IMPACT OF TIME AND ORDER OF MARKET ENTRY, ADVERTISING, AND POSITIONING ON THE EXPECTED MARKET SHARE OF A NEW PRODUCT: AN EMPIRICAL TEST

by

THERESA IDELLA CARTER

B.A., Spelman College
(1980)

and

ZOFIA BARBARA MUCHA

Magister Ekonomii, Akademia Ekonomiczna w Poznaniu
(1978)

SUBMITTED TO THE ALFRED P. SLOAN SCHOOL
OF MANAGEMENT IN PARTIAL
FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1982

Signature of Author ______________________________
Alfred P. Sloan School of Management
May 19, 1982

Signature of Author ______________________________
Alfred P. Sloan School of Management
May 19, 1982

Certified by ______________________________
Glen L. Urban
Thesis Supervisor

Accepted by ______________________________
Jeffrey A. Barks
Director of Master's Programs

IMPACT OF TIME AND ORDER OF MARKET ENTRY, ADVERTISING, AND POSITIONING ON THE EXPECTED MARKET SHARE OF A NEW PRODUCT: AN EMPIRICAL TEST

by

THERESA IDELLA CARTER

and

ZOFIA BARBARA MUCHA

Submitted to the Alfred P. Sloan School of Management on May 19, 1982 in partial fulfillment of the requirements for the Degree of Master of Science in Management

ABSTRACT

An empirical study was conducted to ascertain if there exists a relationship between the expected market share of a new product, its positioning (quality), advertising and order and timing of market entry for low costs consumer products. Upon analyzing the data, such a relationship was detected and a series of alternative models of interaction between variables was constructed. The model selected as the best descriptor of the underlying market mechanism was theoretically compelling met the formal statistical criteria (goodness-of-fit and significance of variables). The model was used to predict ex post market share for a new set of consumer goods. Goodness-of-prediction was assessed with a number of traditional as well as specially designed measures.

Predictive power of the model was very good. Noticeable differences in goodness of predictions for early and late market entrants led to reestimation of the model over an enlarged sample. The resulting descriptor of market mechanisms consists of 2 models. The first model depicts market phenomena for the early entrants and the second for the later market entrants.

Thesis Supervisor: Glen L. Urban
Title: Professor of Management Science

ACKNOWLEDGEMENTS

We wish to thank our advisor, Glen Urban, for guidance, help and friendship, our readers and teachers, John D.C. Little and Richard Schmalensee, for instruction and comments, our colleagues for friendship and good times, and finally our typist, Joanne Sorrentino, for endless patience and excellence. Our two years at Sloan have been a rewarding and enriching experience.

Theresa S. Carter

Zofia B. Suda

For their love and encouragement which are always sources of inspiration, I thank my parents, Kirk and Idella.

Theresa

Mojej Rodzinie za cierpliwość i zachętę,

moim Profesorom za wiedzę,

moim Kolegom za przyjaźń - dziękuję.

Zosia

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT.)

CHAPTER 1

PROBLEM STATEMENT AND OUTLINE OF RESEARCH

Development of a new product and its market introduction is a costly and risky endeavor. Managers have to resolve many portent issues but have few firm inticators that might help them to arrive at good, sound decisions. Some of the issues the company faces are primarily internal, (eg. how to allocate R&D budget between different stages of product research). Certain optimization models have been proposed here. Other issues that the company faces depend, to a large extent, on the environment in which it operates. Indirectly, customers and competitors strongly influence the company's behavior. Firms are very much aware of that and spend considerable resources on market research and analysis of competitive strategies. The obtained information is subjected mostly to qualitative evaluation. Only recently, have several strategic marketing models been developed. One of the least understood problems of strategic marketing is the question of order of market entry. Most of the practitioners and theoreticians believe that order of market entry - whether the company enters the market first with a new product vis. is an innovator, or it enters second or later - has bearing on a company's prospects. However, this is probably where the concensus ends. One group of marketers believe that the company will achieve higher market share and concommitant benefits if it is first to enter the market. This group thinks that later entrants face high costs and lean rewards. The other group of marketers holds that earlier entrants pay the dues of developing the market and educating the public about their products. Later entrants are spared these costly efforts. Also, next-in brands appear only if a given product type wins the acceptance of the consumer (it is not a fad). All in all, the second group says, benefits of procrastination are large.

Confusion on the subject of impact of order of entry on a company's fortunes is hardly astonishing if we consider how many interrelated forces influence the market share of a given product. To make a resolute statement, one would have to isolate the impact of order of entry, advertising, pricing and perhaps many other variables. Unfortunately, little theoretical or empirical work has been done that sheds light on these involved concepts.

In this situation, we have decided to examine empirically the following issues:

(1) Is there any relation between order of market entry and market share?

(2) If evidence shows that such a relationship exists, how is it modified by a simultaneous influence of other magnitudes, such as advertising, product quality, and timing of market entry.

This thesis contains a description of our analysis and obtained results. In our research we went through the following steps:

(1) We searched the literature for clues on possible relationships between market share, order of entry and other variables.

(2) We did an extensive analysis of the data describing market share, order of entry, product perceptions, advertising and timing of market entry. We limited

our analysis to those variables because of difficulties in finding reliable information and to avoid the embarras des richesses at this early stage of research. We constructed the best model describing the relationship between the above-mentioned variables.

(3) We used our model in predictive test over a new, analogous data set. We reasoned that the model that passes five tests of prediction has to be a good descriptor of the underlying market mechanism. We evaluated the predicitve power of the model using a series of goodness of prediction measures, some of them specially tailored to the needs of our work.

(4) We used a combined data set (sample over which the model was estimated and sample employed for predictive tests) to construct a richer and more detailed model of market phenomena.

(5) We suggested how our model can be used in managerial practice and discussed implications of obtained coefficients and trade-offs involved in decision making.

(6) Finally, we pointed out some areas of possible future research.

CHAPTER 2

REVIEW OF RELATED LITERATURE & RESEARCH

Theory and empirical research relating order of market entry, advertising, product quality, and market share is quite scanty. It is rather paradoxical that topics so crucial in the domain of strategic marketing, received so little attention. A few researchers addressed the problem directly by explicit consideration of the entry effect in their theoretical models or empirical research. Many others touched upon the subject while discussing other marketing concepts - advertising, innovation, etc. None of the studies we read uses a model of interaction between market entry, advertising and other phenomena similar to the one developed in this thesis. Still, many of the findings of different researchers corroborate or highlight certain regularities that we have detected in the course of our work.

The following pages contain brief reviews of the literautre related to our own research. In the discussion we emphasized those ideas that directly correspond to concepts we have utilized in our study.

In economic literature, Richard Schmalensee probably gives the most through treatment of an order of entry effect on market share. In his paper entitled "Product Differentation Advantages of Pioneering Brands", R. Schmalensee presents a model of the market consequences of order of entry to the market. The first market entrants obtain demand advantages. If the economies of scale are present, demand advantages can be used to deter prospective later entrants.

Schmaleness's model is built using a series of assumptions concerning the behavior of consumers and strategies available to the companies. The discussed products are assumed to be consumer experience goods, that is, a consumer can only gain knowledge about product effectiveness if he tries the product. Tastes of consumers are fixed, there is no word of mouth information, and all products work. Products have identical characteristics, including cost functions. Companies do not advertise. To concentrate attention on entry effects and demand advantages, Schmalensee assumes that the first entrant does not change his price in responce to the entry of the second compay. A´priori consumers don't know that all products work therefore, they initially attach a certain probability to the possibility that the new product will not work. As a result, some loss is expected if the product fails. Once the consumer buys the goods, he convinces himself about the quality of the product and continues to use it.

The optimal marketing strategy of the first entrant is to charge a low initial price for its product to induce a large number of consumers to try it. Later the price can be raised since consumers have already tried the product and use it. Such a strategy is known as penetration pricing in marketing. It yields higher revenues than a constant pricing strategy.

Schmalensee analyzes pricing strategy of the first entrant under two polar conditions (1) static expectations (consumer expects the most recent price to stay put) and (2) perfect foresight (consumers can predict the path of price changes). Under prefect foresight assumptions, the present value of the revenues of the first firm is always lower, since it is harder to persuade the consumer to buy the product.

The second entrant faces 2 classes of consumers: those who tried Product I and use it and those who never bought Product I. The second firm may try to charge a lower price than the initial price of the first entrant. A low price may induce consumers who never bought Product I to buy Product II. If the initial price of the second entrant is extremely low it may entice even the present, satisifed users of Product I to buy Product II. The success of a very low pricing strategy is highly dependent on the cost function of the product. If there exist economies of scale, then a low pricing strategy for the second entrant is unlikely to be feasible.

The major disadvantage of the second-in brand is that the first-in brand sets the standard against which each of the following brands are compared. The first-in brand can only be compared with unlike items (eg. the first floride toothpaste is compared to a plain toothpaste).

The demand disadvantage of the later entrant remains qualitatively the same under both static expectations and perfect foresight assumptions.

In the last part of his discussion R. Schmalensee relaxes some of his beginning assumptions. For example, he considers a situation when the uncertainty about the quality of the product differs by brands. In such a case the demand gap for later entrants is widened since trial of the first brand alone does not resolve all the uncertainty about the other products. This can be somewhat remedied if the next-in product attempts to differentiate itself purposefully from the first-in brand. R. Schmalensee suggests, that if advertising is included into competitive analysis, the later entrant has an option of combining pricing and advertising to win over the consumer.

One of the cornerstones of Schmalensee's models is the lack of information about the product on the part of the consumer. The author admits that the word of mouth transmission of information changes substantially the position of the consumer. Since information is more eagerly sought and is more readily available for higher priced items, Schmalensee says that his model is more likely to hold for lower cost products.

Eric von Hippel considers lagbetween order of entry, "response time", to the market in the context of appropriability of the benefits of innovation.

Eric von Hippel, in his paper entitled "Appropriability of Innovation Benefit as a Predictor of the Functional Locus of Innovation", discusses what may explain differneces in the locus of innovation in various industries. Von Hippel suggests that depending on the ability of the innovator to capture the benefit of innovation, the prospective innovator could be inclined or disinclined to innovate. The benefit of innovation may be embodied in the firm's output when the innovation imparts values to the firm's product. It may also be non-output-embodied, vis. when the benefit can only be realized by sale or liscensing of the innovation to others. Von Hippel develops theoretical "boundary" cases and describes mechanisms for operating under these extreme, unlikely conditons. Later he proceeds to discuss examples of actual situations in certain industrial goods segments of the economy.

In the extreme case (purely theoretical) when the innovator can reap all the benefits from both output-embodied knowledge and non-output-embodied knowledge, he has every incentive to innovate. In real life however, as evidence show, capturing benefits from output-embodied knowledge may be conditional on some circumstances, and the ability to caputre benefits from non-output-embodied knowledge, despite legal protection afforded by patents and trade secrets, might be very difficult.

An ability of the innovator to benefit from an output-embodied knowledge, according to von Hippel, will depend on the extent to which an innovator can establish a quasi-monopoly with respect to his innovation, either by (1) exclusion of any outside firm (firm engaged in a different line of business but potentially interested in the innovation) from enjoying benefits stemming from his innovation and sharing the benefits with all the other firms already plying the trade, or by (2) selective sharing of knowledge about innovation with just one (few) other firm(s) in his industry.

An innovator may also attempt to reserve the benefits of output-embodied knowledge for himself. Patents and trade secrets laws are of some help. Another mechanism that can be of use in monopolizing benefits is termed by von Hippel as "response time". Competitors of an innovator may be kept in the dark about development of new techniques or products until the product becomes available on the market and can possibly be reverse-engineered. The importance attached to the secrecy surrounding future plans and developments in most of the companies is, according to von Hippel, an indicator of value of the "response time" as an aid to capturing benefit from output-embodied knowledge. The innovator realizes benefits in the form of increased profits and/or larger market share.

Von Hippel warns that the value of the "response time" may be a function of various situation-specific factors. "One such factor is the length of response time divided by the length of customer decision cycle..." The bigger the ratio, the better off the innovator is. "A second factor ... involves a learning curve: the more units produced during the response time period and the steeper the learning curve, the greater the production cost advantage an innovator can accrue relative to potential imitators. The third factor is the size and "indivisibility" of the production plant investment innovation requires relative to market size." In these situations, the lead time advantage may be thought of as a pseudo patent granted by the market to the innovator.

The ideas discussed by von Hippel organize concepts pertaining to innovation into a series of models that allow explanation and perhaps even prediction of the locus of innovation. Von Hippel's theory finds its primary application in the realm of the industrial goods industries and is somewhat less useful in annalysis of the consumer goods industries.

W. J. Lane developed a "descriptive model of a market with differentiated consumers and firms in which both prices and locations were endogenous and in which entry was sequential and endogenous".

In his paper, "Product Differentiation in a market with Endogenous sequential entry," Lane relaxes some of the assumptions about product homogeneity that are common in economic literature. He does this in order to describe how the firm choses the "product variant it will produce and the price it will charge.

Lane contends that market entrants generally secure better locations, near the center of the market, and this positioning results in higher profits. In addition to higher profits, these strategic locations assist first entrants in "deterring further market entry". Lane's model demonstrates that when new firms enter the market established firms experience a reduction in market share and prices. In view of this, early entrants must be concerned about their product positioning. Lane concludes that the extent to which the early entrant is able to discourage market entry corresponds directly to the fixed costs he incurs. For instance, deterring additional market entry is feasible and advantageous when fixed costs are high and at best difficult, when these costs are low.

The influence of advertising on the market share of a given product is widely acknowledged, although there is some variation in interpretation of mechanisms through which advertising operates. Bond & Lean are specifically interested in interaction between advertising and order of market entry in "Consumer Preference, Advertising, and Sales: On the Advantage From Early Entry".

Bond & Lean critically review the "assumption that advertising per se, is a barrier to entry" and offer alternative theories concerning consumer preferences for existing brands. They present a profit-maximizing model, that uses promotion and sales data for brands of prescription drugs, to determine the impact of consumer preferences for exisitng brands on the relationship between advertising and sales. The results of their study imply that when brands are "qualitatively identical" promotion and sales will be greater for the early entrant than for later entrants.

The empirical test of their model is subject to criticisms that the authors themselves identify. First, the data* base consists of promotion and sales figures from only one market, diuretic and combination diuretic antihypertensive prescription drug products. Therefore, generalizations may be unreliable. Secondly, prescription drugs represent a special type of consumer product because physicians act as an intermediary between the producer and the consumer/patient. Therefore, the physician becomes a surrogate consumer. It is obvious that a physician selecting a product among brands of drugs is not readily comparable to the average consumer selecting a brand of shampoo. The product price factor is also reduced in importance since although the consumer pays for the product, it is selected by the doctor. Unless the individual is using the drug for an extended period of time he will probably defer to the physician's judgement and not inquire about alternative (cheaper) brands.

---

*Data collected from 1956 - 1971.

Bond and Lean chose to think of order of entry in two ways. First, they used a chronological ranking of order and second, an "FDA therapeutic gain rating" which identified "brands that were first to incorporate new therapeutic advantages" (e.g. a new and better product).

The empirical results suggest that:

- Physicians are more receptive to the promotion efforts of early market entrants.
- Order of entry is negatively correlated with sales.
- A substantial sales advantage can be expected by first entrants with a product that has a therapeutic gain.

At the risk of making too broad a generalization, we propose that Bond and Lane's last finding may be interpreted as the consumer's (Physician's) requirement that a product be improved above the standards of existing brands before he can be induced to switch. This idea as well as the negative correlation of order of market entry with sales found by Bond & Lean become quite significant when one considers the managerial implications.

One of the most renowed analyses of market-related phenomena came from PIMS research.

In an article entitled "Marketing Costs In Consumer Goods Industries", Buzzell and Farris* present an empirical study that attempts to explain dependent variables defined as ratios of advertising and promotion costs to sales, marketing costs to sales,

---

* This information was gathered from a critique of the Buzzell & Farris article because the full text was not available. The critique was prepared by W. T. Robinson, Ph.D. candidate at the University of Michigan.

and sales force costs to sales using market pioneering, market share and a number (17) of other independent variables. The sample came from the PIMS-Phase III data base that contained information about 103 consumer product manufacturers.

They find a substantial degree of correlation between market share and market pioneering (r=.41). Based on this finding Buzzell & Farris conclude that "the coefficients of these two variables can not be treated as separate, distinct factors influencing marketing costs".

The theory that advertising reenforces brand loyalty and consequently results in restriction of market entry by competition, is well established and usually supported with cross-sectional correlations between advertising and profits. Thomas Nagle, in a working paper "Does Advertising Really Create A Barrier To Entry?", finds this evidence insufficient to conclude that advertising, per se is a barrier to market entry.

Nagle contends that the cost of sampling for consumers is high when they have limited or no information about products. Advertising provides the consumer with information that will allow him to make a more educated decision and hence increase his probability of purchasing a superior brand. Infromation inherent in advertising increases utility per sampling dollar. Along this same line of thinking, Nagle suggests that brand "loyalty" may simply be the consumer's reluctance to sample a potentially inferior brand about which he is ill informed. If this is the case, advertising should promote brand switching, which should make for a more receptive climate for market entry rather than act as a deterent.

In view of the fact that new product development is both risky and expensive, it is important that companies have a well planned strategy for market entry. Thus, consumer products manufacturers are always seeking information that can be used to optimize their marketing strategy (particularly their entry strategy).

