A Brand Share Prediction Model Based on Several Disparate Sources of Data: An Empirical Model of Detergent Choice in Mumbai, India

by *Arindam Banerjee,* Associate Professor, Marketing Area, Indian Institute of Management Ahmedabad, Vastrapur, Ahmedabad, Gujarat, Pin 380015, India

Abstract

We describe the application of a nested logit function for modeling consumer brand choice using household transaction data from the Indian market. This is unique since it is one of the first attempts to integrate disparate consumer information sources available at various levels of aggregation towards developing a prediction model for brand market share in India. We test the usefulness of the model for forecasting brand market share in the premium detergents market in Mumbai, India. The results of the model building exercise reveal the importance of advertising, specifically the role of ad message in influencing brand choice. It is concluded that such modeling initiatives show significant returns for market planning exercises in developing markets. However, the need for streamlining the collection of market data and its subsequent organization in a form that can help develop more portent prediction models is apparent.

Keywords: Brand Development, Prediction Model, Choice Modeling, Nested Logit, Emerging Markets, Advertising

1. Introduction

Prediction models in the marketing context can be used to map input variables like sales promotions, price variations, distribution, and advertising to defined output performance variables like market share, returns etc. These models provide a logical basis to the manager to compute the differential impact of firm's marketing strategies in comparison with its competitors on the market share of its brands using scenario builders. These models could be based on volumetric share analysis, choice share analysis or a combination of the two. Of these, volumetric share based models are very popular in the developed economies because of low complexity and easy interpretability of the output of the models.

Development of such models in emerging markets has not been attempted in the past because of non-availability of large-scale databases, which are necessary to build robust models. In the recent past, many organizations have invested in fairly large sized panels to track customer purchases in India. However, most of the applications based on these databases have been limited to tracking customer purchases on an ongoing basis. There has been little attempt at developing behavior models that help identify drivers of consumer behavior and their relative impact.

This paper is perhaps the first significant attempt to utilize customer transaction data from a large emerging market to build a prediction model based on choice share analysis. The model has high managerial relevance because it directly links marketing initiatives to performance outcomes such as competitive market-shares. One of the key contributions of this paper is to provide a perspective to managers in relatively nascent markets on how disparate sources of data collected with the objective of doing piecewise analysis can be combined to build useful prediction models. The paper describes the challenges of estimating models by integrating market data collected at different levels of aggregation and of varying quality. More specifically, for the detergents market in Mumbai, India, it is able to identify key marketing elements that have significant influence on consumer behaviour. The conclusions reached indicate that brand prediction models have significant utility for planning marketing initiatives even in the evolving markets and there is an advantage to design and manage the collection of appropriate market data to develop robust models.

The rest of the paper is divided into four sections. The second section discusses the rising importance and problems of emerging markets, the third section discusses about choice of modeling technique in such markets and the fourth section discusses model development and results. Conclusions are provided in the fifth section.

2. Modelling Issues in Emerging Markets

Developing markets like India are characterized by higher growth rates, increasing incomes and are supposedly the future drivers of the world economy in the 21st century. India is different from developed countries in terms of its market characteristics and data gathering practices. It is more dynamic and hence evolving at a significant rate. For example, the number of SKU's (stock keeping units) in India in the consumer retail sector has grown from 9093 to 17739 in the period from 1990 to 1996 [1]. There is low penetration of organized retailing outlets and a significantly large concentration of small convenience stores with low or negligible automation.

The heterogeneity in modes of selling and low rate of automation create various problems in data capture and collation [2]. The overall market is organized in a haphazard manner with different agencies capturing separate data using different attributes; hence compatibility of databases is a significant problem. Much of this information is gathered for tracking overall market trends and hence is collected at the aggregated level.

A major source of consumer level data capture in these markets is via panels maintained by individual companies. The household level panel information in these markets is gathered primarily using the diary mode. However the extent and reliability of information collected from the panelists is still scarce compared to their counterparts in more developed markets. While it is essential to improve data collection practices, it would take a considerable amount of time to evolve a system that is comparable to the west. Therefore, it is sensible to focus on developing foundation level models based on available resources and portray the utility of the same prior to emphasizing the need to evolve better data collection and management systems.

3. Prediction Models: Choice Of An Underlying Technique

Prediction models using consumer behavior data normally link marketing mix variables to output performance measured as brand choice or quantity purchased. These measures at a higher level of aggregation such as average probability of choice across a representative sample of customers can be considered as good surrogates for the brand's market share. Typical prediction models using behaviour data as input can be estimated either using store level data gathered from a sample of stores or transaction level data collected from a sample of consumers. In this specific model building initiative, the nature of the product category (large sized detergent packs) used for the estimation made the use of a brand choice model imperative [3], since a large percentage of panelists resorted to the purchase of single packs on any buying occasion.

