Mining Sales Data using a Neural Network Model of Market Response

Thomas S. Gruca	Bruce R. Klemz	E. Ann Furr Petersen
The University of Iowa	University of Nebraska at Kearney	The University of Iowa
Department of Marketing	Department of Marketing/Management	Department of Marketing
Iowa City, IA 52242	Kearney, NE 68849	Iowa City, IA 52242
(319) 335-0946	(308) 865-8170	(319) 335-1015
Thomas-gruca@uiowa.edu	Klemzb@unk.edu	Apetersn@blue.weeg.uiowa.edu

ABSTRACT

Modeling aggregate market response is a core issue in marketing research. In this research, we extend previous forecasting comparative research by comparing the forecasting accuracy of feed-forward neural network models to the premier market modeling technique, Multiplicative Competitive Interaction (MCI) models. Forecasts are compared in two separate studies: (1) the Information Resources Inc. (IRI) coffee dataset from Marion, IN and (2) the A.C. Nielsen catsup dataset from Sioux Falls, SD. Our results suggest neural networks are a useful substitute for MCI models when there are too few observations available to estimate a fully-extended MCI model. Implications are discussed.

Keywords

neural networks, market response model, sales/market share forecasting

1. INTRODUCTION

Modeling aggregate market response is a core issue in marketing research (Cooper and Nakanishi 1988), and the most widely used market share modeling approach is the multiplicative competitive interaction (MCI) model (Cooper and Nakanishi 1988). To model market response, a variety of parametric and non-parametric modeling techniques are available to the researcher, and there is a history of forecasting comparisons between linear, multiplicative and MCI market models. Kumar (1994) summarizes and extends this important stream of research, and recent research suggests the MCI model produces market share forecasts superior to those from simpler linear or multiplicative models (Kumar and Heath 1990).

The MCI model also yields estimates of market share lying between zero and one, while summing over all brands to unity (Naert and Bultez 1973; Nakanishi and Cooper 1974; McGuire and Weiss 1976). This modeling technique offers many alternative model formulations, such as the fully-extended MCI formulation (Carpenter, et al 1988), which can estimate the cross-effects of one brand's marketing mix on the attraction of another brand.

However, estimating such a model can require a vast amount of data due to the large number of parameters that are estimated. For N brands and M marketing mix variables, there are (M*N) variables in each market share equation. If serial correlation is not a concern, then each equation can be estimated separately. However, due to the effects of multi-period promotions and advertising, serial correlation is often pronounced in market response data. In such a case, a system of N-1 equations with (2*M*N) variables must be estimated simultaneously. This requirement often overwhelms the number of observations in the data available.

Additional scanner data issues faced by those modeling market response include non-linearities, asymmetric cross-effects and interactions among the input variables over short time periods (Sharda 1994; Hanssens, Parsons and Schultz 1990: 37-45). Therefore, the MCI model formulation often requires dramatic simplification to ensure that the model parameters can be estimated.

A widely utilized non-parametric method, feed-forward neural networks (Rumelhart and McClelland 1986), have been shown as alternatives to traditional statistical analysis methodologies such as time series techniques (Sharda and Patil 1990), regression based techniques (Refenes, Zapranis and Francis 1994; Gorr, Nagin and Szczypula 1994; Chiang, Urban and Baldridge 1996), ANOVA analysis (Bejou, Wray and Ingram 1996), and traditional clustering techniques (Chen, Mangiameli and West 1996).

Neural networks have been shown to out-perform traditional forecasting techniques in certain conditions (e.g. Hornik et al 1989), such as when non-linearities are present (e.g. Rumelhart and McClelland 1986; White and Stinchcombe 1992), discontinuities are present (e.g. Hill et al 1996), and when there are significant interactions among inputs (e.g. Rumelhart and McClelland 1986). These are the same type of data complexities commonly faced by those modeling market response.

In this research, we extend previous forecasting comparative research by comparing the forecasting accuracy of feed-forward neural network models to several MCI model formulations. Comparisons are performed in two separate studies: the Information Resources Inc. (IRI) coffee dataset from Marion, IN and the A.C. Nielsen catsup dataset from Sioux Falls, SD.