In their paper, "Market Entry Strategy Formulation: A Hierarchical Modeling and Consumer Measurement Approach", Urban, Johnson & Brudnick present a "system of models and measurements designed to support such a strategy". Urban et al. define the competitive structure of the market with "product attributes, usage situations or user characteristics". They estimate purchase probabilities with a logit model and judge hierarchies on their predictive ability relevant to consumer choices of brands other than their first preferences. Data gathered from a simulated shopping setting is also used to determine which hierarchy best depicts the competitive structure. The empirical portion of the study applies these hierarchical models to the coffee market.

The entry model, presented by Urban, Johnson & Brudnick, defines entry share relative to the share obtained by the first market entrant. It does not include entries following the first but preceeding the new product, nor are advertising/promotion expenditures incorporated. However, both of these issues are important and should be addressed using an empirical approach.

The article by Urban et al. concludes our review of relevant literature. Despite a thorough search, including a computerized literature search, we found no material dealing with market share, advertising, perceptions of products, order of entry, and timing of entry in conjunction. We did find interesting studies that dealt with one or a few of the topics separately. The lack of uniform terminology in the literature, even for the basic concepts, is an indication of how new this quantitative, modelling approach is in marketing. Literature that discusses marketing ideas in managerial terms or in an intuitive manner is aboundant but only marginally useful for the purposes of this thesis.

CHAPTER 3

DESCRIPTION OF DATA

Throughout our thesis we will be making constant reference to variables used to model the relationship between market share, order of market entry, timing of entry, advertising and consumer perceptions about products. This chapter contains definitions of the variables used to express the aforementioned notions and notes sources of the data.

Actual national market shares for products cutting across a variety of categories will understandably fluctuate substantially. In order to normalize market share, we use a variable named "Share Index (SHINDEX)", which is a ratio of the market share of a given product to the market share of the category's first entrant. Using SHINDEX makes the share for all first entrants equal to one. This avoids the problems of varying category sizes that can misrepresent an actual market share value (e.g. a market share of 20% in a category of 8 products is better than the same share in a 2 product category).

The data was gathered using ASSESSOR studies on various product categories. The information was carroborated by data in the firm's possession (i.e. warehouse with drawls and shipments).

---

* Silk and Urban have developed a model, ASSESSOR, which is "designed to estimate the sales potential of new packaged goods before they are test marketed". ASSESSOR predicts the new product's steady state rational market share and the sources from which it will obtain that share. Predicted market share is verified by Nielsen and/or SAMI data. For a more detailed description, see Journal of Marketing Research, May 1978, pp. 171-191.

Consumer perceptions about products or product positioning is an important concept in the construction of our model. We refer to this variable as "RELRET". RELRET describes how product positioning influences probability of purchase. RELRET is a ratio of conditional probabilities. The conditional probability referred to is the probability of purchase given the product is evoked. An evoked product is one the consumer presently uses, has used in the past, or would consider buying. Hence,

$$\text{RELRET} = \frac{\Pr(\text{Purchase} \mid \text{Evoked}) \text{ for a later entrant}}{\Pr(\text{Purchase} \mid \text{Evoked}) \text{ for the } 1^{st} \text{ entrant}}$$

Silk and Urban defined the mathematical specifics for estimating probability of purchase based on preference. They define probability of purchase as:

$$P_i(j) = \frac{[\hat{V}_i(j)]^{\beta}}{\sum_{K=1}^{m_i} [\hat{V}_i(K)]^{\beta}} \qquad \hat{V}_i(j) > 0$$

where: $P_i(j)$ = probability that consumer i chooses brand j

β = parameter

m_i = number of brands in the consumers evoked set

$\hat{V}_i(j)$ = estimated preference of consumer i for brand j

Silk and Urban estimate β using data gathered from consumer interviews in which "brands last purchased by respondents are identified and preference measures are obtained for their" evoked sets.

"ORDER" of market entry is another explanatory variable included in our model. "ORDER" takes on positive integer values corresponding to a brand's chronological market entry (e.g. for the first market entrant in a given cateogry order = 1). Information about order of market entry of the various products was obtained directly from manufacturers.

The fourth variable employed in our model is advertising. We gathered advertising expenditure data for a three year period (1978, 1979, 1980). We obtained the information from the Leading National Advertisers Multi-Media Report Service, which provides the dollar amounts spent by companies on advertising in six major media: consumer magazines, newspaper supplements, network TV, spot TV, network radio, and outdoors.

Our basic constructs were 3-year average expenditures and we used these numbers to build all of the other forms of advertising variables. Our rationale for using a 3-year average as the basic building block was as follows:

a) It is doubtful that advertising in distant past years has significant influence on present purchasing behavior.

b) The three year period keeps the dollars spent comparable since some of the products have only been on the market for the last four years. Changing advertising mix and purchasing power per advertising dollar make it difficult to equate outlays made in years separated by long periods of time.

c) Using a three-year average helps to smooth out advertising expenditures. For some of the products, eg. cigarettes, there was pronounced pulsation in advertising expenditures.

d) Averaging allows for an advertising "lag" effect-advertising in immediately preceding periods influencing purchase of subsequent periods.

The following advertising variables were alternately introduced into equations:

$$\text{I. Percentage Advertising} = \frac{\text{average expenditure for next-in (2nd, 3rd, etc.) product}}{\text{Sum of average advertising for all products}}$$

$$\text{II. Relative Advertising} = \frac{\text{average expenditure for next-in product}}{\dfrac{\text{Sum of average advertising for all products}}{\text{Number of products}}}$$

Both percentage and relative advertising variables describe the relative "strength" of advertising for a certain product in a given product category and abstract from the actual dollar amounts spent. This is an important characteristic since advertising in some product categories was at a level of 20 million dollars per product [cigarettes] while other product categories had relatively low expenditures per product [$ hundreds of thousands per year].

The choice of 3-year average expenditures as our basic building block was somewhat arbitrary but not accidental. We tested alternative variables, identical in form with the variables based on 3-year averages but using 4-year average numbers instead. 4-year average expenditures were computed from data for years 1977, 1978, 1979 and 1980. Estimation (fit) results for analogous equations built with the variables based on 3-year averages were always better than estimation results using 4-year average based variables. Such an outcome was not very suprising because the 4-year average excessively penalized very recent products- those with a one or two year market tenure - by attaching too much weight to advertising in early years (1977, 1978). Deducing that with periods longer than four years the weighting problem would only be exacerbated, we settled on the 3-year average as our basic building block for the advertising variables.

The last type of variable, "lag" (timing of market entry) was introduced into the model to capture possible advantages that may accrue to a company when competitors are slow in entering the market.

The presumption was that the longer the competitors procrastinate, before introducing their version of a new product, the smaller the market share they can hope to capture. Information on dates of national launch of brands was gathered by direct questionning of the manufacturers.

The following "lag" variables were alternately employed in the regression equations:

I. "lag" - Time in years between the introduction of first and the next products.

II. "lagbetween" - Time in years between the introduction of the first and second product, second and third product, etc.

All the variables described above were used throughout our research. In our early work we tried some other variables forms, eg. ratio of advertising, absolute advertising, etc. We rejected those variables on either statistical or theoretical grounds. The definition of tested variables and a brief justification for their rejection is presented in the historical note included in Chapter 4, "Exploratory Analysis of Data and Interactive Model Building".

Ratio of advertising is one of the discarded variables that deserves special note at this point.

$$\text{Ratio of advertising} = \frac{\text{average expenditure for next-in product}}{\text{average expenditure for first-in product}}$$

The form of the ratio of advertising variable is logically identical to the forms of the variables shindex and relret. It was very tempting to maintain consistency in the definition of the variables. Yet, unlike shindex or relret, ratio of advertising is susceptible to "degeneration". If the first-in brand has very low or no advertising (as in the case of Corricidin and is likely to occur with other products), ratio of advertising increases infinitely. Using average advertising of all products in the category as the point of reference (divisor) removes that problem, barring highly improbable circumstances. Similar issues would not arise in the case of relret or shindex becuase of the way these two variables were defined. For example, a brand with 1% national market share is not a national brand sensu stricto and it is excluded from consideration in our model. Of course, models employing ratio of advertising are worth exploring, provided one has only "well-behaved" categories of products in the sample (all products in the category advertise at comparable levels). As a final comment, we might add that by choosing relative or percentage advertising over the ratio of advertising variable we in no way compromised the statistical qualities of our models.

CHAPTER 4

EXPLORATORY ANALYSIS OF DATA AND INTERACTIVE MODEL BUILDING

Introduction

A researcher can use empirical data in two philosophically different ways. He can either employ the data to validate his theories or analyze the raw data to uncover the relationships which so far have gone undetected. In the latter approach, with the recently increased availability of statistical packages, it is possible to abdicate responsibility for the definition of the functional relationship between the data to the computer. Certain techniques, such as stepwise regression, forward selection, backward elimination, etc., allow one to structure a model mechanically. Once an à priori selected threshold point is reached the model building process is ceased and the model is completed. Such an approach, though convenient, has its shortcomings. The outlying points included in the regression may have an undue influence on the regression equation. Of course, the computer has no way to judge which of the data points should be included in the analysis and which should be omitted. The functional form of the equation, chosen because of the high value of R^2 or similar characteristic may not easily lend itself to interpretation. Some variables, important from the theoretical point of view, may be completely excluded from the equation because they failed to meet an arbitrary criterion.

It has been argued that the exploratory approach to data analysis and iterative model building may yield more meaningful results than automated regression techniques. In iterative model building, the theoretical knowledge of the scientist is combined with the computing power of the machine. At each stage of the analysis the displays and calculations done by the machine are closely examined by the analyst. The scientist draws on his knowledge about the field to which the data pertains and makes a judgement about the directions the analysis will take in the subsequent step. Researcher's knowledge about special events and unusual circumstances enables elimination of abnormal influences while regularities inherent in the data are captured. Special place in the exploratory analysis is given to the scrutiny of the displays (according to the adage "One picture is worth a thousand words").

The techniques used in the interactive model building include univariate data displays, data transformations, partial regression plots and outlier investigation. Once the model is built, an overall judgement of its quality takes place. Traditional measures of goodness of fit, such as R^2, t-statistics or F-statistics are looked at with some caution, since there has been a considerable amount of researcher input at each stage of the model construction. The ultimate decision about how well the model describes the reality belongs to the analyst. This is another departure from the classical model building procedure.

Brief Historical Note and Description of the Point of Departure in the Exploratory Analysis

Last year, building on the research initiated by Glen L. Urban, we found a series of equations describing the relationship between relative market share, order of market entry, consumer perceptions of products, advertising, and elapsed time between market entries.

We have concentrated our efforts on finding the best ways to introduce additional variables to the equation established by Glen Urban (I).

$$\text{(I)} \quad \ln \text{ share index} = -2.01 - \underset{(-1.59)}{.47 * \text{order}} + \underset{(5.58)}{2.56 * \text{relret}}$$

$$R**2 = 59\%$$

We have tried several forms of advertising and timing variables. The variable forms that proved to be most useful in our analysis (rendred equations with the best fit) were described in the preceding pages. We also tried other variables, e.g.:

I. Absoltue Advertising = Actual amount spent on advertising averaged over a three year period.

II. Ratio of Advertising $= \dfrac{\text{average expenditure of 2nd or 3rd product}}{\text{average expenditure of first product}}$

III. "lag-to-1*order" - this variable was used in an attempt to avoid collinearity between order of entry and "lag".

IV. "lag-between*order" - Same as above.

V. "proxy1" - some of the products were introduced many years after the category leader while other products were put on the market shortly after the first "in." The distinction between soon and late entry was drawn at the fifth year. Products introduced within five years of the leader were coded as zeros and products introduced later were coded as ones.

VI. "sqrt(lag to 1)" - nonlinear transformation of "lag to 1".

VII. "sqrt(lagbetween)" - nonlinear transformation of "lagbetween".

Generally speaking, contribution of the "lag" variables to the explanatory power of the models was not very big. It is not suprising since "lag" variables entered the models as the last ones. Of several lag variables we used, lag-to-1 and lag-between (and their derivatives, such as sqrt(lag between) or lag-to-1*order were the most useful, as the R**2 and t-statistics may attest.

We also tried to incorporate in the new, enlarged model some other concepts, eg. introductory advertising. Sparse data evidence convinced us to abandon these efforts.

As far as the form of the model is concerned we tested linear and exponential equations. Linear equations always gave a worse fit than the corresponding exponential equations. As a result, we decided to settle on the exponential form. Another consideration we had was the comparability of our models with the original study.

Of the large number (about 50) of models, the following were the best, judging by their R**2 and t-statistics.

$$\ln(\text{share index}) = -1.83 - \underset{(-1.21)}{.46} * \text{order} + \underset{(4.90)}{2.19} * \text{relret}$$

$$R^2 = 57.3\%$$

$$\ln(\text{share index}) = -2.89 + \underset{(5.20)}{1.75} * \text{relret} + \underset{(4.49)}{1.04} * \text{rel advert} +$$

$$\underset{(-2.36)}{-0.007} * \text{sqrt(lagbetween)} * \text{order}$$

$$R^2 = 79\%$$

$$\ln\ (\text{share index}) = -1.27 - .71 * \text{order} + 1.71 * \text{relret} + 1.01$$

$$(-2.69) \qquad (5.42) \qquad (4.69)$$

$$\text{rel advert} + -0.17 * \text{sqrt(lagbetween)}$$

$$(-1.80)$$

$$R^2 = 82.5\%$$

Description of Approach Taken in Exploratory Anlaysis

Rather than trying to extend certain predefined models, this time we went through unprejudiced, scrupulous analysis of the data. We wanted to see whether or not there exist any significant relationships between variables, what transformations best captured the detected relationships and which combination of variables model the interrelations in the most convincing fashion. In other words, we wanted to build, "from scratch", a new model which described the relations between relative market share and a set of possible explanatory variables - provided there existed such a model.

In the beginning, the only restriction we used was the considered data set. We employed in our analysis the set of variables used in the earlier studies - relret, order of entry, lag lagbetween, relative advertising, percentage advertising and relative market share (shindex). Since our earlier study involved considerable research on the best form of the variables describing advertising and lag effects, it made little sense to repeat it. We used those forms of the varialbes that had previously rendered the best results. As for relret, shindex and order variables, they incorporated the best information available, so it was only prudent to leave them unchanged.

Our aim was to use exploratory data analysis techniques and interactive model building to arrive at the best possible model without restricting ourselves a priori to any particular regression function or variable combination. The interpretability and comparability of the model would be brought to bear only in the final stages of model selection.

Analysis of Univariate Displays

We used histograms, stem-and-leaf displays,[1] normal plots, box plots, and summary statistics to explore the data. The small number of the data points did not allow the use of statistical tests to ascertain the types of data distributions. In the majority of the displays though, it could be noticed that the distribution is skewed to the right resembling a log-normal distribution. Summary statistics (max, min, average, and standard deviation) corroborated that observation (Exhibits 1-8), Skewed distributions usually suggests that a transformation on the data can bring about symmetry and attendant ease of interpretation. No distribution demonstrated multimodality or other abnormalities. Discrete distribution of the order of entry variable stemed from its nature (integer values of observations).

---

[1] Stem-and-leaf displays visually resemble histograms. In the simplest form of the stem-and-leaf display the first digit of a given number is used as a stem - classifying attribute (base of a histogram) and the less significant digits form the vertical part of a display. Stem-and-leaf displays combine the graphic quality of a histogram with the information inherent in numbers. Stem-and-leaf displays were first introduced by Tukey.

EXHIBIT 1. SHINDEX

No. of Observations - 23

1	0.04286
2	0.14286
3	0.42857
4	0.85714
5	0.56897
6	0.03448
7	1.88889
8	0.57692
9	2.58333
10	0.25000
11	3.14286
12	0.07692
13	1.41379
14	0.64516
15	0.41935
16	0.16667
17	0.22727
18	2.14280
19	0.16667
20	0.16667
21	0.31169
22	0.31579
23	1.63158

Maximum = 3.1429

Minimum = 0.034480

Average = 0.79136

St. Dev. = 0.89337

Middle of Interval	Number of Observations	
0.0	8	********
0.5	8	********
1.0	1	*
1.5	2	**
2.0	2	**
2.5	1	*
3.0	1	*

Stem-and-Leaf Display of Shindex
Leaf Digit Unit = 0.1000
1 2 Represents 1.2

7	+0*	0001111
11	+0T	2233
(4)	+0F	4455
8	+0S	6
7	+0.	8
6	1*	
6	1T	
6	1F	4
5	1S	6
4	1.	8
3	2*	1
2	2T	
2	2F	5
	HI	31

EXHIBIT 2. ORDER

No. of Observations - 23

1	2.
2	3.
3	2.
4	3.
5	2.
6	3.
7	2.
8	2.
9	2.
10	3.
11	2.
12	2.
13	2.
14	2.
15	3.
16	2.
17	3.
18	2.
19	2.
20	3.
21	2.
22	2.
23	3.