A typical consumer choice model uses utility maximization theory as its foundation. Given a set of alternatives, a consumer chooses the brand that provides the maximum utility on a particular occasion. Consumer choice has been extensively modeled in the marketing literature as brand choice or simultaneous decision of both choice and quantity [4-6]. Logit and Probit are most commonly used specifications in the literature to model the phenomenon [7-10]. Consumer choice models are developed using data pertaining to social, demographic, psychometric measures of customers and marketing mix variables of all competing products in a product category gathered over large number of choice occasions (refer Figure 1). The databases available for modeling brand choice were available from various sources, not specifically designed to be compatible to each other. A description of the same follows in the next section

4. Model Development

4.1 Datasets Used In The Study

Considerable effort is required to process data accessed from various sources to make them compatible for use in developing a prediction model. This was

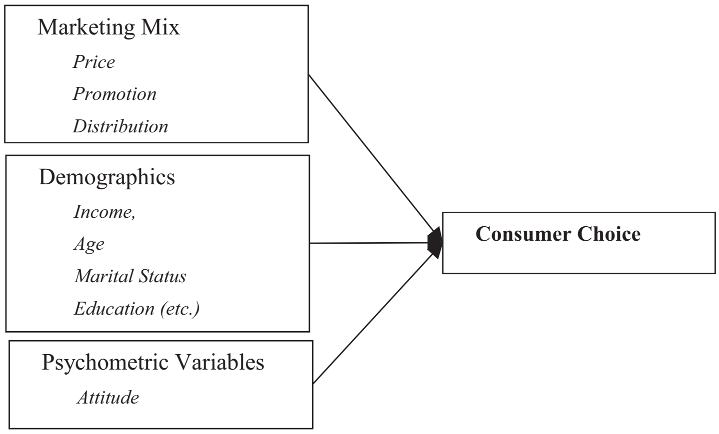

Figure 1 – Data requirements for Choice Models

true in the given context since most of the customer information sources were developed for various independent purposes and not with a concerted plan to use them together for modeling purposes. Hence, incompatibility due to different levels of aggregation across databases, different sample respondents across various panels used for model building, provide significant hurdles in estimating models with reasonable precision.

The data used for the study was made available to the researchers by a leading firm operating in the consumer packaged good market in India. The data provided was from Oct 99 to Dec 2000. It comprised three distinct datasets.

Purchase transaction (behavior) data

The firm maintained a household diary panel with the help of a professional market research agency to collect transaction level information from the randomly selected households in Mumbai, India. The consumer recorded the data manually, in a dairy soon after his/her purchase and the diary was collected by the agency on a monthly basis. Data collected from each panelist had information on the brands bought, size (SKU) bought, amount consumed, frequency of purchase and demographic information pertaining to the panelist. It did not record point of purchase marketing mix elements. The original purpose of maintaining the database was to enable the understanding of loyalty and switching behavior across various brands for individual customers. Main features of data are given in table 1.

Table 1 - Characteristic of Behavior Panel Data	
Total Number of Households	930
Total Number of Transactions	4165
Transactions of Brand Alpha	1987
Transactions of Brand Beta	1689
Transactions of Brand Gamma	429
Transactions of Brand Theta	60
Maximum no. of brands purchased by any customer in one month	12
Average no. of brands Purchased per month	1.013
Average no. of brands Purchased in two years (surrogate for loyalty)	1.4
Average number of transactions made in two years	4.469
Average number of packs bought per transaction	1.235

Attitudinal data

The firm maintained a separate weekly revolving panel of 50 consumers to track brand related image/attitude association with the help of a professional market research agency. The data was collected for all competing brands. The panel members were different from the members in the fixed transaction panel**.** The objective of collecting this data was to independently assess the strength of each brand based on the attitudes associated with them. However, there is no evidence of previous attempts at linking consumer attitudes to their purchase behaviour.

Marketing Mix Variables

Information about aggregate level marketing mix variables was available from retail audit data of ORG-MARG (a leading Market Research Company in India) which captures the information about pricing and distribution of all the brands in a category from a set of representative stores in each market on a monthly basis. These data were available at the aggregate level for the market and not at the transaction level. Additionally, data were available on the number of 'promoted brands' sold in each month but there was no information available on the type of promotional strategy used. Historically, this set of information has been used to track competitor tactics. However, no attempts have been made to evaluate the efficacy of own versus competitor's tactics on brand performance, by linking this data to actual customer transactions.

