Working Paper Series

ANNAM
An Artificial Neural Net Attraction Model to Analyze Market Shares

Harald Hruschka

Working Paper No. 32
April 1999

Working Paper Series

April 1999

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 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

The marketing literature so far only considers attraction models with strict functional forms. Greater flexibility can be achieved by the neural net based approach introduced which assesses brands' attraction values by means of a perceptron with one hidden layer. Using log-ratio transformed market shares as dependent variables stochastic gradient descent followed by a quasi-Newton method estimates parameters. For store-level data the neural net model performs better and implies a price response qualitatively different from the well-known MNL attraction model. Price elasticities of these competing models also lead to specific managerial implications.

Keywords: Market Share Models, Attraction Models, Artificial Neural Networks

1 Introduction

Attraction models are derived from the Market Share Theorem of Bell, Keeney and Little (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 given a change in the attraction 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):

$$MS_{it} = \frac{A_{it}}{\sum_j A_{jt}} \quad (1)$$

Attraction models are logically consistent 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 (Naert and Bultez 1973; McGuire, Weiss and Houston 1977).

This paper considers so called 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 parametric attraction models, i.e. attraction models with strict functional forms (Nakanishi and Cooper 1974; Naert and Bultez 1973; Bultez and Naert 1975; McGuire, Weiss and Houston 1977; Leeflang and Reuyl 1984; Cooper and Nakanishi 1988; Abeele, Gijsbrechts and Vanhuele 1990; Cooper 1993; Chen, Kanetkar and Weiss 1994; Houston, Kanetkar and Weiss 1994). We introduce a more flexible neural net based approach which approximates attraction values without imposing rigid parametric dependence, but still preserves logically consistency.

One can find some examples for estimating aggregate market share response functions by means of artifical neural nets in the Marketing literature (van Wezel and Baets 1995; Wierenga and Kluytmans 1996). But these papers ignore the logical consistency issue metioned above.

The next section discusses the artificial neural net attraction model (ANNAM). Then estimation and model selection methods are dealt with. In the following estimation results (i.e. model performance, price effects) of an empirical study using store-level data are presented. The final section emphasizes managerial implications based on price elasticities which are measured by various models.

2 Artificial Neural Net Attraction Model

ANNAM postulates two parts a brand's attraction can be decomposed into. The first part algebraically corresponds to the exponential attraction of a parametric MNL attraction model, the second part constitutes the artificial neural net extension:

$$(2) \qquad A_{it} = \exp(\sum_{p} a_{pi} x_{pit} + \sum_{k=1}^{K_i} b_{ki} h_{kit} + \epsilon_{it})$$

x_{pit} denotes the p-th predictor for brand i in period t with $p = 1, P$. Errors ϵ_{it} are normally distributed with zero mean and constant variance.

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:

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

An artificial neural net with such a structure is guaranteed to approximate any continuous multivariate function with desired precision given a sufficient number of hidden units (Cybenko 1989; Hornik, Stinchcombe and White 1989; Ripley 1993).

The well-known multinomial (MNL) attraction model represents a special case of ANNAM 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 artifical neural net extension compared to a conventional MNL attraction model.

3 Estimation and Model Selection

McGuire, Weiss and Houston (1977) introduce the so-called log ratio transformation to simplify estimation of attraction models (see also Houston, Kanetkar and Weiss 1994). This approach is equivalent to the well-known log-centering transformation developed by Nakanishi (1972) as well as Cooper and Nakanishi (1974).

Taking the log of equation 1 gives:

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

Without loss of generality we take brand 1 as reference and subtract $\log(MS_{1t})$ from equation 4. This leads to:

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

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:

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

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}:

$$\begin{aligned} MS_{it} &= \frac{A_{it}/A_{1t}}{1+\sum_{j>1} A_{jt}/A_{1t}} \\ &= \frac{\exp(Y_{it})}{1+\sum_{j>1}\exp(Y_{jt})} \end{aligned} \tag{7}$$

For the reference brand this expression simplifies to:

$$MS_{1t} = \frac{1}{1+\sum_{j>1}\exp(Y_{jt})} \tag{8}$$

The error sum of squares E of log ratios serves as estimation objective to be minimized:

$$E = \frac{1}{2}\sum_t \sum_{i>1} (Y_{it} - \hat{Y}_{it})^2 \tag{9}$$

$\hat{Y}_{it}$ for $i = 2, \cdots, I$ symbolizes the estimated log ratio of brand i in period t which for ANNAM can be written as:

$$\begin{aligned} \hat{Y}_{it} &= \log(A_{it}) - \log(A_{1t}) \\ &= \sum_p a_{pi}x_{pit} - \sum_p a_{p1}x_{p1t} + \sum_{k=1}^{K_i} b_{ki}h_{kit} - \sum_{l=1}^{K_1} b_{l1}h_{l1t} \end{aligned} \tag{10}$$