Maximum = 3.000

Minimum = 2.000

Average = 2.3478

St. Dev. = 0.48698

Middle of Interval	Number of Observations	
2.	15	***************
3.	8	********

Stem-and-Leaf Display of order
Leaf Digit Unit = 0.0100
1 2 Represents 0.12

(15)	20	000000000000000
8	21	
8	22	
8	23	
8	24	
8	25	
8	26	
8	27	
8	28	
8	29	
8	30	00000000

EXHIBIT 3. RELRET

No. of Observations - 23

1	0.23564
2	0.43594
3	0.88421
4	0.85474
5	0.75133
6	0.50847
7	1.67230
8	1.16667
9	1.54250
10	1.44610
11	1.37308
12	0.64103
13	0.98300
14	0.92474
15	1.10881
16	0.59475
17	0.76190
18	1.78280
19	0.44929
20	0.52417
21	0.50581
22	0.72398
23	0.97255

Maximum = 1.7828

Minimum = 0.23564

Average = 0.90625

St. Dev. = 0.42542

Middle of Interval	Number of Observations	
0.2	1	*
0.4	2	**
0.6	5	*****
0.8	5	*****
1.0	3	***
1.2	2	**
1.4	2	**
1.6	2	**
1.8	1	**

Stem-and-Leaf Display of Relret
Leaf Digit Unit = 0.1000
1 2 Represents 1.2

1	+0T	2
7	+0F	445555
11	+0S	6777
(5)	+0.	88999
7	1*	11
5	1T	3
4	1F	45
2	1S	67

EXHIBIT 4. PERCENT ADVERTISING

No. of Observations - 23

1	0.193580
2	0.266660
3	0.231370
4	0.395230
5	0.204010
6	0.006610
7	0.676750
8	0.302200
9	0.701250
10	0.034640
11	0.708400
12	0.006710
13	0.575410
14	0.344320
15	0.264050
16	0.197780
17	0.446440
18	0.101940
19	0.099690
20	0.339600
21	0.423610
22	0.010990
23	0.726990

Maximum = 0.72699

Minimum = 0.0066100

Average = 0.31558

St. Dev. = 0.23502

Middle of Interval	Number of Observations	
0.0	4	****
0.1	2	**
0.2	4	****
0.3	5	*****
0.4	3	***
0.5	0	
6.6	1	*
0.7	4	****

Stem-and-Leaf Display of Peradvn
Leaf Digit Unit = 0.0100
1 2 Represents 0.12

5	0	00139
8	1	099
(4)	2	0366
11	3	0349
7	4	24
5	5	7
4	6	7
3	7	002

EXHIBIT 5. RELATIVE ADVERTISING

No. of Observations - 23

1	0.58073
2	0.79998
3	0.69411
4	1.18576
5	0.61204
6	0.01984
7	1.35321
8	0.60440
9	2.10375
10	0.01732
11	1.41680
12	0.01342
13	1.15080
14	1.03295
15	0.79214
16	0.59335
17	1.33932
18	0.20389
19	0.29908
20	1.01879
21	0.84721
22	0.03298
23	1.45400

Maximum = 2.1037

Minimum = 0.013420

Average = 0.78982

St. Dev. = 0.55140

Middle of Interval	Number of Observations	
0.0	4	****
0.2	2	**
0.4	0	
0.6	5	*****
0.8	3	***
1.0	2	**
1.2	2	**
1.4	4	****
1.6	0	
1.8	0	
2.0	0	
2.2	1	

Stem-and-Leaf Display of Reladvn
Leaf Digit Unit = 0.1000
1 2 Represents 1.2

4	+0*	0000
6	+0T	22
8	+0F	55
(5)	+0S	66677
10	+0.	8
9	1*	0011
5	1T	33
3	1F	44
1	1S	
1	1.	
1	2*	1

EXHIBIT 6. LAG

No. of Observations - 23

1	9.
2	11.
3	16.
4	16.
5	1.
6	5.
7	2.
8	2.
9	10.
10	18.
11	1.
12	14.
13	1.
14	7.
15	15.
16	32.
17	44.
18	1.
19	1.
20	2.
21	4.
22	2.
23	14.

Maximum = 44.000

Minimum = 1.0000

Average = 9.9130

St. Dev. = 10.799

Middle of Interval	Number of Observations	
0.	9	*********
5.	3	***
10.	3	***
15.	5	*****
20.	1	*
25.	0	
30.	1	*
35.	0	
40.	0	
45.	1	*

Stem-and-Leaf Display of Lag
Leaf Digit Unit = 1.0000
1 2 Represents 12.

10	+0*	1111122224
(3)	+0.	579
10	1*	0144
6	1.	5688
2	2*	
2	2.	
2	3*	2
	HI	44

EXHIBIT 7. LAGBETWEEN

No. of Observations - 23

1	9.
2	2.
3	16.
4	0.
5	1.
6	4.
7	2.
8	2.
9	10.
10	8.
11	1.
12	14.
13	1.
14	7.
15	8.
16	32.
17	10.
18	1.
19	1.
20	1.
21	4.
22	2.
23	12.

Maximum = 32.000

Minimum = 0.000000000

Average = 6.4348

St. Dev. = 7.3226

Middle of Interval	Number of Observations	
0.	11	***********
5.	3	***
10.	6	******
15.	2	**
20.	0	
25.	0	
30.	1	*

Stem-and-Leaf Display of Lagbet
Leaf Digit Unit = 1.0000
1 2 Represents 12.

7	+0*	0111111
11	+0T	2222
(2)	+0F	44
10	+0S	7
9	+0.	889
6	1*	00
4	1T	2
3	1F	4
2	1S	6
	HI	32

EXHIBIT 8. LNSHINDEX

No. of Observations - 23

1	-3.14982
2	-1.94589
3	-0.84730
4	-0.15415
5	-0.56393
6	-3.36738
7	0.63599
8	-0.55005
9	0.94908
10	-1.38629
11	1.14513
12	-2.56499
13	0.34627
14	-0.43826
15	-0.86905
16	-1.79174
17	-1.48162
18	0.76211
19	-1.79174
20	-1.79174
21	-1.16575
22	-1.15268
23	0.48955

Maximum = 1.1451

Minimum = -3.3674

Average = -0.89931

St. Dev. = 1.2672

Middle of Interval	Number of Observations	
-3.5	1	*
-3.0	1	*
-2.5	1	*
-2.0	4	****
-1.5	2	**
-1.0	4	****
-0.5	3	***
-0.0	1	*
0.5	3	***
1.0	3	***

Stem-and-Leaf Display of LNSHINDEX
Leaf Digit Unit = 0.1000
1 2 Represents 1.2

2	-3*	31
3	-2.	5
3	-2*	
7	-1.	9777
11	-1*	4311
(4)	-0.	8855
8	-0*	41
6	+0*	34
4	+0.	679
1	1*	1

The box plot displays, used to identify outliers, indicated a few possible outliers in the lag, lagbetween and shindex variables. Normal plots of data did not follow straight lines (left and right tails were especially distinctive in shindex, relative advertising, percent advertising, lag and lagbetween). Shapes of normal plots confirmed that data transformations should be considered. However, before proceeding to search for the best transformations, we wanted to analyze bivariate distributions.

Bivariate Displays

We plotted all the possible explanatory variables against the shindex. One method of finding the best data transformation calls for iterating the formula

$$u = \log [(y_2 - y_1)/(y_4 - y_3)]$$

to zero. y_1, y_2, y_3, and y_4 are the counterparts of the equally spaced points x_1, x_2, x_3, and x_4. X's refer to the independent variables, y's to the dependent ones. Unfortunately, this formula, given a small number of observations in our data set, was not helpful.

The general appearance of the eye-fitted curves, penciled-in onto bivariate displays, led us to use a number of possible transformations, such as $y = e^x$, $y = x^2$, $y = ax$, $y = x^3$, $y = 1/x$. It was difficult to decide unequivocally which transformation was the best in each case for the two considered variables. Nonlinear relationships of the form $y = X^2$ or $y = X^3$ are probably better in the case of relret, relative advertising and percent advertising. The relationship between shrindex and lag or lagbetween are rather tenuous. With lag, the transformation $y = 1/x$ was probably the best. For lagbetween, possibly a similar transformation, y = 1/(x+1), could be used (we had to use x+1 becuase lagbetween sometimes takes on the value zero). Admittedly, the transformation 1/(x+1) was somewhat contrived and complicated. Thus, a linear relationship might be considered as a simpler alternative.

Correlations between Variables

To detect the strength and direction of the relationships between variables, we used correlation matrices. (Table 1 and 2).

Generally shindex shows the highest correlations with relret (between .760 and .818. Time-related variables (order, lag and lagbetween) have the lowest correlation values with shindex (from .232 to .404 - numbers are given as absolute values). Relative advertising and percent advertising have correlations with shindex ranging from .532 to .632 and from .615 to .728, respectively. The correlation coefficients vary when the form of shindex is changed. Although relative ranking of correlation values among variables remains roughly the same, ln (shindex) or $3\sqrt{\text{shindex}}$ have, on average, stronger correlations with the other variables than shindex.

Collinearity does not appear to be a major problem. Disregarding correlations between related variables, which never enter into equations simultaneously (eg. see correlations between relative advertising and percentage advertising), the highest correlation between prospective explanatory variables is between ln (lag) and ln (order) and amounts to -.466.

TABLE 1

CORRELATION MATRIX FOR ESTIMATION SAMPLE

	Shindex	Ln(Shin)	Sqrt(Shin)	Cube(Shin)	Relret	Order	Reladv	%Adv	Lag	Lagb	Inverlag	Invlagb	Sqrt(lag)
Relret	.760	.792	.799	.804									
Order	-.272	-.244	-.266	-.261	-.140								
Reladv	.605	.621	.629	.632	.300	.052							
%Adv	.704	.706	.728	.728	.384	-.018	.940						
Lag	-.296	-.236	-.290	-.279	-.140	.395	.123	-.017					
Lagbet	-.216	-.238	-.239	-.243	-.150	-.083	-.023	-.094	.683				
Inverlag	.405	.386	.407	.404	.231	-.451	-.079	.062	-.663	-.575			
Invlagb	.229	.307	.274	.288	.075	.051	.068	.090	-.432	-.688	.567		
Sqrt(Lag)	-.330	-.281	-.326	-.317	-.162	.443	.112	-.039	.965	.718	-.816	-.497	
Sqrt(Lagb)	-.232	-.277	-.264	-.272	-.126	-.052	-.022	-.088	.659	.953	-.647	-.866	.716

TABLE 2

CORRELATION MATRIX FOR ESTIMATION SAMPLE

	Ln(Shindex)	Ln(Relret)	Ln(Order)	Ln(Reladv)	Ln(%(Adv)	Ln(Lab)	Ln(Lagb)	Ln(Invlag)
Ln (Relret)	.818							
Ln(Order	-.244	-.083						
Ln(Reladv)	.532	.121	-.056					
Ln(%Adv)	.615	.249	-.031	.968				
Ln(Lag)	-.327	-.145	.466	-.027	-.028			
Ln(Lagb)	-.291	-.100	-.009	-.119	-.115	.746		
Ln(Invlag)	.327	.145	-.466	.027	.028	-1.0	-.746	
Ln(Invlabg)	.291	.100	.009	.119	.115	-.746	-1.0	.746

Simple Regression Equations

We ran a series of simple regressions between shindex, and explanatory variables. Table 3 gives a summary of the results in terms of R^2, regression coefficients, and t-statistics.

In general, the highest R^2 are for (in descending order) relret, percent advertising, relative advertising, 1/lag, order, and 1/(lagbetween+1). There is some improvement when lag is replaced by 1/lag in a simple regression (R^2 for linear goes from 8.7 to 16.4). Little improvement results from the use of 1/(lagbetween+1) instead of the lagbetween variable (for the linear form respective R^2 are: 4.7 and 5.3). Linear regression gives worse results than any other functional form in the case of relret (the lowest $R^2 = 57.8$ for linear, highest $R^2 = 66.9$ for multiplicative). Improvement of R^2 and t-statistics for other variables, when nonlinear form of regressions are tried, is visible, though less pronounced. In a few cases (percent advertising, relative advertising) some of the nonlinear forms gave worse results than the simple linear regression.

We limited our research to linear, exponential, multiplicative, square root, and cube root functional forms of the simple regression equations because the preceding data analysis did not show any evidence that some other, more complicated functional forms might be useful.

To see how stable coefficients of the regressions were, we examined corresponding resistant line equations[2]. In most cases, the

---

[2] The resistant line uses medians to fit lines to data. This line fitting technique "resists" outliers (i.e. limits the influence of outliers on the fit). For a detailbed discussion of resistant lines see Velleman & Hoaglin, A-B-C's of Explanatory Data Analysis [81].

TABLE 3

STATISTICS OF THE SIMPLE REGRESSIONS

Form of Regression Equation	Relret		Order		% Advertising		Relative Advertising		Lag		Labgetween		1/Lab		1/(Lagbetween+1)	
	R^2	t	R^2	t	R^2	t	R^2	t	R^2	t	R^2	t	R^2	t	R^2	t
Linear	57.8	5.36	7.4	-1.29	49.5	4.51	36.6	3.48	8.7	-1.42	4.7	-1.01	16.4	2.03	5.3	1.08
Exponential	62.7	5.94	6.0	-1.15	49.8	4.56	38.6	3.63	5.6	-1.12	14.9	-1.12	5.7	1.92	9.4	1.48
Multiplicative	66.9	6.52	6.0	-1.15	37.8	3.57	28.4	2.88	10.7	-1.58	8.5	1.4	10.7	1.58	8.5	-1.4
sqrt(shindex) = a + bx	63.9	6.09	7.1	-1.26	53.0	4.87	39.6	3.71	8.4	-1.39	5.7	-1.13	16.6	2.04	7.5	1.3
$3\sqrt{\text{shindex}}$ = a + bx	64.7	6.21	6.8	-1.24	53.0	4.87	40.0	3.74	7.8	-1.33	5.9	-1.15	16.3	2.02	8.3	1.38

coefficient was close, which serves to prove that the results are robust. Neither the various plots of residuals nor the resistant line coefficients indicated an existance of troublesome situations, such as heteroscedascity, autocorrelation, or flagrant outliers. However, with a limited number of data points, it is always difficult to draw firm conclusions.

Selection of Carriers to Multiple Regression Equations and Other Related Issues

To select the carriers for our multiple regressions, we inspected data plots, correlation coefficients, R^2 for simple regression equations, results of the stepwise regression on all variables, and took into account certain theoretical considerations.

Relret and either of two advertising variables have high correlation with shindex and produce good simple regression models. They are prime candidates for entry into the multiple regression models. Correlation between relret and relative advertising is about .3 for raw data and about -.2 for natural logs. Analogous numbers for percent advertising are .384 and .249. Therefore, we should not expect problems stemming from multicollinearity.

The remaining variables - order of entry, lag, and lagbetween along with their transformations 1/lag and 1/(lagbetween+1) showed in previous analysis a limited affinity for shindex. It was difficult to speculate a priori which of these variables should enter the models. It was obvious, that the incremental explanatory power of any of these variables would depend to a great extent on the degree of interdependency of these variables and the remaining variables used in a given model (eg. relret and advertising variables).

To gain some appreciation for the comparative importance of various explanatory variables we looked at several stepwise regressions. Stepwise regression employs certain predefined cutoff criteria and we wanted to find out what kind of model would be constructed in the stepwise process for each of the function classes contemplated. Stepwise regressions on a full set of variables included, in most cases, the variables relret, percent advertising, and 1/lag (lag for the multiplicative function). The highest $R^2 = 90.96$ was for the cube root functional form. Since R^2 for square root function was 90.22 we also checked stepwise results for the dependent variable (shindex) raised to the .25 and to .5 power. The respective R^2s were 90.60 and 89.54. We concluded that for the several tested functions from the family $y^a = x$, for a=1/3 the results were marginally better and further search would add little to our understanding of the relationship between shindex and explanatory variables. Thereafter we focused our attention on models of linear, multiplicative, exponential, square root and cube root forms.

Table 4 contains the summary of results stated in terms of R^2. for 97 different models. For sixteen models, marked in Table 4 by the starlet, we did exhaustive analysis of residuals. We selected these models for scrutiny because they had high R**2 relative to the other models of the same functional form and were likely candidates for the best model of the phenomena under exploration. Analysis of the residuals from these models did not show evidence of any abnormalities.

Generally, nonlinear models give a better fit but their advantage decreases as the number of explanatory variables included in the

TABLE 4

R^2 FOR MULTIPLE REGRESSION EQUATIONS

No.	Explanatory Variables	Form of Equation				
		Linear	Exponential	Multiplicative	sq. rt. (shindx) = f(...)	cube root (shindx) = f(...)
1	Relret, order	60.6	64.5	70.0	66.2	67.0
2	Relret, relat. adv.	73.4	78.9	86.0	80.5	81.5
3	Relret, % adv.	77.7	81.6	84.0	84.6	85.3
4	Relret, order, relat. adv.	77.4	81.9	88.4	84.4	85.2
5	Relret, relat. adv., 1/lag	82.6*	86.5*	-	89.3*	90.1*
6	relret, relat. adv., 1/(lagbetween +1)	75.8	84.1	-	84.3	85.9
7	Relret, order, % adv.	81.1	83.9	87.9	87.7	88.2
8	Relret, % adv., 1/lag	83.9*	86.5*	-	90.5*	91.0*
9	Relret, % adv., 1/(lagbetween + 1)	79.7	86.4	-	88.0	89.3
10	Relret, order, lag	62.5	65.2	72.2	68.0	68.4
11	Relret, order, lagbetween	62.0	66.3	74.6	68.1	68.9
12	Relret, order, relat. adv., 1/lag	83.3*	86.9*	-	90.0*	90.6
13	Relret, order, relat. adv., 1/(lagbetween +1)	80.5	87.7	-	88.7	90.0
14	Relret, order, relat. adv., lag	81.8	84.1	90.7*	88.2	88.6
15	Relret, order, relat. adv., lagbetween	79.5	84.2	91.2*	86.7	88.6
16	Relret, order, % adv., 1/lag	84.6*	86.9*	-	91.1*	91.5*
17	Relret, order, % adv., 1/(lagbetween +1)	83.5	89.2	-	91.5	92.5
18	Relret, order, % adv., lag	83.4	84.8	90.2*	89.7	89.9
19	Relret, order, % adv., labbetween	82.2	85.4	90.9*	89.1	89.7
20	Relret, relat. adv., lag	80.4	82.9	90.2	87.0	87.5
21	Relret, % adv., lag	82.1	83.7	89.5	88.6	88.9

equation grows. Of the nonlinear models, cube root and multiplicative equations tend to have the highest R^2s. Although the highest R^2 = 92.5 is for the equation cube root(shindex) = f(relret, order, percent advertising, 1/(lagbetween+1)), we opted for a multiplicative function as the one that best depicts the relationship between shindex and the other variables,because of its consistently good results and the relative simplicity of the functional form. The variable relret enters all the models. Relative advertising and percent advertising are used interchangeably in analogous equations. Although the contributions of the two advertising variables to the explanatory power of the equations are similar, we think that relative advertising, a variable that is normalized over the number of products in a given category, might be theoretically more compelling. Of the lag variables (lag, lagbetween, 1/lag, 1/(lagbetween+1)), 1/lag and 1/(lagbetween+1) generally render equations with the highest R^2. In the group of the multiplicative equations, owning to the nature of the function, only lag and labgetween variables are used (out of four forms of lag variables). It is difficult to decide which variable adds more to the understanding of the relationships.