Modelling Detergent Choice

4.2 Choice Model

We derived the brand choice model using a logistic function. The general form of model can be described as:

Brand Choice = *f* (Marketing Mix, Psychometric Variables, Demographic variables) (1)

In the literature, marketing mix variables like regular price, promoted price, display, type of promotion have been used to model choice using a conditional choice model where each choice alternative has a unique value associated with each variable. For instance, there will be a unique price associated with each brand alternative, which are all included in the model. The model will provide as an output, the "elasticity" (impact coefficient) of price. However, our model also has attitude variables since explicit treatment of psychological factors in choice models leads to a more behaviorally realistic representation of the choice process, and consequently, better explanatory power [11-13]. We use demographic variables in choice models as it provides managers with identifiable segments and their choice priorities [14,15]. Both attitude and demographic variables are different from marketing mix variables since these are associated with respondents rather than choice alternatives. Hence, there is only one value for each variable associated with each respondent (for example, age would be defined uniquely for each respondent rather than have different values associated with each choice alternative). In this case, the logit model outputs a set of parameters (impact coefficient) associated with each variable. The number of such parameters in each set depends upon the number of choice alternatives. Variables with multiple impact coefficients, one for each choice alternatives are usually defined as polytomous variables. A mixed logit model usually inputs both polytomous and conditional variables. The model presented in this paper conforms to the structure of a mixed logit model. We have incorporated as many relevant market and consumer specific variables accessible to the managers in this model to allow for the development of a robust yet useful specification.

We use the framework depicted in equation 1 to develop our choice model including marketing mix (refer table 2), demographics (refer table 3) and attitude variables. We have modelled consumer choice as choice of brand and pack-size only with number of packs purchased being one at any choice occasion. The quantity decision was not considered because the average number of packs purchased per choice occasion was just over one (1.2).

Let us consider that an individual 'j' is faced with a situation in which s/he has to choose amongst 'm' brands available in market. Let 'P_{jk}' is the probability that individual 'j' will choose brand 'k' from the choice set 'm'; 'X_j' is the demographic characteristics of the individual and 'Z_{jk}' is the characteristic of 'k^{th}' brand as observed by individual 'j' (brand related factors like price, distribution etc.) α and β are parameter estimates for conditional and polytomous variables.

Then

$$P_{jk} \frac{\exp[(\alpha Z_{jk} + \beta_k X_j)}{\sum_{l=1}^{m} \exp(\alpha Z_{jl} + \beta_k X_l)} \tag{2}$$

The model was estimated for the premium laundry detergent category using the data, which was available from October 1999 to December 2000 from the Mumbai market in India. There were 930 households in the dataset with 4165 transactions. The overall choice set in December 2000 constituted four detergent brands – Alpha, Beta, Gamma and Theta[1] (see Box 1 for details on brands). Theta was introduced in the market in June 2000. Alpha and Beta constituted around 85% of market and thus were major brands. Both the brands have been available to consumers for a significant amount of time. Hence, the heritage effect can be assumed to be minimal. Each of the two major brands had three different pack sizes selling in the market (200 gms., 500 gms. and 1000 gms.).

To simultaneously incorporate the brand and pack-size competition, a two stage model (nested logit) was estimated (see Figure 2). In stage 1, the utility because of pack sizes was computed and in stage 2, pack coefficient estimated in stage 1 were used as an additional variable and probability of choice of each brand was estimated [16,17]. This was concluded based on past qualitative research on consumer buying patterns in the market. Besides, the panel data exhibited asymmetric switching behavior across brands and pack sizes, implying that the pack sizes within each brand were more direct substitutes compared to brands.

4.3 Discussion on Variables: Developing Proxies

4.3.1 Marketing Mix Variables

For the current study, data on marketing mix variables (price, distribution) were available at the market level from the retail audit data maintained by ORG-MARG. We operationalized price as Maximum Retail Price per gram (MRP) in each month. Price discounting was not observed to be a frequently employed marketing tactic and hence MRP appeared to be the only operational price variable. All consumers were subjected to the same prices as indicated in the retail survey.