Model estimation consists of two stages, stochastic gradient descent and BFGS. Stochasting gradient descent changes each parameter w by an amount proportional to the gradient of E for a randomly chosen observation (Hertz, Krogh and Palmer 1991; Ripley 1996):

$$\begin{aligned} \Delta w &= -\eta \frac{\partial E}{\partial w} \\ &= -\eta (Y_{it} - \hat{Y}_{it}) \frac{\partial \hat{Y}_{it}}{\partial w} \end{aligned} \tag{11}$$

Random selection from among 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 observations Y_{it}.

We 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 with different normally distributed initial parameter values having zero mean and standard deviation equal to 0.3. The best parameter set obtained by these 100 stochastic gradient runs in terms of E are used as starting values for the nonlinear optimization procedure BFGS of Broyden, Fletcher, Goldfarb und Shanno (Seber and Wild 1989; Bishop 1995). The

BFGS implementation calculates descent directions following a proposal of Saito and Nakano (1997).

Both estimation stages need gradients of the parameters which are given below:

$$\frac{\partial \hat{Y}_{it}}{\partial a_{pj}} = \begin{cases} -x_{p1t} & : j = 1 \\ x_{pit} & : j = i \\ 0 & : \text{else} \end{cases} \tag{12}$$

$$\frac{\partial \hat{Y}_{it}}{\partial b_{kj}} = \begin{cases} -h_{k1t} & : j = 1 \\ h_{kit} & : j = i \\ 0 & : \text{else} \end{cases} \tag{13}$$

$$\frac{\partial \hat{Y}_{it}}{\partial c_{pkj}} = \begin{cases} -b_{k1} h_{k1t}(1 - h_{k1t}) x_{p1t} & : j = 1 \\ b_{ki} h_{kit}(1 - h_{kit}) x_{pit} & : j = i \\ 0 & : \text{else} \end{cases} \tag{14}$$

The ANNAM models studied differ by the number of brand-specific hidden units assuming integer values between zero and four. Selecting a model on the basis of an error measure like E using all the data for estimation is prone to overfitting. That is why we cross-validate as suggested by Bishop (1995) by dividing the total data set into 10 subsets. Each of these subsets is ignored once during estimation and the model determined this way is applied to the left-out subset giving an error measure. Summing error measures over all subsets finally leads to the cross-validated error measure.

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).

We start with models having the same number of hidden units for each brand, i.e. setting each K_i to 0, 1, 2 and 3, respectively. Adding only one hidden unit per brand as artificial neural net part to the conventional MNL attraction model leads to a sharp decrease of cross-validated E from 3.684 to 2.888. ANNAM with two hidden units per brand attains the minimum value of cross-validated E (Table 1).

This result suggests to consider all possible combinations of the number of brand-specific hidden units between 2 and 3. In terms of cross-validated E the best among these models consists of three hidden units for bands A, B, C and two hidden units for brand D (Table 1). In the following we discuss results for this ANNAM model. The conventional MNL model without artificial neural net component serves as standard of comparison.

Figure 1 plots market share for brand A versus its own price (given average prices of competing brands) both for the MNL and the ANNAM model. ANNAM indicates a stronger marginal response for lower prices and a weak marginal response for higher prices. The MNL model on the other hand shows a rather constant marginal response. Market share plots for the other brands have similar shapes.

5 Managerial Implications

We determine (cross-) elasticities for the models studied. On their basis we derive implications for price decisions. From equation 7 we obtain as first derivative w.r.t. to predictor x_{pjt}:

$$\frac{\partial MS_{it}}{\partial x_{pjt}} = MS_{it}(\delta_{ij} - MS_{jt})\frac{\partial Y_{jt}}{\partial x_{pjt}} \tag{15}$$

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

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

Using expression 15 market share (cross-) elasticities el_{it} of brand i in period t w.r.t. predictor x_{pjt} may be written as:

$$el_{ijt} = (\delta_{ij} - MS_{jt})x_{pjt}\frac{\partial Y_{jt}}{\partial x_{pjt}} \tag{16}$$

Just like for the parametric MNL model cross-elasticities for all brands $i \neq j$ are equal.