Theoretical consideratons (proposition that the order of entry has an impact on the relative market share of the product) led us to introduce order of entry variable into several models. Contribution of this variable to the explanatory power of the tested equations is smaller than the contribution of the other variables. Such an outcome may be a consequence of the relatively high value of the correlation coefficient between the order and lag or lagbetween. It seems

TABLE 5

SUMMARY TABLE OF THE SELECTED MODELS - MULTIPLICATIVE EQUATIONS

No.	Intercept	Ln(Order)	Ln(Relret)	Ln(Relative Advertising)	Ln(Lag)	Ln(Lagbetween)	R^2	Adjusted R^2
1	0.41 (0.35)	-1.57 (-1.15)	-	-	-	-	6.0	1.5
2	0.48 (0.71)	-1.14 (-1.44)	2.24 (6.54)	-	-	-	70.0	67.1
3	0.64 (1.48)	-1.01 (-2.01)	1.90 (9.54)	0.35 (5.49)	-	-	88.4	86.6
4	0.50 (1.25)	-0.51 -0.97	1.85 10.03	0.36 (5.99)	-0.18 (-2.10)	-	90.7	88.6
5	1.02 (2.45)	-1.03 (-2.30)	1.86 (10.43	0.34 (5.85)	-	-0.24 (-2.41)	91.2	89.3

TABLE 6

SUMMARY TABLE OF THE SELECTED MODELS - EXPONENTIAL EQUATIONS

No.	Intercept	Order	Relret	Relative Advertising	1/lag	1/(lagbetween+1)	R^2	Adjusted R^2
1	0.59 (0.45)	-0.63 (-1.15)	-	-	-	-	6.0	1.5
2	-2.15 (-2.25	-0.35 (-1.01)	2.30 5.74	-	-	-	64.5	61.0
3	-2.32 (-3.31)	-0.46 (-1.80)	1.89 (6.13)	1.01 (4.28)	-	-	83.9	81.4
4	-3.19 (-4.56)	-0.19 (-0.75)	1.73 (6.25)	1.08 5.19	0.86 2.60	-	86.9	84.0
5	-2.54 -4.24	-0.50 -2.29	1.89 (7.02)	0.99 (4.93)	-	1.30 (2.91)	87.7	85.0

TABLE 7

SUMMARY TABLE OF THE SELECTED MODELS - LINEAR EQUATIONS

No.	Intercept	Order	Relret	Relative Advertising	1/lag	1/(lagbetween+1)	R^2	Adjusted R^2
1	1.96 (2.13)	-0.50 (-1.29)	-	-	-	-	7.4	3.0
2	0.12 (0.16)	-0.31 (-1.19)	1.54 (5.19)	-	-	-	60.6	56.6
3	0.00 (-0.00)	-0.39 (-1.92)	1.25 (5.21)	0.71 (3.82)	-	-	77.7	74.2
4	-0.66 (-1.18)	-0.18 (-0.90)	1.14 (5.18)	0.76 (4.58)	0.64 (2.46)	-	83.3	79.6
5	-0.11 -0.20	-0.40 (-2.08)	1.24 (5.30)	0.69 (3.90)	-	0.63 (1.60)	80.5	76.1

TABLE 8

SUMMARY TABLE OF THE SELECTED MODELS - SQUARE ROOT EQUATIONS ($\sqrt{shindex}$ = a + b*order + ...)

No.	Intercept	Order	Relret	Relative Advertising	1/lab	1/(lagbetween+1)	R^2	Adjusted R^2
1	1.36 (2.83)	-0.25 (-1.26)	-	-	-	-	7.1	2.7
2	(0.35) 1.02	-0.15 (-1.20)	0.85 (5.95)				66.2	62.9
3	0.29 (1.20)	0.19 (-2.18)	0.69 (6.61)	0.38 (4.70)	-	-	84.4	81.9
4	-0.05 (0.23)	-0.08 (-1.05)	0.63 (7.12)	0.41 (6.06)	0.33 (3.15)	-	90.0	87.7
5	0.21 (1.03)	-0.20 (-2.64)	0.68 (7.36)	0.37 (5.26)	-	0.41 (0.61)	88.7	86.2

TABLE 9

SUMMARY TABLE OF THE SELECTED MODELS - CUBE ROOT EQUATIONS ($\sqrt[3]{\text{shindex}}$ = a + b*order + ...)

No.	Intercept	Order	Relret	Relative Advertising	1/lag	1/(lagbetween+1)	R^2	Adjusted R^2
1	1.22 (3.57)	-0.18 -1.24	-	-	-	-	6.8	2.4
2	0.50 (2.05)	-0.10 (-1.17)	0.61 (6.04)	-	-	-	-	-
3	0.45 (2.71)	-0.13 (-2.17)	0.50 (6.85)	0.27 (4.83)	-	-	85.2	82.8
4	0.40 (2.63)	-0.08 (-1.37)	0.48 (7.22)	0.28 (5.66)	0.003 -2.39	-	88.6	86.1
5	0.40 (2.82)	-0.14 (-2.75)	0.49 (7.91)	0.26 (5.61)	-	(2.96)	90.0	87.8

superfluous to have two time-related variables in some models. Contribution of order of entry tends to be bigger when the number of variables in a model is smaller. Given that we are interested both in the order of entry effect and relative timing of entry, (as described by the lag and lagbetween variables) we may want our model to include two such time-related variables, which are least akin to one another. Of the order, lag and lagbetween variables, order and lagbetween seem to represent motions which are less related than the concepts captured by order and lag. Hence inclusion of order of entry and lagbetween entrants seems more logically justified.

With all the statistical evidence and theoretical considerations in mind we selected the following equation as the model of the studied phenomena:

$$\text{V.}\quad \ln(\text{shindex}) = \underset{(2.45)}{1.02} - \underset{(2.30)}{1.03}\ln(\text{order}) + \underset{(10.43)}{1.86}\ln(\text{relret})$$

$$+ \underset{(5.85)}{0.34}\ln(\text{relative advertising}) - \underset{(-2.41)}{0.24}\ln(\text{lagbetween})$$

$$R^2 = 91.2 \qquad R^2\text{adj} = 89.3$$

Two other models that are possible alternative descriptions of phenomena in question are:

$$\text{VI.}\quad \ln(\text{shindex}) = \underset{(1.25)}{0.50} - \underset{(-0.97)}{0.51}\ln(\text{order}) + \underset{(10.03)}{1.85}\ln(\text{relret})$$

$$+ \underset{(5.99)}{0.36}\ln(\text{relative advertising}) - \underset{(-2.10)}{0.18}\ln(\text{lag})$$

$$R^2 = 90.7 \qquad R^2\text{adj} = 88.6$$

$$\text{VII.}\quad \ln(\text{shindex}) = \underset{(-4.24)}{-2.54} - \underset{(2.29)}{0.50} * \text{order} + \underset{(7.02)}{1.84}\ \text{relret}$$

$$+ \underset{(4.93)}{0.99}\ (\text{relative advertising}) + \underset{(2.91)}{1.30}\ (1/\text{lagbetween}+1)$$

$$R^2 = 87.7 \qquad R^2\text{adj} = 85.0$$

Extraordinary Data Points

One of the purposes of the univariate and bivariate data displays is to aid in the identification of outliers. Once such data points are spotted, one can try to find out whether there is any particular reason for their existance. Sometimes outliers can help to detect new and unexpected aspects of the studied phenomenon. On other occasions the outlier is just an erractic data point that obscures the underlying regularity and distorts predictions. Such points can be removed form further analysis.

While looking at the bivariate dispalys of shindex and advertising variables, one can notice a point that can be classified as an outlier. The point in question has the following values: shindex = 2.14, relret = 1.78, order = 2, percent advertising = 0.11, relative advertising = 0.20, lag = 1, and lagbetween = 1 (Axion).

To evaluate the influence of that point we reestimated four of the five models after removal of Axion from the considered data set. Four models were reestimated so that we could test for robustness and stability. The summary of results is in Table 10. Exponential models include the following variables: relret, order, percent advertising 1/lag and (1/lagbetween+1). The multiplicative models use the lag or labgetween variable instead of 1/lag or (1/lagbetween+1).

TABLE 10

COMPARISON OF COEFFICIENTS FOR REGRESSIONS WITH A DIFFERENT # OF OBSERVATIONS

Model	No. of Observation	Order	Relret	Relative Advertising	Lag	Lagbetween	Intercept	R^2	Adjusted R^2
multiplicative	23	-1.03 (-2.30)	1.86 (10.43)	0.34 (5.85)	-	-0.24 -2.41	1.02 (2.45)	91.2	89.3
	22 without 1	-1.05 (-2.20)	1.84 (8.34)	0.34 (5.65)	-	-0.23 (2.30)	1.02 (2.37)	89.7	87.3
	22 without 10	-1.00 (-2.03)	1.87 (9.36)	0.33 (4.80)	-	-0.23 (-2.29)	0.99 (2.17)	91.2	89.1
	22 without 18	-1.02 (-2.18)	1.84 (9.45)	0.34 (5.63)	-	-0.23 (-2.23)	0.99 (2.22)	90.5	88.3
multiplicative	23	-0.51 (-0.97)	1.85 (10.03)	0.36 (5.99)	-0.18 (-2.10)	-	0.50 (1.25)	90.7	88.6
	22 without 1	-0.53 (-0.94)	1.83 (8.03)	0.36 (5.80)	-0.17 (-2.00)	-	0.52 1.21	89.1	86.5
	22 without 10	-0.48 -0.86	1.87 (8.99)	0.35 (4.86)	-0.17 (-1.99)	-	0.47 (1.11)	90.7	88.5
	22 without 18	-0.50 (-0.93)	1.83 (9.11)	-0.50 (-0.93)	-0.17 (-1.93)	-	0.48 1.15	89.9	87.6
exponential	23	-0.50 (-2.29)	1.84 (7.02)	0.99 (4.93)	-	1.30 (2.91)	-2.54 (-4.24)	87.7	85.0
	22 without 1	-0.58 (-2.76)	1.67 (6.28)	1.00 (5.32)	-	1.18 (2.78)	-2.13 -3.50	87.9	85.0
	22 without 10	-0.45 (-1.80)	1.91 (6.12)	0.94 (3.97)	-	1.25 (2.65)	-2.65 (-3.99)	87.8	84.9
	22 without 18	-0.49 (-2.19)	1.76 (5.49)	1.03 (4.52)	-	1.25 (2.66)	-2.53 (4.11)	86.8	83.7
exponential	23	-0.19 (-0.75)	1.73 (6.25)	1.08 (5.19)	0.86 (2.60)	-	-3.19 (-4.56)	86.9	84.0
	22 without 1	-0.30 (-1.22)	1.58 (5.62)	1.08 (5.48)	0.75 (2.37)	-	-2.71 (-3.73)	86.8	83.6
	22 without 10	-1.14 (-0.52)	1.82 (5.49)	1.01 (4.08)	0.82 (2.37)	-	-3.30 (-4.49)	87.0	83.9
	22 without 18	-0.19 (-0.75)	1.66 (5.01)	1.12 (4.80)	0.82 (2.35)	-	-3.15 -4.34	85.9	82.5

Removal of an outlying observation worsened the values of R^2. A logical explanation of this somewhat puzzling outcome is that although the data point is an outlier with respect to advertising values, it must otherwise contribute positively to the definition of the relationship between shindex and the rest of the varialbes, and removal of Axion can only diminish the precision of our estimates.

For the sake of completeness, we proceeded to compute the projection matrix $H = X\,(X'X)^{-1}X'$. According to Welsch and Hoaglin, replacement of y by y+1 in the regression changes the corresponding predicted value y to y+n. They suggest that in a regression with p carriers and n data points, leverage value (i-th diagonal element of H) greater than 2p/n deserves attention as a possible extreme. In our case critical value is:

$$\frac{2p}{n} = \frac{2 * 4}{23} = .3478$$

and for Axion

$$18\ (\text{Axion}) = .1712 < .3478$$

As we previously found, removal of Axion did not improve our estimates. The highest diagonal numbers of H matrix are for:

$$1(\text{Head \& Shoulders}) = .3607 > .3478$$

$$10(\text{Miracle White}) = .3478 = .3478$$

All the other diagonal elements are lower then the critical value of 2p/n. Table contains reestimated multiplicative and exponential models after observation no. 1 or observation no. 10 was removed from the data set. As previously noticed for Axion, removal of either

Head Shoulders or Miracle White from the data set, did not have a positive effect on the quality of the obtained equations. The conclusion of this question is that the outliers do not play a significant role in our data set and do not obscure the results of estimates.

Concluding Remarks on the Exploratory Analysis Phase of the Study

Findings of the careful and extensive analysis of the data and the robustness of results for estimates of a large number of models, led us to believe that there is a strong relationship between relative market share, positioning of the product, advertising, order of entry, and lagbetween market entries. We tentatively concluded that this relationship finds its best expression in a multiplicative model (V).

The contribution of the explnatory variables is uneven. The strongest relation exists between shindex and relret and shindex and advertising. Time-related variables seem to play less important roles in the explantion of relative market shares. We suppose it may be connected to the large range of values of the time-depenent variables. The appearance of some products on the market was separated by time lapses measured in tens of years while other products appeared separated by mere months. Market mechanisms operating in a short vs. a long period might be different and lumping together observations pertaining to these two groups might be errorneous. It could be useful, provided enough data is available, to reestimate the equations after dividing the data into classes with similar time differentials between introductions. The best models delivered quite a satisfactory fit, in the neighborhood of 0.9 for R^2. The final version of the model is simple, robust, and potentially helpful in a simulation of conditions under which the desired market share can be obtained.

CHAPTER 5

COMPARISON OF DATA IN TWO SAMPLES AND THE IMPLICATIONS FOR FURTHER ANALYSIS

We have used the same data sources to gather information for sets employed to estimate model equations and for the set utilized for predictive tests. One invariate we had in our study, was the composition of the first data set (let us call it the "estimation sample", for brevity). Although we supplemented the original information with data on advertising and timing of entry (lags), we never changed the products G. Urban had orginally used in his research. At the same time we were using the estimation sample in our quest for the best model, we also went through a long and laborious process of collecting new data for the "predictive" sample.

All products that we considered were low cost popular consumer goods. Product categories included household cleaners, coffees, detergents, feminine hygiene products, antacids, etc., etc. The first sample consisted of 15 categories, (38 products) and a second sample of 40 categories (50 products). All the categories were well-defined, vis. products in each category had a distinct, common characteristic and were clearly different from other products of similar type (eg. the coffee in our sample was freeze-dried, thus is different from regular ground coffee or instant coffee). Several more detailed requirements, necessitated by the ASSESSOR model, had to hold. The ratio of usage to purchase for the new brand had to be comparable to that of existing brands and consumer preferences had to reach equilibrium within a reasonably short time span.

TABLE 11

RESULTS OF F-TEST FOR ESTIMATION AND PREDICTIVE SAMPLE

Variable Name	All Brands $F_{0.05}(1,61) = 4.00$ $F_{0.01}(1,61) = 7.08$	3 Brands Only $F_{0.05}(1,41)$ 4.08 $F_{0.01}(1,41)$ 7.31
Shindex	0.24	1.31
Relret	0.84	0.24
Order	15.67	0.45
Relative Advertising	2.77	0.95
Lag	13.08	4.54
Lagbetween	.50	2.80

One obvious difference between estimation and prediction samples was that the estimation data contained at most 3 products in a given product category, while the prediction sample had as many as 10 products in one category (cat foods). This fact aroused our suspicion about direct comparability of the two data sets. We ran a series of analysis of variance (ANOVA) tests on corresponding data from the two samples. The ANOVA test compares means of two populations, assuming that both populations have normal distributions and the same variance. The first assumption is more important and fortunately, by and large, holds for our sets. We ran the ANOVA test on two variations a) we compared complete estimation and prediction samples b) we compared the estimation sample with the prediction sample from which brands 4 and further were removed. Results of the ANOVA test are summarized in Table 11.