We used aggregate level dealer stocking position (% of dealers stocking the brand) as a proxy for availability of the brand at the transaction level. This is usually not the best variable to capture availability of brands at the transaction level (Point of Sale information), however this was the best

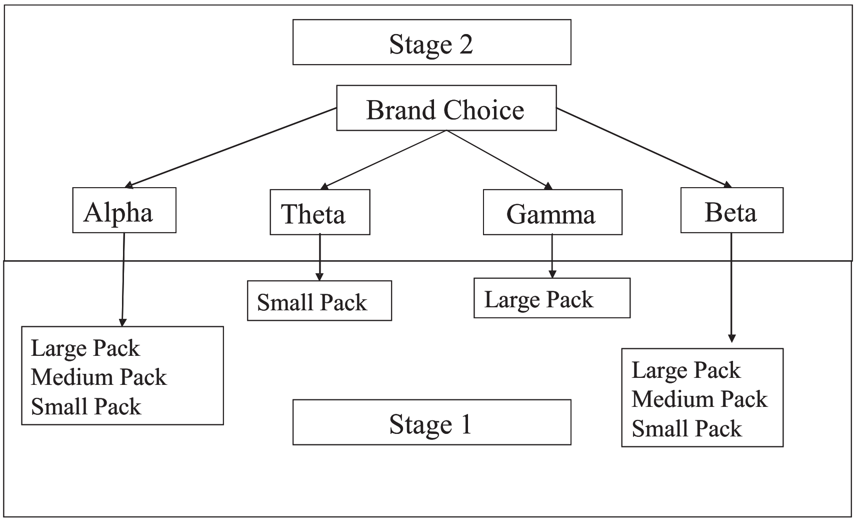

Figure 2 - Structure of Market

that was available to us. We operationalised promotion intensity as a continuous variable measuring the proportion of promoted brand sold in a month to overall promotion brands sold during the total period in the transaction panel data.

Summary of the operationalized marketing mix variables used to estimate the model is provided in table 2.

Table 2 – Marketing Mix variables

Name	Unit	Type	Collection Level	Data Source
Price	Rs per gm	Ratio Scaled	Monthly market level aggregate	Retail Survey
Promotion	%promotion	Ratio Scaled	Monthly market level aggregate	Retail Survey
Distribution	% Dealer stocking	Ratio Scaled	Monthly market level aggregate	Retail Survey

4.3.2 Attitudinal Variables

We have modeled these variables as consumer attitude towards various brands in the choice set. We believe that this is superior and more appropriate as compared to use of advertising budgets because, (1) attitude have a more direct relationship with choice than advertising budgets (2) inclusion of advertisement budgets assumes that all brand advertisements are equally effective, which may not be true. Also brand level advertising budgets are not available in the public domain in India.

The attitude data were available as scores that were measured on 18 statements on a dichromatic (yes/no) scale. However these attitudinal statements were highly correlated. On applying factor analysis (after Varimax Rotation) the 18 attitude statements converged into three distinct factors. We incorporated three statements (variables) which had the highest loading on these factors dropping all other statements.

4.3.3 Demographic Variables

Demographic variables were available for each household for head of household and respondent. We applied a classification tree analysis using brand choice as the independent variable and demographics as dependent on CART (see Appendix 2 for details) to create segments with interaction of demographic variables. This analysis was done on a calibration sample. These interactions were coded as dummy variables and included in the model. This was done to capture significant interactions within demographic variables. Apart from the demographic variables, brand specific dummy variables were introduced to capture any residual influence of brand on choice, which are not captured by any of the other variables. This may include order of entry effect, especially for the late entrants in the market such as brands “Gamma” and “Theta”.

A schematic representation of cart dummies is shown in figure 3.

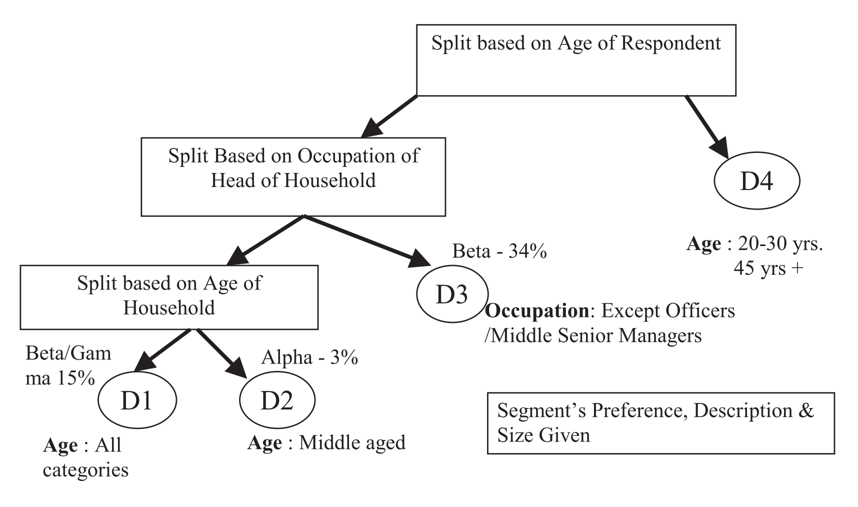

Figure 3 - CART Structural Dummies

The operationalization of demographic variables is given in table 3.