This paper is organized as follows. A detailed description of the market models used in this study is followed by a description of the comparison criteria. Following this is a description of the studies. The paper closes with a discussion of the results and concluding comments.

2. Market Models and Estimation Methods

In this section we briefly describe the structural market models used in this comparative study.

2.1. Neural Network[1] Market Model

The neural network market models utilized in this study are fully-connected 3-layer feed-forward neural networks estimated using Backpropagation (Rumelhart and McClelland 1986).

The subjectivity of specifying network architecture and training parameters are at the core of many criticisms of neural networks (Tam and Kiang 1992). To estimate these architecture parameter and estimation values, we first partitioned the total data sample into estimation and hold-out samples. Following Gorr, Nagin and Szczypula (1994), we (1) trained a network with the parameter values of a grid point and calculated the resulting SSE for the estimation set, (2) performed a complete enumeration of SSE for all grid points and selected the minimum SSE grid point, and (3) trained the resulting network to this minimum SSE and forecast the hold-out sample using a one-step-ahead rolling design.

For the coffee dataset, seven feed-forward fully-connected neural networks were estimated, one for each brand[2]. The input for each network include price, feature index, display index for each of the 7 brands plus lagged market share, resulting in 22 (3*7+1) inputs, the output was market share for that brand. The best-fit model configuration resulting from the grid search (detailed previously) contained 7 intermediate nodes, learning constant of 0.2, and momentum constant of 0.8. Market share forecasts were made using a one-step-ahead rolling design.

For the catsup dataset, four feed-forward fully-connected neural networks were estimated, one for each brand. The input for each network include price for each of the 4 brands plus lagged market share, resulting in 5 (1*4+1) inputs, the output was market share for that brand. The best-fit model configuration resulting from the grid search (detailed previously) contained 4 intermediate nodes, learning constant of 0.2, and momentum constant of 0.8. Market share forecasts were made using a one-step-ahead rolling design.

2.2. MCI Market Model

MCI model formulation and estimation has been extensively detailed in the forecasting and modeling (e.g. Cooper and Nakanishi 1988) literatures. The general form of a market share attraction model has the following structure:

(1) $m_i = a_i / \Sigma_j a_j$

where: m_i are the attraction shares or, in practice, market share, for brand i, and:

(2) $a_i = \exp(\alpha_i + \varepsilon_i) \Pi_h \Pi_j (x_{hj})^{\beta_{hij}}$

where: α_j denotes brand specific constraints, X_{hj} denotes mix variable h for brand j, ε_i is a white noise error term, and a_i is the attraction value of brand i.

Bell, Keeney and Little (1975) showed that the attraction model given by equation (1) yields logically consistent estimates of market share for each brand, i.e. $0 \leq m_i \leq 1$ and $\Sigma_n m_n = 1$, under the following assumptions: $a_i \geq 0$, $\Sigma_n a_n > 0$, if $a_i = 0$ then $m_i = 0$, if $a_1 = a_2$ then $m_1 = m_2$, and if the attraction of a competitor of a product i increases by some amount τ, then the new market share of product i will not depend on which competitor made the increase.

This model allows for many different MCI model formulations. The constant-effects MCI model assumes that all brands have the same coefficients ($\beta_{hii} - \beta_{hjj} = \beta_h$ and $\beta_{hij} = 0$ for $i \neq j$). In the differential effects model, brands have different coefficients for each marketing mix variable ($\beta_{hij} = 0$ for $i \neq j$). The fully-extended asymmetric cross-effects model has no restrictions on the β parameters. The complex denominator of the MCI model requires a transformation in order to derive estimation equations. In these applications, we use the log-ratio transformation (Theil 1969), developed in a marketing context by McGuire, Weiss and Houston (1977).

Serial correlation of the error structures, caused by the influence of past marketing actions on current consumer behavior, requires modification of the basic model formulation. Following Durbin (1960), we formulated a regression model with spherical disturbances by including lagged dependent and independent variables and an autocorrelation coefficient. Estimation of the resulting equations is accomplished using Zellner's (1962) Seemingly Unrelated Regression (SUR) procedure.