As one notices, there are statistically significant differences between the estimation sample and full prediction sample. Both at 0.05 at 0.01 significance levels, we have to reject the null hypothesis that lag and order variables have equal means (respectively), in the two examined sets. Differences between data of the two samples indicate that the samples might not have come from the same general population. Hence, we might expect that predictive power of the model, tested over a somewhat different data set than the data set it was estimated over, might be diminished. When we inspect results of the ANOVA test for the abridged predictive sample, we see that only one of the variables, lag, is significantly different at the 0.05 level. At the 0.01 level we can not reject the null hypothesis about the equality of the lag variable in 2 samples. This serves to show that estimation and abridged prediction samples are very similar (presumably they came from the same population). We may surmise that predictions for the abridged sample should be better than for the complete sample. Still the lag variable may prove to be the cause of distortions in predictions. Upon investigation we learned the cause for having no more than three brands in each category in the original sample. Most of the categories were so recent that few alternative products were sold when the data was collected. In several cases, when there were more than three products on the market, the later entrants were either small local brands or generic products. Since the products considered for the analysis had to be national brands with distinctive brand identity, by assumption local brands and generic products were excluded.

If we limit the second sample to 3 brands only in each category, we obtain a comparable data set and we are fully justified in our inferences about market shares for the predicton sample. If we use the model to predict values over the set that transcends boundaries of the data domain previously established, we make certain assumptions that may or may not hold. Having that in mind we decided to pursue the following strategy.

(1) We will predict values of market share, using the model over the predicitve sample limited to 3 brands. Thus we will retain strict comparability of data and we will be fully justified in extrapolations.

(2) We will predict values of product share over the full new sample. We recognize it will be formal (model estimated on 3 brands is used to predict distant

brands) and a logical (we implicitly assume that the market mechanism stays the same) extension of the model. It's quite likely that results will be spotty, but even so, a ball park estimate is better than none.

(3) Upon completion of predictive testing we intend to make some value judgements about the model and its predictive power for market entrants of different order. After the evaluation, we will consider updating the model. We decided that such a sequence of steps is scientifically more defendable than mixing of the data from 2 samples and random redrawing of the estimation and prediction sample. The latter method assumes certain à priori knowledge, which we have actually gained à posteriori, and hence, redrawing might be intellectually dishonest.

CHAPTER 6

ANALYSIS OF THE PREDICTIVE POWER OF THE MODEL

Introduction

In the first phase of our research, we found the best model of the relationship between market share, perceptions of products, order of market entry, lagbetween market entries, and advertising. Now, we shall proceed to test the predictive power of our model.

Predictive tests can only be valid if one decides unequivoacally which model is to be tested. Previously we found that many models seemed to depict the relationship between analyzed phenomena almost equally well, if we were to judge on statistical characteristics alone. We chose to interpret our results as proof of the strength of the underlying relationships. From among several candidate equations we selected equation (V) as our model:

$$\text{(v)}\quad \ln(\text{shindex}) = \underset{(2.45)}{1.02} - \underset{(-2.30)}{1.03} * \ln(\text{order})$$

$$+ \underset{(10.43)}{1.86} * \ln(\text{relret}) + \underset{(5.85)}{.34} * \ln(\text{relative adv.})$$

$$- \underset{(-2.41)}{.24} * \ln(\text{lagbetween})$$

$$R^2 = 91.2 \qquad R^2\text{adj.} = 89.3$$

The selected model has several desirable characteristics.

(1) It is multiplicative. Multiplicative models, as the reader may compare in Table 4, proved no worse than equations of any other functional form. The multiplicative model has a formal advantage, over square root or similar models, of being simple and elegant.

(2) The variable lagbetwen is simplier and more convincing than its transformation 1/(lagbetween+1), which entered some of the models.

(3) All the variables that entered the selected model are statistically and/or theoretically more desirable than alternative variables of the same type (e.g. relative advertising is theoretically better than percent advertising).

(4) All the variables included in the model are significant at t = 2.101. In all subsequent discussions and tests we shall use model (v),

Goodness of Fit Measures

After selecting equation (V) as the best model, we turned our attention to testing the predictive capability of the other model.

When doing a predictive test, it is important to realize that there is no concise, definitive prediction methodology that can be used to test the predictive power of a considered model. If one examines a few models and finds they offer similar results, then some positive inferences about predictions can be made. Also, there exist measures such as Theil* coefficients, predicted R^{2}**, etc. that give the analyst some sense of the validity of the predictions. However, such measures have to be considered in reference to actual observations or they are meaningless.

---

* Theil coefficient is defined as abs$\left(\frac{y - \hat{y}}{y}\right)$ where y = actual market share, $\hat{y}$ = predicted market share.

**$R_p^2 = \frac{\Sigma(y - \hat{y})}{\Sigma(y - \bar{y})}$, y = actual market share, $\hat{y}$ = predicted market share,

$\bar{y}$ = average actual market share

In our evaluation of goodness of prediction we used certain traditional measures such as predicted R^2 and Theil coefficients. Since these measures proved to be of limited value, we went one step further and designed some new measures of goodness of prediction $\breve{\tilde{R}}$, $\overset{\square}{R}$, % turns. We tailored these measures to the particular needs of our research.

Predicted R^2 (later referred to as R_p^2) is a popular measure of goodness of prediction. In this analysis, the authors found that R_p^2 was insufficient and somewhat misleading. If we had relied on R_p^2 alone, we would have concluded that the model's predictive capability is at best meager. Such a conclusion would be inconsistent with the information inherent in the graphs of actual vs. predicted market share for our prediction data sets (see Appendices). The graphs show that the model is able to track very well downturns and upswings in market share. Not only does the model do an excellent job of tracking these changes in direction, it also yields predictions that are surprisingly close to actual market shares.

Since some of our variables are expressed in relation to the first entrant (recall the definitions of relret and shindex), the first entrants are excluded from regressions. As a result, R_p^2 by definition disregards all the information contained in the model's ability to track market share from the first to the second entrant. Consequently, R_p^2 does not give the model all the credit it deserves for its predictive power. In an effort to capture this lost information, the authors offer an alternative measure of predictive power, $\tilde{R}$. $\tilde{R}$ is defined in a manner very similar to R_p^2.

$$\tilde{R} = 1 - \frac{(y - \hat{y})^2}{(y - 1)^2} \quad \text{where } y = \text{actual market share}$$
$$\hat{y} = \text{predicted market share}$$

The advantage of $\tilde{R}$ is that it makes no assumptions about market share of brands trailing the first entrant as does R_p^2. That is to say, $\tilde{R}$ compares actual market share of later entrants with the market share of the first-in (see denominator), while R_p^2 compares these actual shares with an average market share of products in a given category. Use of an average pulls downward values obtained from R_p^2 as compared with $\tilde{R}$. R_p^2 takes for granted that we know market share of next-in entrants and hence steals away part of the credit due the model. The concept of comparing market share of later entrants with first entrant share = 1, may be thought of as a zero level prediction or null hypothesis. $\tilde{R}$ is a measure of the sensitivity of the model's predictions.

The lower order test of predictive capability of the model is %turns. %turns measures how many directional changes in market share were correctly tracked by the model (by directional change we mean relative decrease or increase of market share of the next product as compared with the market share of the immediately preceding product - relative upturn or downturn). Although, %turns offers no insights about absolute accuracy of predictions, it does convey an idea of how agile the model is. %turns may be thought of as a measure of model responsiveness.

One may consider predicted "turns" in a conceptual framework that lends itself to a statistical sign test. We applied the sign test to the number of "turns" correctly predicted (see Table 12). It is quite heartening, that in the light of the sign test, the predictions of turns are significant at 0.01 level.

SUMMARY TABLE 12. MEASURES OF GOODNESS OF PREDICTION

No.	Data	$\tilde{R}$	R^2 Fitted		R^2 Predicted	Mean Square Error	Sign* Test	% turns
			R^2	Adj. R^2				
1	Fitted estimation sample	97.4	91.2	89.3	-	.4144	Signif.	99
2	Fitted pooled samples, all brands	92.0	2.7	-4.0	-	1.0812	Signif.	-
3	Fitted prediction sample 3 brands	89.7	78.3	72.5	-	.5191	Signif.	96
4	Fitted pooled samples 3 brands	95.1	79.4	77.2	-	.5591	Signif.	92
5	Predicted prediction sample all brands	26.1	-	-	-1.10	1.8302	Signif.	76
6	Predicted prediction sample 5 brands	73.2	-	-	26.5	.9953	Signif.	-
7	Predicted prediction sample 3 brands	91.9	-	-	78.7	.6645	Signif.	90

*At 0.01 significance level.

The measures discussed above focus either on the model's ability to predict "turns" in market share or the model's ability to forecast the magnitude of market share. We propose yet another, spacial measure of goodness of predictions, $\overset{\square}{R}$. $\overset{\square}{R}$ combines evaluation of precision of the prediction of both turns and magnitudes. $\overset{\square}{R}$ take the area between the predicted and actual values represented in the graphs and compares it to a rectangular area, defined by the extremeties of the actual and predicted values. Ideally, the model would predict market share perfectly and the area of disparity, (space enclosed between actual and predicted market shares) would be zero. When this does not happen, the closer to zero the ratio of areas is, the better. Obviously, the lower boundary of $\overset{\square}{R}$ is ϕ. There exist an upper boundary for $\overset{\square}{R}$ and it is calculable. However, this boundary will change for each product category (it is a function of relative values of actual market shares). Computing upper boundaries would add little to our analysis, since the role we envisioned for $\overset{\square}{R}$ is to be a comparative measure across categories. In that capacity, $\overset{\square}{R}$ has a desirable characteristic. It is normalized over each category so cross-comparisons are valid (see $\overset{\square}{R}$ for categories 1 and 5 as examples of $\overset{\square}{R}$ values for good and poor predictions, respectively).

Table 13 contains a full set of $\overset{\square}{R}$ values for the selected models. Precision of prediction varies, but on average it is good. In several cases, values of $\overset{\square}{R}$ equal ϕ. Comparison of $\overset{\square}{R}$, for the same categories but different model equations, indicates which equation best predicted market shares within a given category. Comparison of $\overset{\square}{R}$ between different categories for the same model allows one to analyze a given model's strength of prediction for different types of products.

Table 12 contains a summary of measures of goodness of prediction. Alongside, where applicable, similar measures were computed for data over which the model has been fitted. The latter values can be used as points of reference in the evaluation process.

The value R_p^2 looks poor when computed for the prediction sample with all brands. It improves considerably once the prediction sample is limited to the first five brands. Values of R_p^2 goes up even higher when the prediction sample contains only the first 3 brands. In fact, R_p^2 of 78.7% comes rather close to the fitted model's R^2 = 91.2%.

A similar tendency is noticeable in mean square error values (MSE). Very high M.S.E. for predictions over the full set of brands gives way to much smaller errors as the number of brands considered decreases. M.S.E. for predicted values, 3 brands, is rather close to the M.S.E. for the original model.

The sign test is significant over all considered prediction data sets (3,5 and all brands) %turn hovers between 76% and 90% for the predictive data set, depending on the number of brands included.

Compared to R_p^2, $\tilde{R}$ gives understandably higher values. As we have argued before, $\tilde{R}$ is a better (fair) measure of goodness of prediction than R_p^2. $\tilde{R}$ is 26.1 for all brands, 73.2 for 5 brands and 91.9 for 3 brands. $\tilde{R}$ for the original model equals 97.4. These numbers serve to prove that model predictions are very good and reliable for the first three brands. They grow increasingly worse when we attempt to predict market share of more distant (later) brands. An identical conclusion can be drawn from the values of previously discussed measures of fit. Predictive power of the model deterioriates when we extrapolate predictions beyond early brands (vis. outside brand domain of the estimation

TABLE 13. SUMMARY TABLE-$\bar{R}$ FOR SELECTED MODELS.

Sample	Plot #	Category	Pooled Data 3 br. only Refitted	Old Data Fitted	4th-nth brands Fitted
Estimation Sample	1	Shampoo	.02	.01	-
	2	Liquid Detergent	.35	.11	-
	3	Prewash	.05	0.0	-
	4	Coffee	.25	.275	-
	5	Floor Wax	.50	.50	-
	6	Dry Bleach	.03	.12	-
	7	Fabric Softner	.23	.16	-
	8	Nonasprin	0.0	0.0	-
	9	All Purpose Cleaner	.07	.31	-
	10	Deoderant Soap	.42	.07	-
	11	Glass Cleaner	.18	.02	-
	12	Pre-soak	.15	.10	-
	13	Light Beer	.10	.03	-
	14	Dry Soup	.04	.05	-
	15	Cigareets	.19	.21	-
Prediction Sample	16	Dry Cat Food	.27	-	.12
	17	Air Freshner	.50	-	.16
	18	Antacide Tablets	.11	-	-
	19	Cough Remedy	.31	-	.24
	20	Tampons	.25	-	-
	21	Maxi/Mini Pads	.49	-	-
	22	Furniture Polish	.30	-	.15
	23	Steak Sauce	0.0	-	-
	24	Foot Powder	.21	-	.10
	25	Sleeping Aid	.28	-	-
		Average	0.212	0.131	.154
		Maximum	0.5	0.5	.24
		Minimum	0.0	0.0	.10

sample - as one recalls estimation of the model was done over a sample that included categories with up to three product brands).

If one combines the information conveyed by a series of graphs displaying actual vs. predicted market shares with the assessment of goodness of fit supplied by a variety of measures described above, the unavoidable impression one gets is that the predictive power of the tested model is good, especially for the first few brands. Our model is able to predict well both upturns and downturns in market shares of consecutive market entrants and the magnitude of their market shares. Given that the prediction and estimation samples were not entirely homogeneous, the strength of the predicitons is most encouraging and impressive.

CHAPTER 7

REFITTING THE MODEL

Exploratory analysis of the original sample uncovered a model of the relationship between market share, order of entry to the market, consumer perceptions of products, relative advertising and lagbetween market entries. The model (equation V) is a very good though not perfect, predictor. Upon testing, we concluded that the model depicts well the underlying market mechanism. However, one has to recognize that the coefficients of the regression are a function of the sample. Hence, for the purpose of gaining more generalized results, we chose to pool the two data samples and refit the enlarged sample.

In general, refitting the data over the pooled sample gave worse results. It is hardly surprising, since the combined sample contained more varied products than the original estimation sample. The differences between the two samples were both quantitative (these were easily detected by ANOVA test) and qualitative. Some of the more apparent differences between two samples include values of lag variables and age of products. The old estimation sample consisted of products whose lag factors were on average smaller than lag factors of the new (prediction) sample. The new sample consisted of products that were "ancient". Many of them (Sleep-eze, steak sauces, foot powders) were introduced at the turn of the century, and quite a few in the 20's or 30's. The products from the old (estimation) sample were much more "contemporary" (most of them were introduced in the 60's or 70's). Another issue is that of the surrogate buyer. In the new sample there are a few catagories in which the ultimate consumer may not select the product. Over the counter drugs (e.g. antacids, sleeping pills, cough remedies) constitute an important part of the new sample. These products are often chosen based on the advice of a physician or a pharmacist. In those instances, it is not really the consumer's preference but the medical practicioner's preference that is significant. Also, the category of dry cat food is included in the second sample. Here the pet owner makes the purchase. We do not know if the owner is making speculations about the cat's preferences or is driven by his own considerations (price etc.). In any event, all this poses certain questions that were nonexistent in the old sample.

Although once the samples are pooled we lose some statistical precision of estimation (compare R^2 and t statistics of the original and reestimated models), we gain increased diversity of data in the sample. As a result, we have a more realistic sample and most likely, a more realistic model. Such a model should be a better proxy for the real market environment. Furthermore, formally, a larger sample increases our confidence in the statistical findings.

The summary of the results of refitting are listed in Tables 16 through 23. The tables contain equations reestimated for all brands, pooled data; 3 brands, pooled data; 3 brands, new data only; 5 brands, new data only; all brands, new data only; 4^{th} thru n^{th} brands.

We reestimated the model over the new sample (with market entrants later than the third included) to see how well any previously found regularity held for the new data.

We refitted our model over the combined sample (for both 3 brands only and all brands) to build a new more general version.

Finally, we used the 4^{th} thru n^{th} brands to explore market mechanisms influencing late market entrants.

If we focus our attention on results from refitting the model over the new data, we quickly notice that equations constructed using all brands yield dismal results. None of the coefficients are significant and R^2 is basically null. The only conclusion one can draw is that the model does not hold. If we limit the number of brands included in the estimation to 5, results are suprisingly better. R^2 hovers between 50 and 60% and relret and relative advertising become significant. We get even better results for the new sample when we do estimation for the first 3 brands only. R^2 grows to 78%, relret and relative advertising are very significant. Order and lag variables are not significant (lagbetween even has the "wrong" sign). This serves to prove that the new sample indeed must be qualitatively different with respect to the time-related variables (as compared with the original estimation sample) (see Tables 14 through 18).