Table 3 Demographic Variables

Name	Unit	Type	Level	Source
Age of head of household	Dummy	Nominal (12 categories)	Disaggregate	Household panel
Age of Respondent	Dummy	Nominal (12 categories)	Disaggregate	Household panel
Education of Head of household	Dummy	Nominal (9 Categories)	Disaggregate	Household panel
Education Respondent	Dummy	Nominal (9 Categories)	Disaggregate	Household panel
Respondents Knowledge of English	Dummy	Nominal (6 categories)	Disaggregate	Household panel
Medium of Education of Head of Household	Dummy	Nominal (5 categories)	Disaggregate	Household panel
Medium of Education of respondent	Dummy	Nominal (5 categories)	Disaggregate	Household panel
Marital status respondent	Dummy	Nominal (5 categories)	Disaggregate	Household panel
Occupation head household	Dummy	Nominal (15 categories)	Disaggregate	Household panel
Occupation respondent	Dummy	Nominal (15 categories)	Disaggregate	Household panel
Household income	Rs. per family member	Ratio Scaled	Disaggregate	Household panel
Mother Tongue	Dummy	Nominal (19 categories)	Disaggregate	Household panel
Working Status	Dummy	Nominal (5 categories)	Disaggregate	Household panel
Owner of Washing Machines	Dummy	Nominal (2 categories)	Disaggregate	Household panel
Segment Dummy	Dummy	Nominal(2 categories)	Disaggregate	CART

4.4 Model Estimation

The estimation of the model was done in stages. Initially, the transaction level data was combined with the market level data, which provided details on marketing mix elements of all the relevant competitors. Brand choice was explained using relevant and available marketing mix elements and demographic details of the panelists in the transaction database. Additionally, market level information on attitudes towards various brands was incorporated by aggregating the scores of all panelists in the database. This provided the time dependent variation in attitude towards various alternatives.

The results of stage I estimation routine were used to identify demographic variables that were found to affect brand choice significantly. A matching routine was formulated to merge the transaction panel with the attitudinal panel using the demographic variables that significantly affected choice. The objective was to append disaggregated attitudinal scores of panelists to see if cross sectional variability of attitudes provided better explanation of the choice phenomenon. Since the transaction panel and the attitude panel had different sets of respondents, it was necessary to match them based on similarities in their profiles defined by the select set of demographic variables. If there were more than one panelist in the attitudinal panel that matched the profile of any specific panelist in the transaction database, the average attitude score across the matched members were used to append with the transaction panelist. Likewise, if any specific member in the transaction panel did not have a representation in the attitudinal panel, the matching was performed at a higher level of aggregation. This methodology ensured that all panelists in the transaction panel had a matched attitudinal score for every period.

The model was estimated using PHREG[2] procedure of SAS 8.01. The likelihood function was customized to model three alternatives (brands) from Oct-99 to May-00 and four alternatives (brands) from June-00 to Dec-00.

4.5 Quality of Fit

Model Fit was comparable with typical choice estimation in the research literature. The pseudo R^2 of the estimated model was 0.49 [17,18]. Fit results reported are at individual transaction level and not at the market aggregate level. Hit rate was computed to obtain predictive validity of the model. The hit-rate of the model was 70.23%, i.e. in 70.23% of records in our sample the model predicted choice matched with actual choice of consumer. The model was also validated on data provided for the Jan 2001- June 2001 period.

Aggregate market shares were computed at monthly, quarterly and semi-annually level using estimated coefficients and was compared to actual market share of the panel data. The charts comparing actual market shares to predicted market shares for both calibration sample and validation sample are given in figure 4 and 5. In the calibration sample the mean of the absolute difference[3] in the predicted and actual market share of the alpha brand was 3.47 and the beta brand was 2.77. In the validation sample the mean of the absolute difference in predicted and actual share was 3.06 for brand alpha and 4.51 for brand beta. Since the share of the other two brands was considerably less, the graphs were of little significance. Also, with less number of data points, the volatility of the shares are much higher and the predictability is much less.

4.6 Discussion of Coefficients

The impact coefficients of significant[4] marketing mix, attitude variables (conditional variables) and polytomous variables are reported in table 4.