For the coffee dataset, the market consisted of seven brands and three (price, feature and display) marketing mix variables. A fully-extended MCI market share model (adjusted for serial correlation) would require the estimation of 259 parameters. Estimating this many parameters would require at least 260 observations. Since we have only 49 weeks of data in the estimation sample, the fully-extended MCI model could not be estimated from the available data. Therefore, we estimated a reduced form of the MCI model, specifically the differential effects model. By restricting our analysis to the simplified differential-effects MCI model, it may appear that we have *stacked the deck* in favor of the neural network modeling approach, an issue which is addressed in the catsup studies. We estimated the log-ratio (McGuire, Weiss and Houston 1977) version of the differential effects MCI model adjusted for serial correlation (Durbin 1960) using Zellner's (1962) SUR procedure as implemented in SAS.

For the catsup dataset, we estimated the fully-extended MCI model with price as the only current period marketing mix variable. We estimated a four brand, one variable fully-extended (main and cross effects) MCI model adjusted for serial correlation. As in the coffee market study, we used the log-ratio transformation, and estimated the coefficients using SUR as implemented in SAS.

2.3. Comparison Criteria

Forecasting accuracy for the estimation and hold-out samples were evaluated using MAPE - the mean absolute percentage error (Makridakis 1993).

3. CASE STUDIES

Two studies were performed using two scanner panel data sets. Each study will be discussed separately.

3.1. Information Resources Inc. (IRI) coffee dataset

Using the IRI BehaviorScan scanner panel data provided from the Marion, IN market, we computed the weekly shares for the top 7 brands in the ground regular coffee market by aggregating store level scanner panels data across all grinds, sizes and stores. These brands represent more than 90% of all panelist purchases from March 1981 to April 1982, resulting in 58 week of data[3]. We split the 58 week sample into a 49 week estimation sample and a 9 week hold-out sample.

All models estimated in this research used the same dependent and independent variables. The dependent variable, market share, was aggregated across all grinds, sizes and stores. The independent variables include price, feature, display, and lagged market share. Price was measured in dollars per ounce, net of any discounts.

To include the effect of retail feature and display activities, we utilized the distinctiveness index (Nakanishi, Cooper and Kassarjian 1974) which takes on values of 1.0 if all or none of the brands are on feature (display) in a given week. For other circumstances, the index is:

$X_{f,i} = N/n_f$ if brand i is on feature (display)

$X_{f,i} = (1-n_f/N)$ for all brands not on feature (display)

where there are N brands in the market and n_f is the number of brands on feature (display) in a given week. The sample profile for the independent and dependent variables are presented in Table 1.

3.2. A.C. Nielsen catsup dataset

Using the A.C. Nielsen catsup dataset provided from the Sioux Falls, SD market, we computed the weekly shares for the top 4 brands by aggregating store level scanning records across sizes and stores. These brands represent more than 90% of all purchases from August 1985 to August 1988, resulting in 156 weeks of data. We split the 156 week sample into a 146 week estimation sample and a 10 week hold-out sample.

All models estimated in this research used the same dependent and independent variables. The dependent variable, market share, was aggregated across all sizes and stores. The independent variables were price and lagged market share. Price was measured in cents per pound. The sample profile for the independent and dependent variables are presented in Table 2.