Reestimation of equation (V) using the pooled data was intended to generalize the model. Refitting using the sample which included all brands produced, as in the similar case described earlier, deplorable results. Refitting of the model on pooled data, 3 brands only gave equations whose R^2s ranged from 70-80%. Relret and relative advertising were strongly significant. Order and lagbetween were barely significant, but they did have the same sign as in the original model. The model (*)

TABLE 14

CORRELATION MATRIX ALL BRANDS, POOLED DATA

	Shindx	Ln(Shindx)	Relret	Order	Reladv	%Adv	Lag	Lagbet.	Invlag
Relret	.501	-.030							
Order	-.234	.008	.213						
Reladv	.600	.073	.180	-.041					
%Adv	.601	.099	.132	-.337	.662				
Lag	-.071	-.022	.139	.383	.035	-.136			
Lagbetween	.042	-.133	.004	-.276	-.045	-.041	.452		
Invlag	.166	.059	.032	-.372	-.035	.205	-.602	-.366	
Invlagbetween	.042	-.133	.004	-.276	-.045	-.041	.452	1.0	-.366

TABLE 15

CORRELATION MATRIX ALL BRANDS, POOLED DATA

	Ln(Shindx)	Ln(Relret)	Ln(Order)	Ln(Reladv)	Ln(%Adv)	Ln(Lag)
Ln(Relret	-.084					
Ln(order	.043	.188				
Ln(Reladv)	.063	.043	.007			
Ln(%Adv)	.148	-.032	-.260	.853		
Ln(Lag)	-.034	.110	.505	-.010	-.175	
Ln(Lagbetween)	-.118	-.010	-.254	.006	.046	.546

TABLE 16

REGRESSION EQUATIONS FOR NEW DATA ONLY, ALL BRANDS

No.	Intercept	Order	Relret	Relative Advertising	Lag	Lagbetween	R^2	Adjusted R^2
1	-1.00 (-1.94)	0.27 (0.64)	-	-	-	-	1.1	-1.5
2	-1.24 -2.08	0.36 (0.80)	-0.31 -0.81	-	-	-	2.8	-2.5
3	-1.22 (-2.03)	0.39 (0.87)	-0.32 -0.83	0.09 (0.80)	-	-	4.5	-3.5
4	-1.31 (-1.55)	0.36 (0.76)	-0.33 (-0.84)	0.10 (0.80)	0.04 (0.16)	-	4.6	-6.4
5	-0.93 (-0.87)	0.30 (0.56)	-0.30 (-0.76)	0.09 (0.79)	-	-0.09 (-0.32)	4.8	-6.1

TABLE 17

REGRESSION EQUATIONS FOR NEW DATA ONLY, 5 BRANDS

No.	Intercept	Order	Relret	Relative Advertising	Lag	Lagbetween	R^2	Adjusted R^2
1	-0.59 (-1.02)	0.18 (0.35)	-	-	-	-	0.4	-2.8
2	-0.15 (-0.33)	-0.07 (-0.17)	1.19 (4.54)	-	-	-	41.0	37.0
3	-0.14 (-0.36)	0.09 (0.29)	1.14 (5.31)	0.26 (3.98)	-	-	61.8	57.9
4	-0.16 (-0.34)	0.08 (0.23)	1.14 (5.01)	0.26 (3.90)	0.01 (0.09)	-	61.8	56.4
5	-0.04 (-0.07)	0.06 (0.17)	1.15 (5.21)	0.26 (3.92)	-	-0.03 (-0.25)	61.9	56.5

TABLE 18

REGRESSION EQUATIONS FOR NEW DATA ONLY, 3 BRANDS

No.	Intercept	Order	Relret	Relative Advertising	Lag	Lagbetween	R^2	Adjusted R^2
1	0.19 (-0.19)	-0.14 (-0.12)	-	-	-	-	0.1	-5.5
2	0.58 (0.78)	-0.77 (-0.95)	1.27 (4.32)	-	-	-	52.3	46.7
3	0.42 (0.81)	-0.39 (-0.69)	1.07 (5.09)	0.39 (4.37)	-	-	78.3	74.2
4	0.22 (0.35)	-0.46 (-0.77)	1.00 (3.98)	0.41 (4.30)	0.10 (0.64)	-	78.8	73.2
5	0.30 (0.32)	-0.33 (-0.46)	1.05 (4.39)	0.39 (4.23)	-	0.03 (0.16)	78.3	72.5

TABLE 19

REGRESSION EQUATIONS FOR POOLED DATA, ALL BRANDS

No.	Intercept	Order	Relret	Relative Advertising	Lag	Lagbetween	R^2	Adjusted R^2
1	-0.77 (-2.04)	0.12 (0.34)	-	-		-	0.2	-1.5
2	-0.85 (-2.15)	0.15 (0.47)	-0.20 (-0.73)	-	-	-	1.1	-2.2
3	-0.81 (-2.02)	0.46 (0.15)	-0.20 (-0.75)	0.04 (0.52)	-	-	1.5	-3.5
4	-0.75 (1.76)	0.24 (0.63)	-0.20 (-0.73)	0.04 (0.51)	-0.07 (0.47)	-	1.9	-4.9
5	-0.48 (-0.86)	0.08 (0.23)	-0.19 (-0.71)	0.04 (0.52)	-	-0.14 (-0.83)	2.7	-4.0

$$(*)\ \ln(\text{shindex}) = \underset{(0.69)}{0.36} - \underset{(-0.54)}{0.16}\ \ln(\text{order} + \underset{(5.74)}{1.03}\ \ln(\text{relret})$$

$$+ \underset{(5.62)}{0.30}\ \ln(\text{relative advertising}) - \underset{(-0.46)}{0.06}\ \ln(\text{lagbetween})$$

$$R^2 = 71.8\% \qquad R^2\text{adj} = 67.6\%$$

estimated on pooled data, 3 brands only (43 products) can be regarded as a modified version of the previously established model. We may also consider alternative equation (*1) as the model of the phenomena in question.

$$(*1)\ \ln(\text{shindex}) = \underset{(1.45)}{0.55} - \underset{(-1.71)}{0.74}\ \ln(\text{order}) + \underset{(9.02)}{1.46}\ \ln(\text{relret})$$

$$+ \underset{(6.27)}{0.38}\ \ln(\text{relative advertising})$$

$$R^2 = 78.8 \qquad R^2\text{adj} = 77.2$$

Equation (*1) has only one time-related variable, order, and as a result, the significance of that variable is increased as compared with model (*). (see Tables 19 and 20).

We previously discussed some of the possible reasons why equations estimated over the pooled data have worse statistical characteristics.

The last issue we tackled was an exploration of market phenomena for the 4th and later entrants. Consistantly inferior estimation results for samples including all brands led us to believe that the nature of the market mechanisms not only changes as the market becomes

TABLE 20

REGRESSION EQUATIONS FOR POOLED DATA 3 BRANDS

No.	Intercept	Order	Relret	Relative Advertising	Lag	Lagbetween	R^2	Adjusted R
1	-0.02 (-0.03)	-0.71 (-0.78)	-	-	-	-	1.5	-0.9
2	0.50 (0.94)	-0.95 (-1.57)	1.66 (7.26)	-	-	-	57.5	55.4
3	0.55 (1.45)	-0.74 (-1.71)	1.49 (9.02)	0.38 (6.27)	-	-	78.8	77.2
4	0.56 (1.47)	-0.63 (-1.33)	1.50 8.97	0.38 (6.21)	-0.05 (-0.64)	-	79.1	76.8
5	0.36 (0.69)	-0.16 (-0.54)	1.03 (5.74)	0.30 (5.62)	-	-0.06 (-0.46)	71.8	67.6

TABLE 21

CORRELATION MATRIX BRANDS 4 THRU n

	Shindx	Ln(Shindx)	Relret	Order	Reladv.	Lag	Lagbetween
Relret	-.002	-.089					
Order	-.450	.558	.346				
Relative Adv.	.725	.589	-.053	-.147			
Lag	-.304	-.372	-.321	-.053	-.045		
Lagbetween	-.035	-.038	-.198	-.220	-.066	.583	
Invlag	.214	.325	.457	-.103	.042	-.812	-.453

saturated with many new brands, but indeed the forces acting on later entrants may have opposite direction to those influencing earlier-in brands. (see Tables 21 through 23).

Statistical findings corroborated our deductions (see Table 23 for a summary of the estimations). First, the best model obtained was linear.

$$(*)\ \text{shindex} = \underset{(3.48)}{1.36} - \underset{(-2.78)}{0.16} * \text{order} + \underset{(0.56)}{0.09} * \text{relret} + 0.43 * \text{relative advertising} - \underset{(1.91)}{0.01} * \text{lag}$$

$$R^2 = 73.7 \qquad R^2\text{adj} = 66.7$$

Second, in all models estimated for 4^{th} and later entrants, the hierarchy of importance of explanatory variables has been altered as compared to the models for 3 brands only. Recall that in the 3 brands model, relret and relative advertising contributed most to the explanatory power of the model while order of entry and lagbetween were less significant. Now, the hierarchy of relative importance of explanatory variables is: for the linear and exponential model - 1) relative advertising 2) order 3) lag or 1/lag, respectively 4) relret; for the multiplicative model - 1) order 2) relative advertising 3) relret 4) lagbetween.

The multiplicative equation ln(Shindex) = f(ln(relret * relative advertising * order * lagbetween) exhibits this hierarchial change in the most pronounced manner. The authors believe that the following scenario offers a possible explanation for the change. Consider a consumer eager to purchase an effective dishwasher detergent that won't

TABLE 22

CORRELATION MATRIX BRANDS 4 THRU n

	Ln(Shindex)	Ln(Relret)	Ln(Order)	Ln(Reladv)	Ln(Lag)
Ln(Relret)	-.009				
Ln(Order)	-.568	.287			
Ln(Reladv)	.445	-.206	.040		
Ln(Lag)	-.368	-.415	.025	-.102	
Ln(Lagbetween)	.102	-.220	-.344	.023	.531

TABLE 23

REGRESSION EQUATIONS FOR 4^{th} AND LATER ENTRANTS

Model	No.	Intercept	Order or Ln(order)	Relret or Ln(relret)	Relative Advertising or Ln(relative advertising)	Ln(lag) or 1/lag	Ln(lagbetween) or 1/(lagbetween+1)	R^2	Adjusted R^2
Multiplicative	1	2.19 (2.19)	-1.78 -2.93	-	-	-	-	32.2	28.1
	2	2.45 (2.31)	-1.94 (-3.01)	0.29 (0.82)	-	-	-	34.8	27.1
	3	2.98 (3.50)	-2.12 (-4.14)	0.50 (1.77)	0.28 (3.35)	-	-	61.6	54.5
	4	3.97 (3.56)	-1.99 (-3.90)	0.30 (0.95)	0.25 (3.02)	-0.39 (-1.33)	-	65.7	56.5
	5	3.22 (3.28)	-2.20 (-4.02)	0.48 (1.65)	0.28 (3.27)	-	0.08 (-0.52)	62.3	52.3
Exponential	1	0.90 (1.54)	-0.31 -2.85	-	-	-	-	31.1	27.3
	2	0.83 (1.36)	-0.33 (-2.82)	0.17 (0.56)	-	-	-	32.4	24.4
	3	0.20 (0.37)	-0.29 (-3.03)	0.18 (0.70)	0.45 (3.20)	-	-	58.7	51.0
	4	-0.24 (-0.40)	-0.24 (-2.53)	-0.05 (-0.17)	0.44 (3.26)	9.26 (1.49)	-	64.0	54.4
	5	0.21 (0.41)	-0.30 (-2.96)	0.15 (0.57)	0.45 (3.15)	-	0.19 (0.47)	59.3	48.0
Linear	1	1.66 (3.48)	-0.19 (-2.14)	-	-	-	-	20.2	15.8
	2	1.58 (3.21)	-0.21 (-2.25)	0.20 (0.77)	-	-	-	22.9	13.8
	3	0.96 (2.69)	-0.17 (-2.68)	0.20 (1.16)	0.44 (4.66)	-	-	67.3	61.2
	4	1.30 (3.48)	-0.16 (2.78)	0.09 (0.56)	0.43 (4.92)	-0.01 (-1.91)	-	73.7	66.7
	5	1.02 (2.47)	-0.18 (-2.62)	0.19 (1.07)	0.44 (4.47)	-	-0.01 (-0.33)	67.5	58.9

leave spots on glasses. Conceivably, he will try a few brands, say up to three, and perhaps develop a first preference and identify a second preference to be purchased when his first choice is not in stock, when the second choice is on sale, etc. In that light, the model seems to tell us that if a brand is not among those first few market entrants, the order of its entrance is critical. Order is crucial because consumers will have little or no incentive to risk trial of a new product after already having identified at least two products thatwill leave glasses spotless or at some reasonable level of cleanliness.

The most important facts we discovered during a detailed analysis of the equations can be summarized as follows:

I. The model we have constructed, through exploratory analysis of the original data, correctly captures forces operating in the consumer products market. When we reestimated this model over a richer, more diverse sample, coefficients change somewhat and statistical qualities of the model decline, but we obtain a more genral, realistic model of consumer product markets.

II. The reestimated model is valid for the first 3 brands. Equations deteriorate rapidly if we add more brands to the estimation sample. For 4^{th} and later entrants we established a new, linear model that shows a dramatic shift in relative importance of explanatory variables. The market mechanism is clearly different for earlier and later brands. We suggest that the two models be used in conjunction to analyze market share behavior for the full spectrum of brands.

Depending on the situation, a manufacturer can use results generated by the "first" or "second" model to help him understand his specific market position.

CHAPTER 8

MANAGERIAL IMPLICATIONS

After developing a model that relates market share, product positioning, advertising, and market entry timing, the next logical step was to explore its managerial implications.

We define managerial implications to encompass questions related to the concepts explicitly included in the model. Certain managerial issues are strategic in scope and are dealt with only at the corporate level. These are questions of research and development outlays, pricing, timing of new product introduction, and advertising expenditures. Other issues (e.g. couponing) are more detailed in nature and likely to be the responsibility of a brand manager. At both levels of decision making, it is important to understand the role of factors included in our model in determining potential market share. The specific decisions may be fined-tune and allocation of resources optimized when the market situation of a given company (product) is well comprehended thanks to the established model.

Based on our study the following hierarchy of significance of the variables can be established for the first few (three) brands:

(1) product positiioing - relret

(2) advertising

(3) order of market entry

(4) entry lag

This hierarchy should give manufacturers some guidelines for assessing the trade-offs involved when the market is "new". For instance, companies must decide on the trade-off point between the time a product remains in research and development stage and the advertising costs. If the R&D phase is lengthy, chances are that the developed product will be superior. However, it will also be late entering the market. Given the importance of relret to the market share of the first few brands, high quality translates into big market rewards. A high quality product combined with vigorous advertising will outweigh by far the negative effects of belated entry.

Nevertheless, if a company takes too long to introduce its product, competitors may proliferate. With several similar products already on the market, a company faces the world as described by the second, linear model. For late entrants, the hierarchy of variable significance changes to:

(1) advertising

(2) order of entry

(3) relret

(4) entry lag

Advertising continues to be important. Quality of products becomes less relevant while penalty for late entry increases rapidly. Very late entrants will have to make an extraordinary advertising effort to make a dent in the market.

Since well-timed market entry precipitates substantial financial rewards, market intelligence on new product development becomes a very important consideration for management. If several companies are developing similar new products parallely, information about competitor's progress would help determine the correct tempo for R&D work and the optimal timing for product release. Similarly, if a company discovers that competitor is developing an entirely new product, it may be able to develop an analogous product and possibly preempt the "originator". Intelligence is crucial if a "weak" company (company with limited financial and other resources), is about to launch a product that is likely to be in competition with a "strong" company's product. The order of entry effect might save a "weak" company a substantial amount of money in advertising and R&D outlays.

The "wait and see" tactic may be a viable alternative for a resourceful company. Such a compnay can afford to spend substantial amounts of money on advertising and research and development if the need exists. Therefore, it can afford a late market entry. In fact, the resourceful company can think of the other company's product as being a guinea pig. In effect, its strategy will be to let the pioneer make all the mistakes and then introduce a product free of the shortcomings of the first-in brand. However, the late entrant must remember that it is the first entrant that sets the standard and gives consumers a point of reference. If the first-in product is high quality, the followers may have difficulty developing better products since marginal improvement becomes more difficult and marginal benefit to consumers less obvious. Furthermore, when the new product falls in the category of "cheap" or low involvement products, the later entrant may expect problems stemming from consumer laziness. It may not be worthwhile, from the consumer's point of view, to retain information about all the additional brands that trail the market pioneer. The company's advertising campaign, and to some extent R&D, has to be mindful of that possibility while designing and promoting the next-in product.

The later entrant is apt to require a very elaborate campaign incorporating advertising, couponing, varying pricing, sweepstakes and other techniques, to assure sufficient impact on consumers. Pricing strategy will be an important element of the campaign. In order to induce trial purchase of the new product, a low initial pricing strategy is often used. Such a strategy helps the first-in gain the highest possible trial rate and hook the price-conscious consumers. A low initial pricing strategy is even more important for later entrants. People who had not tried the first entrant's product might find the reduced price of the next-in product sufficiently tempting to elicit trial. Those who had tried the pioneering brand might also be cajoled to switch if the pioneer's campaign has failed to foster strong brand loyalty.

To conclude, our model allows the company to analyze the environment and improve its decision making process by simulating consequences of different marketing/manufacturing strategies.

CHAPTER 9

FUTURE RESEARCH

Because of limitations imposed on us by small sample sizes and time constraints, we were unable to pursue certain interesting issues that surfaced during our study. Below we mention a few of the topics that may be fruitfully explored in the future.

Just as we have divided products into two groups (e.g. 3rd and earlier entrants, and 4th and later entrants), almost every variable may be used to segment the sample in order to see whether or not models estimated over different subsamples vary. Substantial increases in sample size will be necessary before extensive segmentation can be undertaken. Otherwise the results of modelling may be spurious.

Advertising is a prime candidate for segmentation. Our sample includes some categories where the average advertising level exceeds $20 million per year (beer) while other categories spend less than $3.5 million per year (foot powders). It would be interesting to estimate models over the sample containing highly advertised products only and over a sample consisting of minimaly advertised products. To repeat an earlier remark, one might try to explore a model employing the ratio of advertising variable. Using ratio of advertising, keeps the definition of all variables in the model consistent.

Further analysis of the timing of product entry to the market, as expressed by lag and lagbetween variables, might prove enlightening. Products which have been around since time immemorial might well be subject to different market laws than are the relative newcomers.

Then of course, there is an issue of whether or not the four or so particular variables we have chosen explain all aspects of market phenomena. We may safely assume that other concepts such as absolute (or relative) prices of products, promotions, distribution networks or size and strength of the manufacturer might be usefully introduced to an enlarged model.