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

$$el_{ijt} = (\delta_{ij} - MS_j)x_{pjt}(a_{pj} + \sum_{k=1}^{K_j} b_{kj}c_{pkj}h_{kjt}(1 - h_{kjt})) \tag{17}$$

It subsumes the well-known equation for the parametric MNL model as special case:

$$el_{ijt} = (\delta_{ij} - MS_j)x_{pjt}a_{pj} \tag{18}$$

Table 2 contains price elasticities and price cross-elasticities for the average price of each brand (with prices of the other brands also set to their average values). Contrary to MNL ANNAM indicates lower absolute values both for elasticities and cross-elasticities. For our data set MNL misleads a brand manager to overestimate the effect of her/his own price changes as well as of competitors' price changes on market share.

References

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

Bell, David E., Ralph L. Keeney, and John D.C. Little (1975). "A Market Share Theorem", *Journal of Marketing Research* 12, 136–41.

Bishop, Christopher M. (1995). *Neural Networks for Pattern Recognition.* Oxford: Clarendon Press.

Bultez, Alain V. and Philippe A. Naert (1975). "Consistent Sum-Constrained Models", *Journal of the American Statistical Association* 70 , 529–35.

Chen, Youhua, Vinay Kanetkar, and Doyle L. Weiss (1994). "Forecasting Market Shares with Disaggregate or Pooled Data: A Comparison of Attraction Models", *International Journal of Forecasting*, 10, 263–76.

Cooper, Lee G. (1993). "Market-Share Models". In Jehoshua Eliashberg and Gary L. Lilien (eds.), *Marketing.* Amsterdam: North-Holland.

Cooper, Lee G. and Masao Nakanishi (1988). *Market-Share Analysis.* Boston: Kluwer Academic Publishers.

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

Hertz, John, Anders Krogh and Richard G. Palmer. (1991). *Introduction to the Theory of Neural Computation.* Redwood City, CA: Addison-Wesley.

Hornik, Kurt, Max Stinchcombe, and Halbert White. (1989). "Multilayer Feedforward Networks are Universal Approximators", *Neural Networks* 3, 359–366.

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

Leeflang, Peter S.H. and Jan C. Reuyl. (1984). "On the Predictive Power of Market Share Attraction Models", *Journal of Marketing Research* 21, 211–5.

McGuire, Timothy W., Doyle L. Weiss, and Franklin S. Houston. (1977). "Consistent Multiplicative Market Share Models". In B.A. Greenwood and Danny N. Bellinger (eds.), *Contemporary Marketing Thought*, Chicago: AMA.

Naert, Philippe A. and Alain V. Bultez. (1973). "Logically Consistent Market Share Models", *Journal of Marketing Research* 10, 334–40.

Nakanishi, Masao. (1972). "Measurement of Sales Promotion Effect at the Retail Level". AMA Proceedings, Spring and Full Conference. Chicago: AMA.

Nakanishi, Masao and Lee G. Cooper. (1974). "Parameter Estimation for a Multiplicative Competitive Interaction Model – Least Squares Approach", *Journal of Marketing Research* 11, 303–11.

Ripley, Brian D. (1993). "Statistical Aspects of Neural Networks". In Ole E. Barndorff-Nielsen, Johan L. Jensen, and Wilfried S. Kendall (eds.), *Networks and Chaos - Statistical and Probabilistic Aspects.* London: Chapman & Hall.

Ripley, Brian D. (1996). *Pattern Recognition and Neural Networks.* Cambridge: Cambridge University Press.

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

Seber, George A.F. and C.J. Wild. (1989). *Nonlinear Regression.* New York: Wiley.

van Wezel, Michiel C. and Walter R.J.Baets (1995). "Predicting Market Responses with a Neural Network", *Marketing Intelligence & Planning* 13, 7, 23–30 .

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

Table 1: Model Performance

Hidden Units Brand				E	Cross-validated E
A	B	C	D		
0	0	0	0	3.396	3.684
1	1	1	1	2.545	2.888
2	2	2	2	2.475	2.716
3	3	3	3	2.375	2.731
2	2	2	3	2.287	2.418
2	2	3	2	2.244	2.399
2	2	3	3	2.347	2.822
2	3	2	2	2.207	2.431
2	3	2	3	2.193	2.358
2	3	3	2	2.272	2.434
2	3	3	3	2.226	2.418
3	2	2	2	2.287	2.472
3	2	2	3	2.466	2.587
3	2	3	2	2.474	2.649
3	2	3	3	2.246	2.460
3	3	2	2	2.275	2.420
3	3	2	3	2.202	2.396
3	3	3	2	2.140	2.300

Table 2: Elasticities

Brand	Average Price	Own Elasticity	Cross Elasticity
A	42.18	-3.83	1.19
		-0.83	0.28
B	36.22	-3.48	1.26
		-2.60	0.82
C	35.14	-1.29	0.35
		-0.70	0.21
D	40.10	-3.07	1.22
		-2.37	0.85

first line MNL, second line ANNAM results

Figure 1: Market Share (Brand A)