Table 1: Marion, IN Ground Coffee Market

	Price (¢ per oz.)						
	Folg.	Folg. Flake	Chase & Sand.	Max. House	Store 1	Store 2	Mast. Blend
Avg	17.49	18.12	15.70	17.62	12.80	13.07	17.40
Std dev	0.78	0.89	0.84	0.63	1.24	0.92	1.01
Min	15.59	15.61	12.72	15.58	10.57	11.19	14.57
Max	18.89	19.63	16.94	18.90	15.10	14.94	20.69
	Feature Index						
	Folg.	Folg. Flake	Chase & Sand.	Max. House	Store 1	Store 2	Mast. Blend
Avg	1.74	1.53	0.88	1.35	2.02	0.95	1.14
Std dev	1.27	1.04	0.71	1.00	1.12	0.76	0.94
Min	0.43	0.43	0.14	0.43	0.43	0.29	0.29
Max	7.00	3.50	3.50	3.50	3.50	3.50	3.50
	Display Index						
	Folg.	Folg. Flake	Chase & Sand.	Max. House	Store 1	Store 2	Mast. Blend
Avg	1.17	0.97	2.57	1.27	2.50	0.76	0.76
Std dev	1.29	0.62	1.66	1.07	1.81	0.15	0.15
Min	0.57	0.57	0.57	0.43	0.57	0.43	0.43
Max	7.00	3.50	7.00	7.00	7.00	1.00	1.00
	Market Share						
	Folg.	Folg. Flake	Chase & Sand.	Max. House	Store Brand 1	Store Brand 2	Mast. Blend
Avg	0.21	0.15	0.04	0.18	0.23	0.09	0.10
Std dev	0.08	0.04	0.05	0.07	0.08	0.04	0.06
Min	0.06	0.07	0.00	0.08	0.08	0.01	0.01
Max	0.37	0.26	0.30	0.39	0.38	0.19	0.30

Table 2: Sioux Falls, SD Catsup Market

	Price (¢ per lb.)			
	Hunts	Del Monte	Heinz	Store Brand
Avg	0.53	0.56	0.64	0.47
Std dev	0.04	0.06	0.05	0.02
Min	0.43	0.39	0.39	0.35
Max	0.62	0.70	0.70	0.54
	Market Share			
	Hunts	Del Monte	Heinz	Store Brand
Avg	16.80	8.11	71.12	4.23
Std dev	0.08	0.06	0.10	0.02
Min	2.72	1.43	37.62	0.87
Max	55.90	33.59	90.10	21.2

4. RESULTS

4.1. IRI Coffee Dataset

We compared forecasting accuracy for the estimation and hold-out sample. These forecasting results are presented in Table 3. For the estimation sample, the neural network had a lower MAPE for six of seven brands. For the hold-out sample, the forecasts from the neural network had lower MAPE for all brands. Neither model predicted the hold-out sample for Chase & Sandborn well, due, in part, to its very low market share.

For the estimation sample, the neural network had a lower MAPE for six of seven brands. For the hold-out sample, the forecasts from the neural network had lower MAPE for all brands. Neither

model predicted the hold-out sample for Chase & Sandborn well, due, in part, to its very low market share.

Table 3: Coffee Market Forecasting MAPE

	Estimation sample NNET model	Holdout sample NNET model	Estimation sample MCI model	Holdout sample MCI model
Folgers	16.41	11.46	19.83	26.55
Folgers Flaked	16.53	13.21	19.31	25.45
Chase & Sand.	87.07	137.68	70.18	169.39
Maxwell House	19.04	16.58	25.38	59.78
Chain 1	14.74	31.23	17.25	78.32
Chain 2	31.94	30.22	28.90	130.06
Master Blend	17.31	23.67	25.51	186.91

The differences between the accuracy measures for the two models were small for the estimation sample. However, there were large differences in forecast accuracy in the hold-out sample. The neural network models fit this data better for at least four of the brands.

The forecasting accuracy of the neural networks was better than the MCI model (for 7 of 7 brands using MAPE). This is an indication that the neural networks were not over-fit to the estimation data. If they were, the neural networks would be unable to predict well in the hold-out sample (e.g. Refenes, et al 1994).

4.2. A.C. Nielsen catsup dataset

The forecasting accuracy for the fully-extended MCI and neural network models were compared for both the estimation and hold-out samples. The results are presented in Table 4.

Table 4: Catsup Market Forecasting MAPE

	Estimation sample NNET model	Holdout sample NNET model	Estimation sample MCI model	Holdout sample MCI model
Hunts	24.72	25.98	18.30	11.33
Del Monte	35.26	49.02	20.58	26.19
Heinz	7.14	3.81	5.12	4.19
Store brand	21.88	27.79	17.22	22.92

Comparing the fully-extended MCI model and the neural network model directly, the MCI model fit the estimation sample better for all 4 brands. In the hold-out sample, the neural networks forecast better than the MCI model only for the largest brand, Heinz.