In particular we suggest research that considers the concepts of brand loyalty. Insufficient data restricted us from examining brand loyalty vs. market share in a statistical manner. Still, scattered evidence seems to indicate that such a relationship might be possible to quantify.

Another variable that deserves attention is introductory advertising. In our data set we did not notice patterns in advertising expenditures that would imply existence of introductory advertising. This observation runs contrary to well-publicized advertising strategies and further research would be valuable.

There may be some other interesting variables that will enhance the explanatory power of the model. However, as often done in the physical sciences, we may elect to operate at a different level of abstraction, trading off richness of detail and situation specific precision vs. complexity.

One important formal stumbling block we encountered in our study was lack of adequate measures of goodness of prediction. We tried to remedy the situation by proposing several measures tailored to our specific needs. It would be most helpful if future statistical studies further developed the theory of measures of goodness of prediction.

CHAPTER 10

CLOSING REMARKS

In closing, we would like to repeat that we believe our two-part model is a very good descriptor of market mechanisms. As such, it can be very useful for further theoretical investigation of market behavior and market strategy. What perhaps is even more important (if importance is measured in dollars) is that the model offers an excellent managerial tool which along with other techniques can help develop effective market entry strategies for new consumer products.

APPENDICES

APPENDIX A

DATA

Category Number	Product Number	Name	Shindex	Relret	Order	3-year Average Advertising	Relative Advertising	Percent Advertising	Year of Introduction	Lag	Lagbetween
1		Dandruff shampooes									
	1	Head & Shoulders	1	1	1	12,398.9	-	-	62	-	-
	2	Tegrin	0.043	0.236	2	4,446.6	0.581	0.194	71	9	9
	3	Selsun Blue	0.143	0.436	3	6,125.4	0.800	0.267	73	11	2
2		Liquid detergent									
	4	Wisk	1	1	1	10,562.3	-	-	56	-	-
	5	Dynamo	0.429	0.884	2	6,545.1	0.694	0.231	72	16	16
	6	Era	0.857	0.855	3	11,181.2	1.186	0.395	73	16	0
3		Prewash									
	7	Spring & Wash	1	1	1	7,747.8	-	-	68	-	-
	8	Shout	0.569	0.751	2	2,002.4	0.612	0.204	69	1	1
	9	Miracle White	0.035	0.508	3	64.9	0.020	0.007	73	5	4
4		Coffee									
	10	Maxim	1	1	1	4,599.5	-	-	64	-	-
	11	Tasters Choice	1.889	1.672	2	9,623.2	1.353	0.677	66	2	2
5		Floor Wax									
	12	Mop & Glow	1	1	1	2,812.4	-	-	71	-	-
	13	Step Saver	0.577	1.167	2	1,219.0	0.604	0.302	73	2	2
6		Dry Bleach									
	14	Snowy	1	1	1	2,729.6	-	-	60	-	-
	15	Clorox 2	2.583	1.543	2	7,247.5	2.104	0.701	70	10	10
	16	Miracle White	0.250	1.446	3	358.0	0.017	0.035	78	18	8
7		Fabric Softener									
	17	Cling free	1	1	1	5,716.8	-	-	71	-	-
	18	Bounce	3.143	1.373	2	13,888.4	1.417	0.708	72	1	1

Category Number	Product Number	Name	Shindex	Relret	Order	3-year Average Advertising	Relative Advertising	Percent Advertising	Year of Introduction	Lag	Lagbetween
8		Non-aspririn									
	19	Tylenol	1	1	1	15,355.1	-	-	61	-	-
	20	Datril	0.077	0.641	2	103.7	0.013	0.007	75	14	14
9		All purpose cleaner									
	21	F-409	1	1	1	3,658.0	-	-	67	-	-
	22	Fantastic	1.414	0.983	2	4,957.4	0.151	0.575	68	1	1
10		Deodorant soap									
	23	Dial	1	1	1	10,351.5	-	-	50	-	-
	24	Zest	0.645	0.925	2	9,100.7	0.033	0.344	57	7	7
	25	Safeguard	0.419	1.109	3	9,979.1	0.792	0.264	65	15	8
11		Glass Cleaners									
	26	Windex	1	1	1	4,327.6	-	-	33	-	-
	27	Ajax	0.167	0.595	2	2,405.8	0.593	0.198	65	32	32
	28	Glass Plus	0.227	0.762	3	5,430.4	1.339	0.446	75	42	10
12		Presoak									
	29	Biz	1	1	1	4,535.0	-	-	67	-	-
	30	Axion	2.143	1.783	2	514.8	0.204	0.102	68	1	1
13		Light beer									
	31	Lite	1	1	1	27,845.9	-	-	74	-	-
	32	Schlitz	1.667	0.449	2	4,951.0	0.299	0.099	75	1	1
	33	A/B	1.667	0.524	3	16,865.0	1.019	0.340	76	2	1
14		Dry soup									
	34	Cup-a-soup	1	1	1	3,337.2	-	-	71	-	-
	35	Souptime	0.312	0.506	2	2,452.6	0.847	0.424	75	4	4
15		Cigarettes									
	36	True	1	1	1	8,020.3	-	-	66	-	-
	37	Doral	0.316	0.724	2	336.5	0.033	0.011	68	2	2
	38	Vantage	1.632	0.973	3	22,253.8	1.454	0.727	70	14	12

Category Number	Product Number	Name	Year of Introduction	Lag	Lagbetween	3-brands Relative Advertising	5-brands Relative Advertising
16		Dry cat food					
	39	Little Friskies	60	-	-	-	-
	40	Purina cat chow	61	1	1	1.786	1.2050
	41	9-lives	67	7	6	1.000	1.1240
	42	Purina sp. dinners	68	8	1		.070
	43	Meow-mix	74	14	6	-	1.557
	44	Chef's blend	78	18	4	-	-
	45	Country blend	78	18	4	-	-
	46	Good mews	79	19	1	-	-
	47	Ocean blend	80	20	1	-	-
	48	Crave	82	22	2	-	-
17		Air freshener					
	49	Wizard	60	-	-	-	-
	50	Airwick solid	70	10	10	1.932	1.9890
	51	Renuzit	72	12	2	0.039	.0410
	52	Glade Solid	74	14	2	-	.0250
	53	Twice as fresh	78	18	4	-	1.8850
18		Antacid tablets					
	54	Tums	29	-	-	-	-
	55	Digel	64	35	35	0.392	.5200
	56	Rolaids	72	43	8	1.562	2.0720
	57	Alka-2	75	46	3	-	0.0240

Category Number	Product Number	Name	Shindex	Relret	Order	3-year Average Advertising	Relative Advertising	Percent Advertising
16		Dry cat food						
	39	Little friskies	1	1	1	964.7	-	-
	40	Purina cat chow	1.333	1.385	2	8,064.4	2.798	0.274
	41	9-lives	0.333	0.692	3	4,515.9	1.533	0.153
	42	Purina sp. dinners	0.733	1.980	4	281.7	0.956	0.010
	43	Meow-mix	1.467	1.165	5	6,254.0	2.123	0.212
	44	Chef's blend	0.467	1.200	6	4,540.6	1.542	0.154
	45	Country blend	0.133	1.440	6	162.5	0.055	0.006
	46	Good mews	0.200	0.720	7	3,247.0	1.102	0.110
	47	Ocean blend	0.200	2.700	8	254.8	0.090	0.009
	48	Crave	0.267	1.200	9	1,167.3	0.396	0.040
17		Air freshener						
	49	Wizard	1	1	1	2,638.3	-	-
	50	Airwick solid	1.286	0.742	2	4,958.0	1.989	0.398
	51	Renuzit	0.429	0.925	3	101.8	0.041	0.008
	52	Glade solid	0.857	1.004	4	67.5	0.025	0.005
	53	Twice as fresh	1.190	1.356	5	4,697.0	1.885	0.377
18		Antacid tablets						
	54	Tums	1	1	1	6,027.4	-	-
	55	Digel	0.471	1.357	2	2,262.0	0.520	0.130
	56	Rolaids	2.411	2.128	3	9,006.9	2.072	0.518
	57	Alka-2	0.294	1.003	4	106.1	0.024	0.006

Category Number	Product Number	Name	Shindex	Relret	Order	3-year Average Advertising	Relative Advertising	Percent Advertising
19		Cough remedies						
	58	Corricidin	1	1	1	0.0	-	-
	59	Dristan	0.787	0.532	2	17,851.9	2.517	0.420
	60	Contac	2.134	0.704	3	14,249.1	2.009	0.335
	61	Co-tylenol	0.740	0.864	4	2,313.8	0.324	0.054
	62	Nyquil	1.528	0.577	4	6,589.9	0.929	0.155
	63	Daycare	0.370	0.624	5	1,549.9	0.216	0.036
20		Tampons						
	64	Tampax	1	1	1	6,279.6	-	-
	65	Kotex	0.603	0.697	2	1,198.6	0.248	0.062
	66	Playtax	0.994	1.025	3	5,489.5	1.128	0.282
	67	OB	0.276	0.378	4	6,477.7	1.332	0.333
21		Pads						
	68	Stayfree	1	1	1	12,828.6	-	-
	69	Kotex	0.211	0.238	2	4,088.8	0.537	0.179
	70	New Freedom	0.255	0.444	3	5,988.9	0.783	0.261
22		Furniture Polish						
	71	Old English	1	1	1	1,671.0	-	-
	72	Pledge	4.543	1.558	2	5,152.5	2.382	0.397
	73	Endust	0.526	0.607	3	1,821.7	0.840	0.140
	74	Favor	0.440	0.820	4	299.0	0.138	0.023
	75	Behold	0.793	0.653	5	2,013.8	0.930	0.155
	76	Scott's Liquid Gold	0.414	0.521	6	2,032.8	0.936	0.156

Category Number	Product Number	Name	Year of Introduction	Lag	Lagbetween	3-brands Relative Advertising	5-brands Relative Advertising
19		Cough remedies					
	58	Corricidin	49	-	-	-	-
	59	Dristan	57	8	8	1.668	-
	60	Contac	61	12	4	1.332	-
	61	Co-tylenol	69	20	6	-	-
	62	Nyquil	69	20	6	-	-
	63	Daycare	78	29	9	-	-
20		Tampons					
	64	Tampax	36	-	-	-	-
	65	Kotex	65	29	29	0.277	.2480
	66	Playtax	69	33	4	1.270	1.1280
	67	OB	76	40	7	-	1.3320
21		Pads					
	68	Stayfree	72	-	-	-	-
	69	Kotex	75	3	3	0.537	.5370
	70	New Freedom	78	6	3	0.783	.7830
22		Furniture Polish					
	71	Old English	46	-	-	-	-
	72	Pledge	58	12	12	1.788	2.3510
	73	Endust	62	16	4	0.632	.8310
	74	Favor	67	21	5	-	.1360
	75	Behold	68	22	1	-	.9190
	76	Scott's Liquid Gold	69	23	1	-	-

Category Number	Product Number	Name	Shindex	Relret	Order	3-year Average Advertising	Relative Advertising	Percent Advertising
23		Steak sauces						
	77	A-1	1	1	1	5,115.7	-	-
	78	Heinz	0.220	0.387	2	2,890.6	0.722	0.361
24		Foot powders						
	79	Dr Schols	1	1	1	569.5	-	-
	80	Absorbine Jr.	0.190	0.531	2	6.5	0.006	0.001
	81	NP-27	0.190	1.061	3	7.1	0.006	0.001
	82	Desenex	2.524	1.295	4	3,368.8	3.612	0.602
	83	Tinactin	0.333	1.083	5	128.8	0.138	0.023
	84	Aftate	0.238	0.844	6	1,518.2	1.626	0.271
25		Sleeping aids						
	85	Sleep-eze	1	1	1	1,318.1	-	-
	86	Nytol	1.489	1.068	2	1,977.4	0.720	0.180
	87	Sominex	1.922	1.366	2	4,043.1	1.473	0.363
	88	Excedrin PM	2.511	2.120	3	3,641.2	1.327	0.332

Category Number	Product Number	Name	Year of Introduction	Lag	Lagbetween	3-brands Relative Advertising	5-brands Relative Advertising
23		Steak sauces					
	77	A-1	07	-	-	-	-
	78	Heinz	13	6	6	0.722	.7220
24		Foot powders					
	79	Dr Schols	20	-	-	-	-
	80	Absorbine Jr.	37	17	17	0.033	.0080
	81	NP-27	40	20	3	0.037	.0090
	82	Desenex	45	25	5	-	4.127
	83	Tinactin	68	48	23	-	.1580
	84	Aftate	76	56	8	-	-
25		Sleeping aids					
	85	Sleep-eze	31	-	-	-	-
	86	Nytol	55	24	24	0.720	.7200
	87	Sominex	55	24	24	1.473	1.4730
	88	Excedrin PM	69	38	14	1.327	1.3270