Table 4 – Parameter Coefficient

Variable	Parameter Estimate	Signifi-cance Level
Economy	0.87	0.02
Performance	0.55	0.04
Distribution	-0.02	0.04
PACK Composite	0.29	0.00
Total Awareness	0.03	0.00
CART1_Beta	-0.70	0.00
CART1_Gamma	-0.80	0.00
CART1_Theta	-1.70	0.00
CART2_Beta	0.87	0.00
CART4_Beta	-0.16	0.07
CART4_Theta	-0.94	0.00
CART6_Gamma	-1.04	0.01
CART6_Theta	-2.03	0.05
CART7_Gamma	1.26	0.02
Beta_Repondent Speaks & Reads English	-0.28	0.03
Beta_Repondent Speaks English	0.44	0.02
Beta_Medium of Education_Hindi	0.21	0.01
Beta_Married Housewife	-0.54	0.02
Beta_Occupation_Business_0Employee	0.29	0.01
Beta_Occupation_Other	0.45	0.02
Beta_washing machine owner	0.25	0.00
Beta_Full time working	-0.39	0.01
Beta_Part time working	-0.28	0.07
Gamma_Age of Head Household 41-44 YEARS_AGE	0.69	0.00

Gamma_Age of Head Household 59 + YEARS_AGE	0.26	0.06
Gamma_Age of Respondent_Others	-0.96	0.01
Gamma_Medium of education_other	0.54	0.00
Gamma_Mother Tongue_Bengali	0.82	0.04
Gamma_Mother Tongue_Malyalam	1.59	0.02
Gamma_Mother Tongue_Others	-0.58	0.09
Gamma_Mother Tongue_Sindhi	-0.77	0.07
Gamma_Married Housewife	0.85	0.00
Gamma_Occ_Business_1-9Employee	-2.05	0.05
Gamma_Occupation_Other	-1.02	0.03
Gamma_Occupation_Self Employed	-0.62	0.03
Gamma_Occupation_Shop	-0.67	0.10
Gamma_Occ_Unskiled Worker	-0.66	0.06
Gamma_Full Time Working	-0.96	0.00
Gamma_Part Time Working	-0.74	0.01
Theta_Age of Respondent_Others	-1.73	0.00
Theta_Occupation_Unskilled	0.95	0.01