5. SUMMARY AND CONCLUSIONS

The critical issue addressed in this research is how well the neural network models perform compared to existing leading-edge models of market response. The empirical comparisons presented here were not chosen at random. The coffee dataset was utilized because there were too few observations to estimate the fully-extended MCI model, and the catsup dataset was utilized because there were enough observations to exploit the power of the fully-extended model. Overall, the neural network models performed better than the MCI models when insufficient data forced the estimation of an under-specified MCI model. When there was sufficient data to estimate a fully-extended MCI model, the MCI model had marginally better performance than the neural network.

Many studies that compare neural networks to regression techniques tend to model human decision tasks as in Gorr, et al (1994). In their review, Hill et al (1994) claim there are no studies which compare neural networks and regression models using real multivariate forecasting data. While there is one such marketing application (Hruschka 1993), it compares a simple neural network to a linear sales model in a non-competitive (monopoly) situation using monthly data. Our study uses state-of-the-art data sources (household-level and store-level data) and market response models in competitive markets.

These two studies suggest a niche where neural networks might be a useful substitute for MCI models. Neural networks are trained using an iterative optimization procedure, where the MCI model was estimated using SUR that is based on asymptotic theory. If a marketing researcher suspects that significant cross-effects are present, and there are too few observations to estimate a fully-extended MCI model, our research shows that a neural network will outperform a simplified MCI model. Otherwise, we found that there is no advantage to neural networks in stable markets with few brands. This finding is consistent with previous research, which suggests that the advantages of neural networks are restricted to specific types of forecasting problems (Hill et al 1994).

Even in this age of abundant scanner data, scenarios such as the catsup example cited in this research are rare in practice. Namely, few product categories exist where there are no product introductions or withdrawals for enough periods to estimate a fully-extended MCI model. In these instances, neural networks offer the brand manager a rich and powerful tool that can be used to model complex relationships and forecast sales.

6. ACKNOWLEDGEMENTS

7. ENDNOTES

[1]Neural network formulation and estimation issues have been extensively detailed in the forecasting literature (Gorr 1994), and are therefore omitted from this paper for brevity.

[2]The neural networks were written in Visual Basic 4.0 using NeuroWindows 4.0 DLL. All networks were trained on an IBM-type Pentium II 233 PC running Windows '95.

[3]In IRI-coded week 78, there was an entry of a major new brand - Master Blend. Some previous research using this database seems to have ignored this event (e.g. Gupta 1988).

8. REFERENCES

[1] Bejou, D., Wray, B., and Ingram, T.N. Determinants of Relationship Quality: An Artificial Neural Network Analysis. Journal of Business Research, 36 (1996), 137-143.

[2] Bell, D.E., Keeney, R.L. and Little, J.D. A Market Share Theorem. J. of Marketing Research, 12 (1975), 136-141.

[3] Carpenter, G.S., Cooper, L.G., Hanssens, D.M., and Midgley, D.F. Modeling Asymmetric Competition. Marketing Science, 7 (1988), 393-412.

[4] Chen, S.K., Mangiameli, P., and West, D.The Comparative Ability of Self-Organizing Neural Networks to Define Cluster Structure. Omega, 23 (1996), 271-279.

[5] Chiang, W.C., Urban, T.L., and Baldridge, G.W. A Neural Network Approach to Mutual Fund Net Asset Value Forecasting, Omega, 24 (1996), 205-215.

[6] Cooper, L.G. Market Share Models. In Handbook of Operations Research & Management Science, Vol 15, Marketing, Eliashberg, J. and Lilien, G.L., eds., Elsevier (New York, 1993).

[7] Cooper, L.G. and Nakanishi, M. Market-Share Analysis. Kluwer Academic Press (Boston, MA, 1988).

[8] Durbin, J. Estimation of Parameters in Time-Series Models. Journal of the Royal Statistics Society, 22 (1960), 139-153.

[9] Gorr, W.L. A Research Prospective on Neural Network Forecasting. Inter. Journal of Forecasting, 10 (1994), 1-4.

[10] Gorr, W.L., Nagin, D., and Szczypula, J. Comparative Study of Neural Networks and Statistical Models for Predicting Student Grade Point Averages. International Journal of Forecasting, 10 (1994), 17-34.

