1	0	3.078	< 0.00001
1	0	0	1	3.034	< 0.00001
0	1	1	0	2.763	< 0.00001
0	1	0	1	2.861	< 0.00001
0	0	1	1	3.396	< 0.00001
1	1	1	0	2.430	0.00001
1	1	0	1	2.430	0.00001
1	0	1	1	3.176	< 0.00001
0	1	1	1	2.671	< 0.00001
1	1	1	1	2.430	< 0.00001
1 or 2 hidden units per brand at least one brand with 2 hidden units					
1	1	1	2	2.430	< 0.00001
1	1	2	1	2.324	< 0.00001
1	1	2	2	2.239	< 0.00001
1	2	1	1	2.391	< 0.00001
1	2	1	2	2.391	< 0.00001
1	2	2	1	2.239	< 0.00001
1	2	2	2	2.239	< 0.00001
2	1	1	1	2.266	< 0.00001
2	1	1	2	2.266	< 0.00001
2	1	2	1	2.266	< 0.00001
2	1	2	2	2.266	< 0.00001
2	2	1	1	2.266	< 0.00001
2	2	1	2	2.266	< 0.00001
2	2	2	1	2.200	< 0.00001
2	2	2	2	2.200	< 0.00001

Table 2: *Model Evaluation (Continued)*

Number of Hidden Units per Brand				Error Sum of Squares E	Posterior Probability $p(m)$
A	B	C	D		
2 or 3 hidden units per brand at least one brand with 3 hidden units					
2	2	2	3	2.200	< 0.00001
2	2	3	2	2.200	< 0.00001
2	2	3	3	2.200	< 0.00001
2	3	2	2	2.200	< 0.00001
2	3	2	3	2.194	< 0.00001
2	3	3	2	2.200	< 0.00001
2	3	3	3	2.194	< 0.00001
3	2	2	2	2.200	< 0.00001
3	2	2	3	2.200	< 0.00001
3	2	3	2	2.200	< 0.00001
3	2	3	3	2.200	< 0.00001
3	3	2	2	2.200	< 0.00001
3	3	2	3	2.194	< 0.00001
3	3	3	2	2.141	< 0.00001
3	3	3	3	2.141	< 0.00001
4	4	4	4	2.141	< 0.00001

Table 3: *Elasticities at Average Prices*

Brand	Average Price	Own Elasticity	Cross-Elasticity
A	42.18	-3.83	1.19
		-0.99	0.35
B	36.22	-3.48	1.26
		-1.09	0.32
C	35.14	-1.29	0.35
		-1.09	0.29
D	40.10	-3.07	1.22
		-2.79	1.19

first line multinomial logit results, second line average over ANNAM models

Table 4: *Optimal Solutions for Fictitious Play*

MNL Model			ANNAM Models		
Optimal		Profit	Optimal		Profit
Price	Profit	according to ANNAM models	Price	Profit	Increase (%)
37	2993.56	2580.16	42	2856.35	10.70
36	1407.89	1287.95	41	1489.03	15.61
40	1061.05	1351.93	41	1360.22	0.61
37	2559.51	3103.09	38	3116.41	0.43

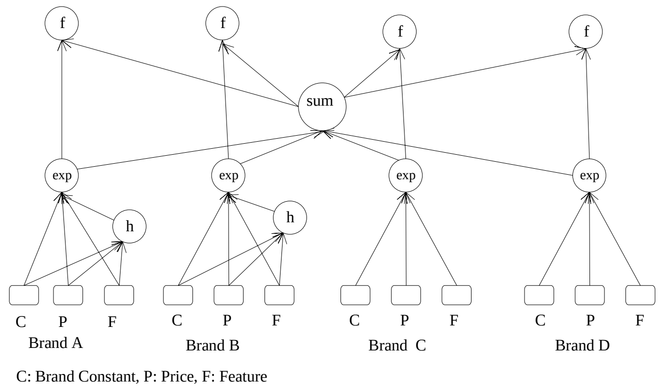

Figure 1: *Path Diagram of ANNAM with one hidden unit for Brand A and Brand B*

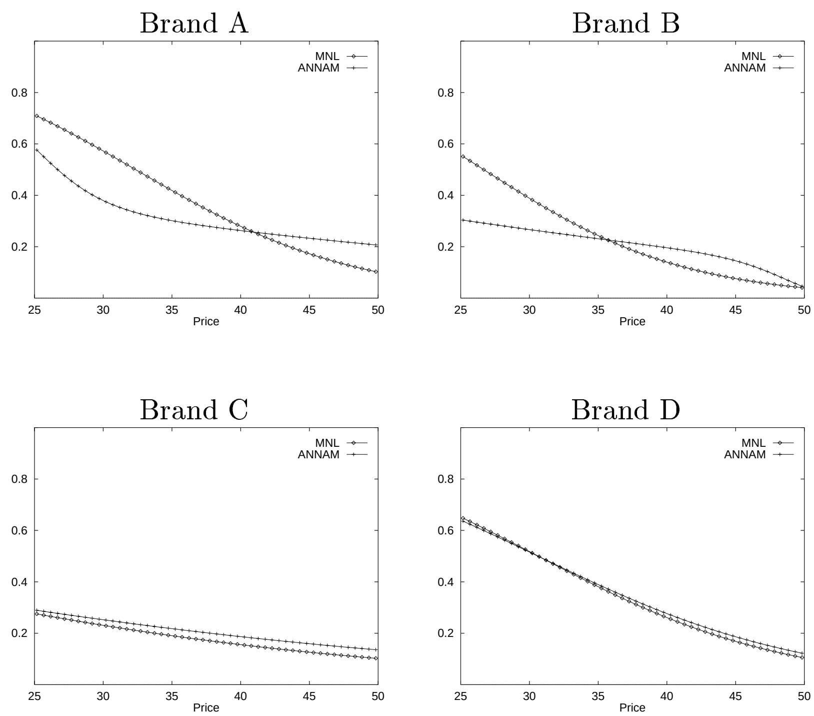

Figure 2: *Market Share vs. Price (MNL and ANNAM models)*