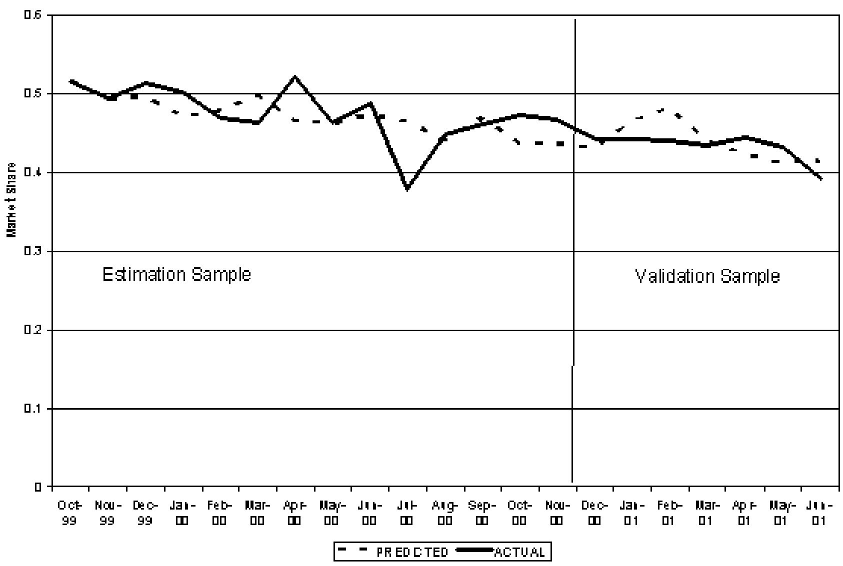

Figure 4 - Brand Alpha Monthly Aggregated Market Share

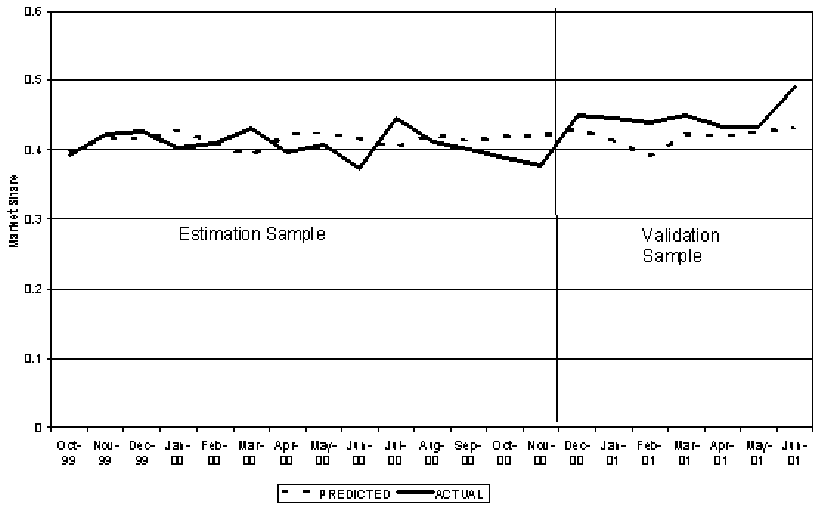

Figure 5 - Brand Beta Monthly Aggregated Market Share

The two composite attitudinal variables Economy and Performance came significant and positive in direction. The pack level composite variable based derived from stage one logit model was also significant, thereby suggesting that the pack level tactics had a significant effect on overall brand choice. The parameter estimate of distribution turned out to be negative. A plausible explanation could be the aggregated nature of the marketing mix variables (available at the market level). Some enquiries in the market revealed that the firm marketing brand Alpha had reduced its distribution reach but at the same time had intensified its sales initiative in a limited geographical area. One of the limitations of the databases available is that they do not capture the intensity of field level selling activity, other than the span of distribution.

Qualitative assessment of the Indian market points at the relevance of distribution as a driver in making choices. Since, a large proportion of stores do not carry stocks of a specific brand, distribution would hold an enormous influence on the purchase pattern. This contradiction between hypothesis regarding market behavior and the empirical finding points at the possibility of limitations in the data collection methodology used in India.

4.7 Managerial Implications: Developing a framework to evaluate marketing spends

The major variables influencing premium detergent brand choice in the Mumbai market seem to be attitudinal variables such as perception regarding the efficacy (PERFORMANCE) of the brand, closely followed by the perception on the value-for-money (ECONOMY). Field level promotional activities such as price-offs, freebies associated with different pack sizes of the same brand also seem to impact the choice of the brand, although the impact was low. Base price reduction is generally resorted to quite infrequently and hence the price elasticity measure was found to be statistically insignificant. Surprisingly, the distribution variable had a negative impact on brand choice.

A critical result drawn from the model is the importance of attitudes in affecting behavior in the purchase of premium detergents. More importantly, the fact that only two attitudinal factors were relevant in driving behavior and of them, the perception of product efficacy is more important than the perception of value-for-money. This finding will be relevant for marketing managers responsible for premium detergents in designing effective brand development initiatives. In the Indian market, it is generally believed that brand attitudes are primarily influenced by advertising. If this assertion is assumed to be true, this research makes a case for better management of consumer attitudes through the deployment of appropriate advertisements.

5. Conclusion

In this paper we have presented a novel combination of tools and techniques to develop a predictive model for emerging markets linking marketing mix variables to market share performance. The model development can help managers to evaluate the performance of the initiatives like brand building programs or promotion programs. It can also assist managers to apportion the limited marketing resources for meaningful investment to long term market-building activities. This research initiative is perhaps the first attempt at linking consumer attitude data along with marketing mix information to explain actual consumer behavior.

While it is difficult to significantly modify investments made in information gathering in the near term, proper tooling to suit the poor data quality can provide significant gains to managers in building decision support infrastructure. We provide a way to exploit large volumes of available consumer data with the companies for developing appropriate decision support systems for managers in emerging markets.

Due to limitations in the existing data gathering practices, the model does not incorporate various relevant marketing mix variables. For instance, qualitative findings indicate the importance of distribution variable (availability of brand at POP) in the Indian markets. However, the pertinent information is available only at the market aggregate level and hence its ability to explain transaction level choice behavior is limited. It highlights the need for a planned investment in customer data in the emerging markets for managers to derive maximum benefits of prediction based models.

Endnotes

1. Brand names have been camouflaged to maintain confidentiality.

2. Is a statistical procedure in SAS, which estimates the coefficient of a Hazard Model. The model is suitably modified to estimate a Multinomial Logit model.

3. This measure simply averages the difference (modulus of difference) between the predicted and the actual market shares over all months. This is directly related to the explanatory power of the estimated model.

4. Mostly at 95% significance levels, although variables at 90% significance were also reported since the variables were considered managerially important enough to be reported.

5. For details see Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. Classification and Regression Trees. *Wadsworth International Group*, Blemont, California, 1984.

References

1. Banerjee B. Raghuram G. Koshy A. Foodworld (C): The Road Ahead. Working Paper, Ahmedabad: Indian Institute of Management, 1999.

2. Banerjee A. Banerjee B. Effective Retail Promotion Management: Use of Point of Sales Information Resources. *Vikalpa* 2000;25(October):51-59.

3. McFadden D. The Choice Theory Approach to Market Research. *Marketing Science* 1986;5:275.

4. Chiang J. A Simultaneous Approach to the Whether, What and How Much to Buy Questions. *Marketing Science* 1991;10:297-315.