[11] Gupta, S. Impact of Sales Promotion on When, What and How Much to Buy. J. Marketing Research, 25 (1988), 342-355.

[12] Hanssens, D.M., Parsons, L.J., and Schultz, R.L Marketing Response Models: Econometric and Time Series Analysis. Kluwer (Boston, MA, 1990).

[13] Hill, T., O'Connor, M., and Remus, W. Neural Network Models for Time Series Forecasting, Management Science, 43 (1997).

[14] Hornik, K., Stinchcombe, M., and White, H. Multilayer Feedforward Networks are Universal Approximators. Neural Networks, 4 (1989), 251-257.

[15] Hruschka, H. Determining Market Response Functions by Neural Network Modeling: A Comparison to Econometric Techniques. Euro. J. Operations Research, 66 (1993), 27-35.

[16] Kumar, V. Forecasting Performance of Market Share Models: An Assessment, Additional Insights and Guidelines. International Journal of Forecasting, 10 (1994), 295-312.

[17] Kumar, V. and Heath, T.B. A Comparative Study of Market Share Models Using Disaggregate Data. International Journal of Forecasting, 6 (1990), 163-174.

[18] Makridakis, S. Accuracy Measures: Theoretical and Practical Concerns. Inter. Journal of Forecasting, 9 (1993), 527-529.

[19] McGuire, T.W. and Weiss, D.L. Logically Consistent Market Share Models II. J. Marketing Research, 13 (1976), 296-302.

[20] McGuire, T.W., Weiss, D.L., and Houston, F.S. Consistent Multiplicative Market Share Models. In Contemporary Marketing Thought, Greenwood and Bellinger, eds. American Marketing Association, Chicago (1977), 129-134.

[21] Naert, P.A. and Bultez, A. Logically Consistent Market Share Models. Journal of Marketing Research, 10 (1973), 333-334.

[22] Nakanishi, M. and Cooper, L.G. Parameter Estimation for a Multiplicative Competitive Interaction Model - Least Squares Approach. J. of Marketing Research, 11 (1974), 303-311.

[23] Nakanishi, M., Cooper, L.G., and Kassarjian, H. Voting for a Political Candidate Under Conditions of Minimal Information. Journal of Consumer Research, 1 (1974), 36-43.

[24] Refenes, A.N., Zapranis, A., and Francis, G. Stock Performance Modeling Using Neural Networks: A Comparative Study with Regression Models. Neural Networks, 7 (1994), 375-388.

[25] Rumelhart, D.E., and McClelland, J. Parallel Distributed Processing: Explorations in the Microstructures of Cognition, MIT Press (Cambridge, MA, 1986).

[26] Sharda, R., Neural Networks for the MS/OR Analyst: An Application Bibliography. Interfaces, 24 (1994), 116-130.

[27] Sharda, R., and Patil, R.B. Neural Network as Forecasting Experts: and Empirical Test. International. Joint Conference on Neural Networks, 2 (1990), 491-494.

[28] Tam, K.Y. and Kiang, M.Y. Managerial Applications of Neural Networks: The Case of Bank Failure Predictions. Management Science, 38 (1992), 926-947.

[29] Theil, H. A Multinomial Extension of the Linear Logit Model. Int. Journal of Economics, 10 (1969), 251-259.

[30] White, H. and Stinchcombe, M. Approximating and Learning Mappings Using Multilayer Feedforward Networks with Bonded Weights. In Artificial Neural Networks: Approximations and Learning Theory, H. White, ed. Blackwell (Oxford, UK, 1992).

[31] Zellner, A. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. J. of the American Statistics Association, 57 (1962), 979-992.

About the Authors:

Thomas S. Gruca Ph.D. is an Associate Professor in Marketing at the University of Iowa. His areas of research interest include marketing strategy and market structure.

Bruce R. Klemz is an Assistant Professor of Marketing at the University of Nebraska at Kearney. His areas of research interest include advanced software tools to address such areas as competitor analysis, positioning and customer service.

E. Ann Furr Petersen is a doctoral student in Marketing at the University of Iowa. Her dissertation research focuses on consumer choice among the complex options present in cafeteria-style health plans.