5. Guadagni PM. Little JDC. A Logit Model of Brand Choice Calibrated on Scanner Data. *Marketing Science* 1983;2:203.

6. Chintagunta, Pradeep K. Variety Seeking, Purchase Timing, and the 'Lightning Bolt' Brand Choice Model. *Management Science,* Apr99, Vol. 45 Issue 4, p486, 13p

7. Seetharaman, P.B. Probabilistic versus random-utility models of state de pendence: an empirical comparison.*International Journal of Research in Marketing,* Mar2003, Vol. 20 Issue 1, p87, 10

8. Kamakura WA. Srivastava RK. Predicting Choice Shares Under Conditions of Brands Interdependence. *Journal of Marketing Research* 1984;21:420.

9. Moore WL. Lehmann DR. A Paired Comparison Nested Logit Model of Individual Preference Structures. *Journal of Marketing Research* 1989;26:420.

10. Raju JS. Dhar SJK. Morrison DG. The Effect of Package Coupons on Brand Choice. *Marketing Science* 1994;13:145.

11. Ben-Akiva M. Boccara B. Discrete choice models with latent choice sets. *International Journal of Research in Marketing* 1995;12:9.

12. Ben-Akiva M. Walker J. Bernardino AT. Gopinath D. Morikawa T. Polydoropoulou A. Integration of Choice and Latent Variable Models. *Proceedings of 8th International Conference on Travel Behavior.* Austin, USA, 1999

13. Kalidas A. Dillon WR. Yuan S. Extending Discrete Choice Models to Incorporate Attitudinal and Other Latent Variables. *Journal of Marketing Research* 2002;39:31.

14. Gupta S. Chintagunta PK. On using demographic variables to determine segment membership in logit mixture models. *Journal of Marketing Research* 1994;31:128.

15. Kalyanam K. Putler DS. Incorporating demographic variables in brand choice models: An indivisible alternatives framework. *Marketing Science* 1997;16:166.

16. Jovicic, Goran; Hansen, Christian Overgaard. A passenger travel demand model for Copenhagen. *Transportation Research: Part A,* May 2003, Vol. 37 Issue 4, p333, 17p

17. Maddala GS. *Limited Dependent and Qualitative Variables in Econometrics:* Cambridge University Press, 1986.

18. Guadagni, Peter M.; Little, John D.C.When and what to buy: A nested logit model of coffee purchase.*Journal of Forecasting*, Jun/Jul 98, Vol. 17 Issue 3/4, p303, 24p

APPENDIX 1

Description of the Brands:

“Alpha” and “Beta” are premium detergent brands in the Indian market, which are marketed by two formidable consumer goods companies (both multinational companies). Both brands have been in the market for over 4-5 years and command a very high awareness rating. Premium detergent segment comprises of brands of detergent that can be safely used in a washing machine, an appliance that has very recently been adopted by the middle-class Indian consumer. “Alpha” is the first entrant in this premium segment followed by “Beta” and then “Gamma” and lastly “Theta” that has been introduced in mid-2000.

Brand “Gamma” is the second offering from the same company that markets brand “Beta” and was introduced in October 1999. It is perceived to be slightly inferior to the other two formidable brands, which are internationally well known. On the contrary, “Gamma” is a “home-grown” brand and is not perceived to have the same performance characteristics as “Alpha” and “Beta”. The price of “Gamma” is significantly lower than the other two, however, the market in general perceives it to be in the same price segment as “Alpha” and “Beta”.

Brand “Theta” is an internationally known brand that the makers of brand “Alpha” had just introduced in mid-2000. It is expected to provide stiff competition to both the established players, Internationally, it is a better known brand than “Alpha”, although in India it is a novice. It is also priced in the same range as “Alpha” and “Beta”.

APPENDIX 2

Classification and regression tree analysis (CART) is conceptually similar to thee AID techniques (Automatic Interaction Detection). It is a tree-based model that uses a criterion variable, categorical or interval-scaled, and attempts to split the sample into homogeneous clusters while at the same time ensuring maximum heterogeneity across different clusters WRT (the criterion variable). The methodology used to achieve this objective adopts a sequential splitting process to separate the sample into smaller groups based on the homogeneity within the clusters. The algorithm identifies the predictor variable and the splitting rule using the above methodology.

A general exposition of CART models is given below:

RESPONSE = f (variables describing respondents' characteristics, past credit handling Behaviour)

f = non parametric function

A CART model is similar to regression based models since it explains causal relationships. However, unlike regression models which explain a mathematical relationship among predictor variables that predict a dependant variable, tree-based models such as CART5 identify splitting criteria that regroup the sample into homogeneous cells. A similar technique used quite extensively in marketing research literature is CHAID.