APPENDIX B

PLOTS OF FITTED VS. ACTUAL MARKET SHARES -

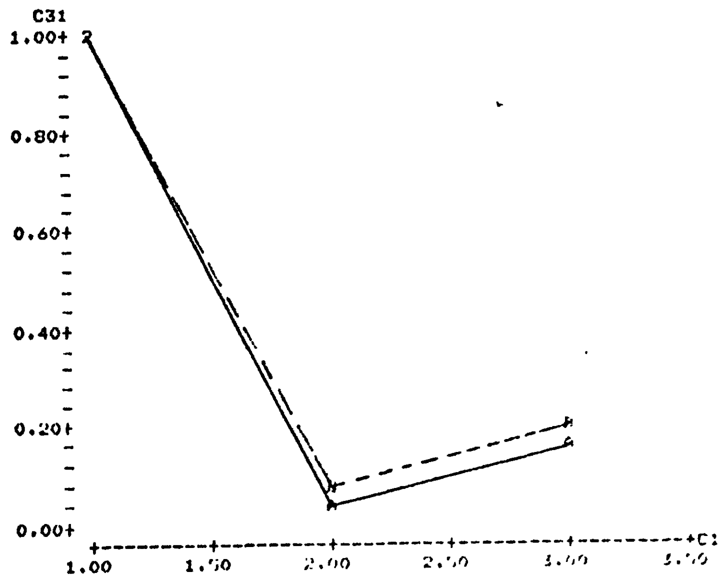
REFITTED POOLED DATA SAMPLE

3-BRAND MAXIMUM -

- - - = Predicted Market Share

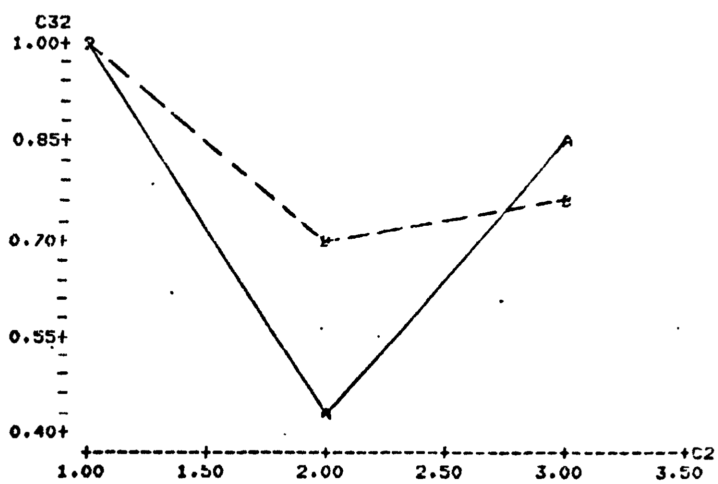
_____ = Actual Market Share

Shindex



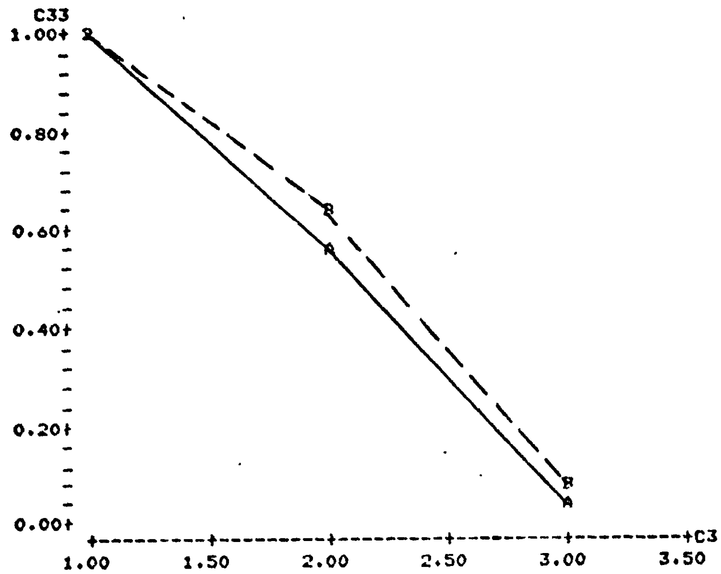
Head & Shoulders
Tegrin
Selson Blue

mplot c32 c2,c92 c2



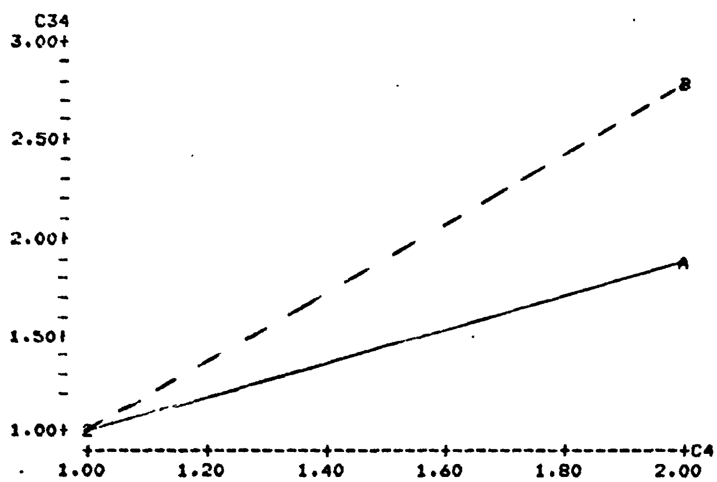
Wisk
Dynamo
Era

Shindex



Spray & Wash
Shout
Miracle White

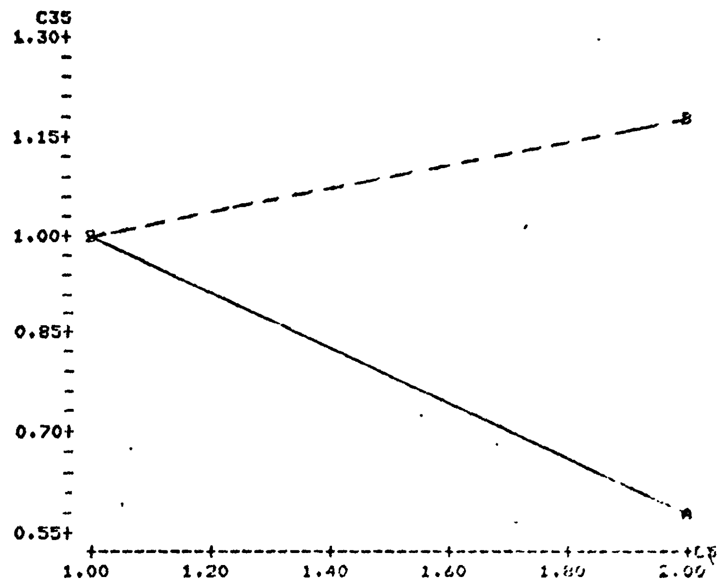
mplot c34 c4,c94 c4



Maxim
Tasters Choice

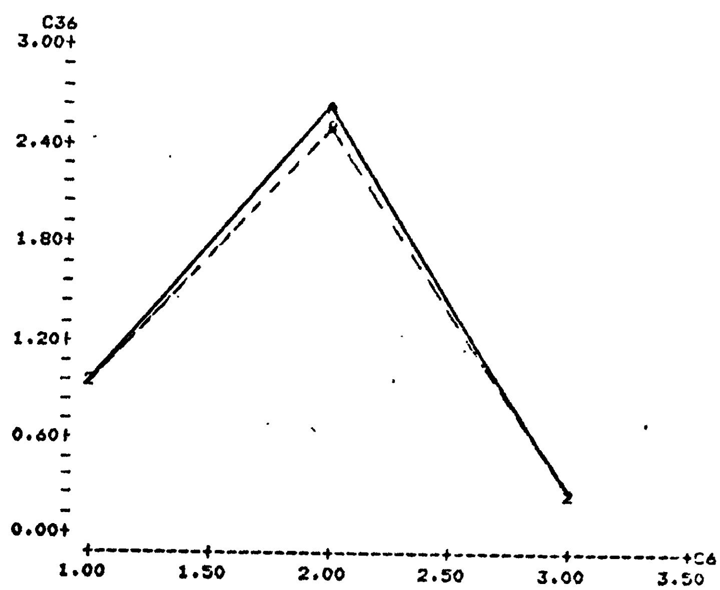
Order

Shindex



Mop & Glow
Step Saver

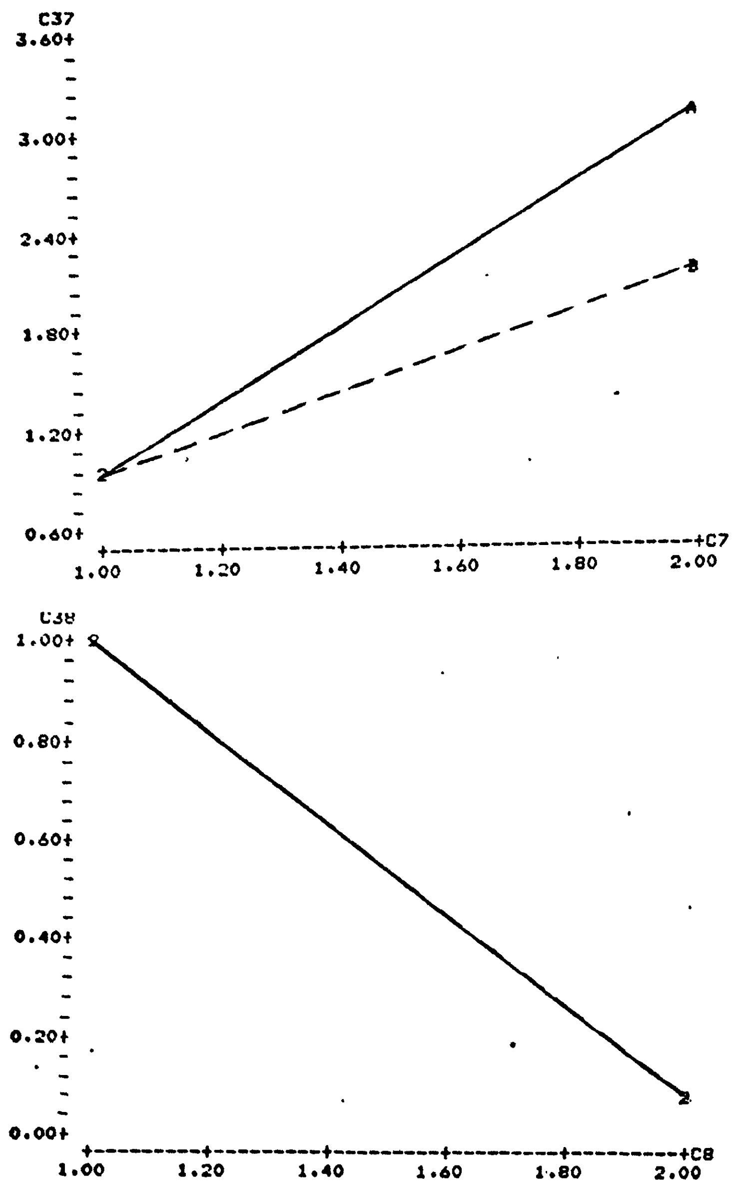
MPlot c36 c6,c96 c6



Snowy
Clorox 2
Miracle White

Order

Shindex

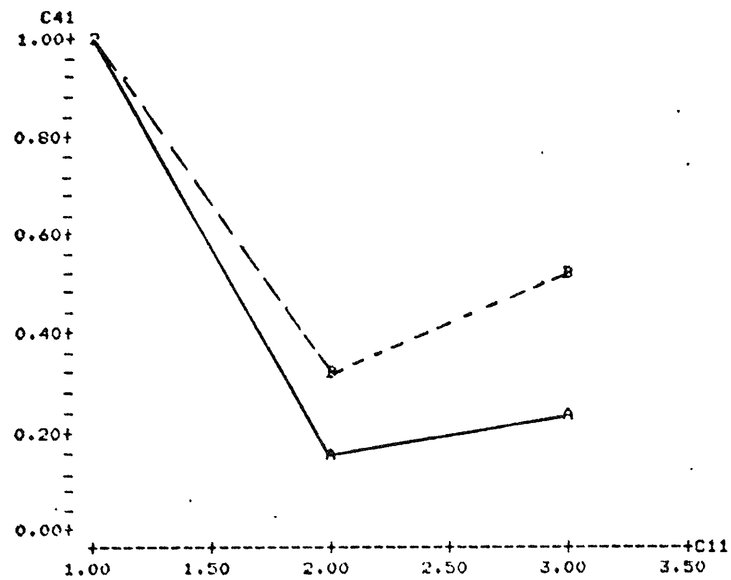


Cling Free
Bounce

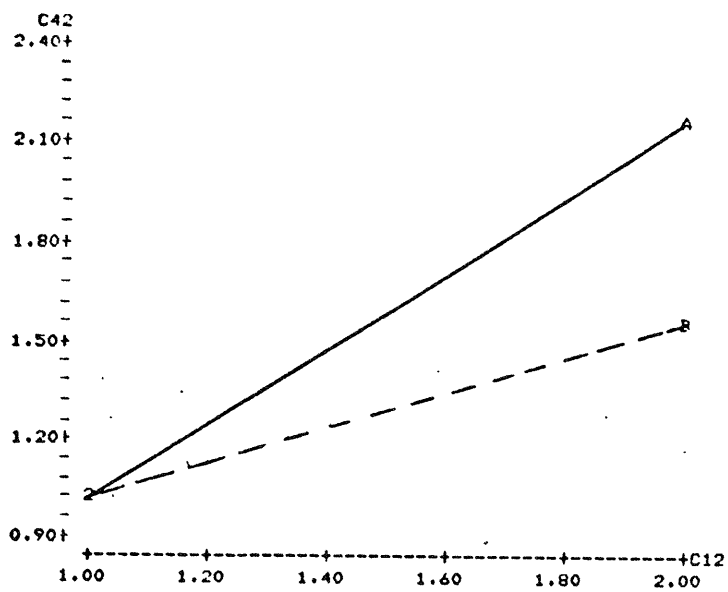
Tylenol
Datril

Order

Shindex



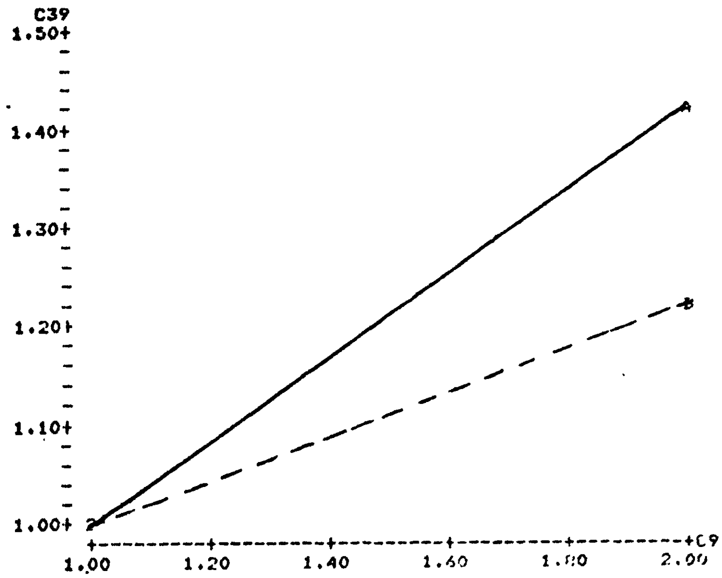
Windex
Ajax
Glass Plus



Biz
Axion

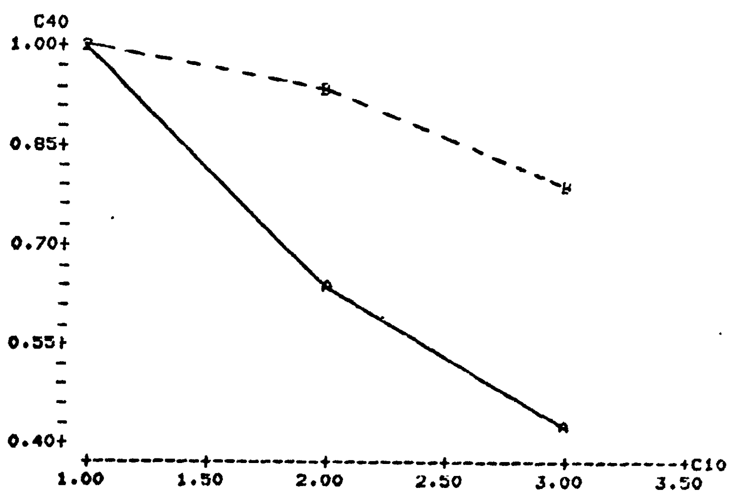
Order

Shindex



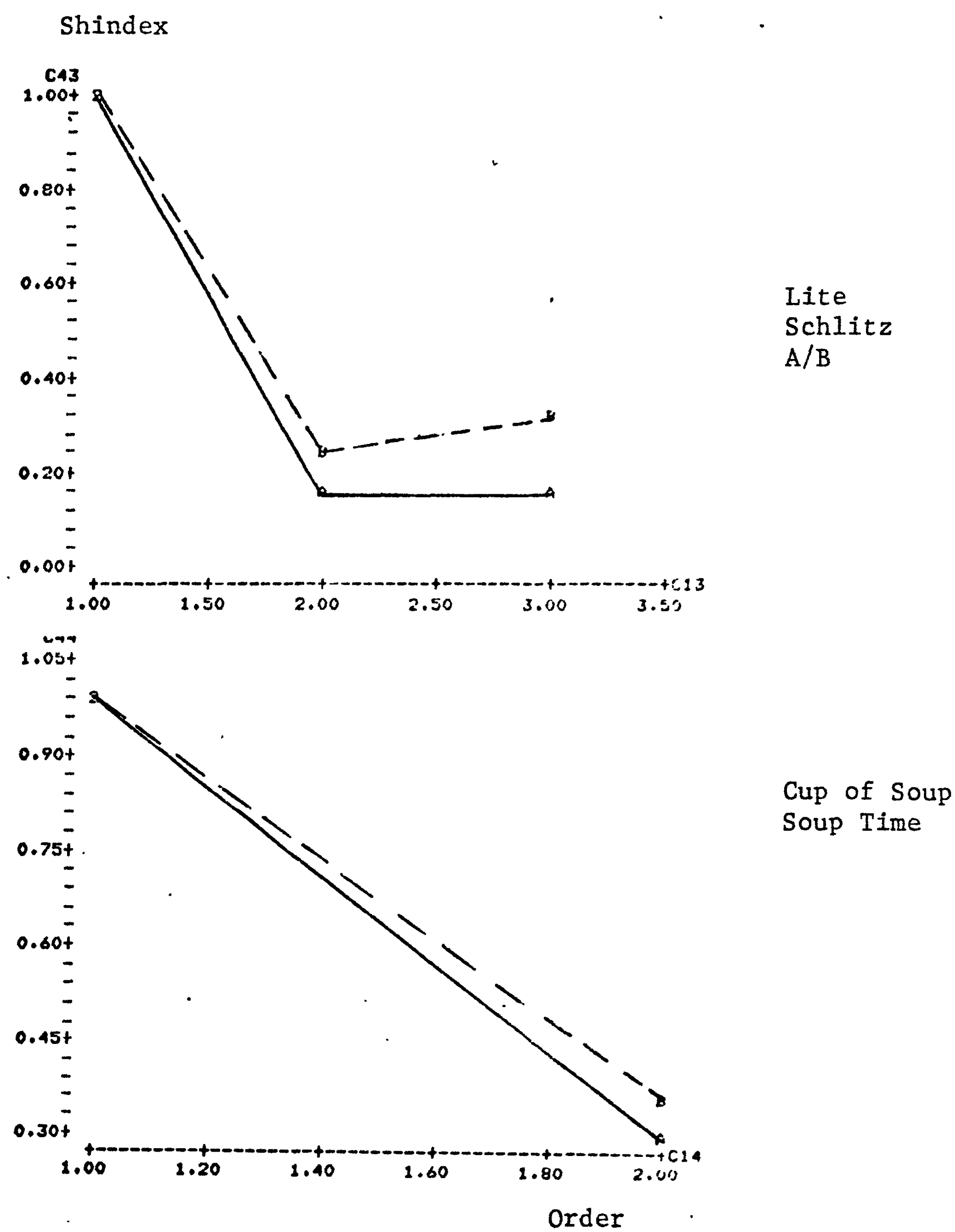
409
Fantastic

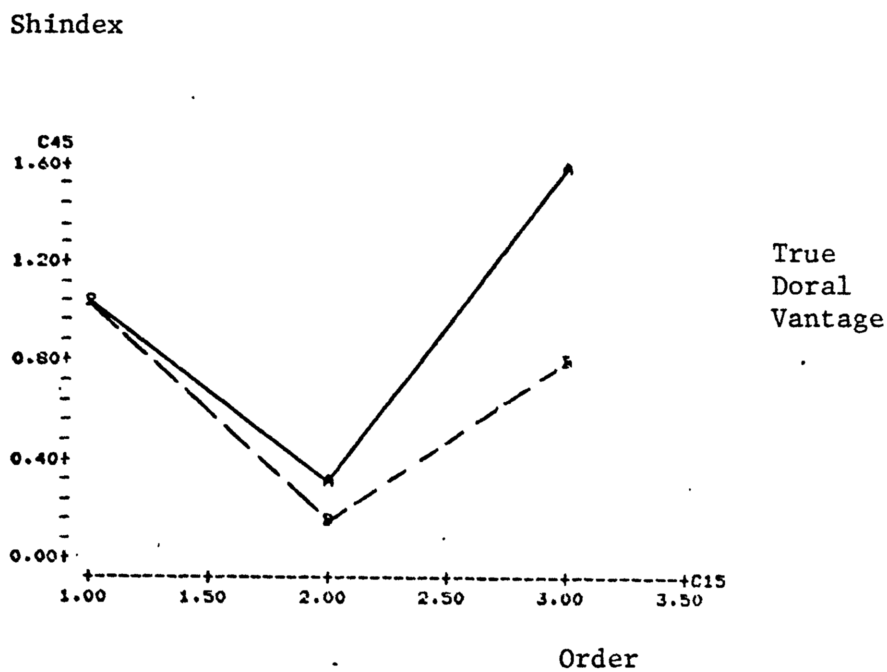
mplot c40 c10 c100 c10



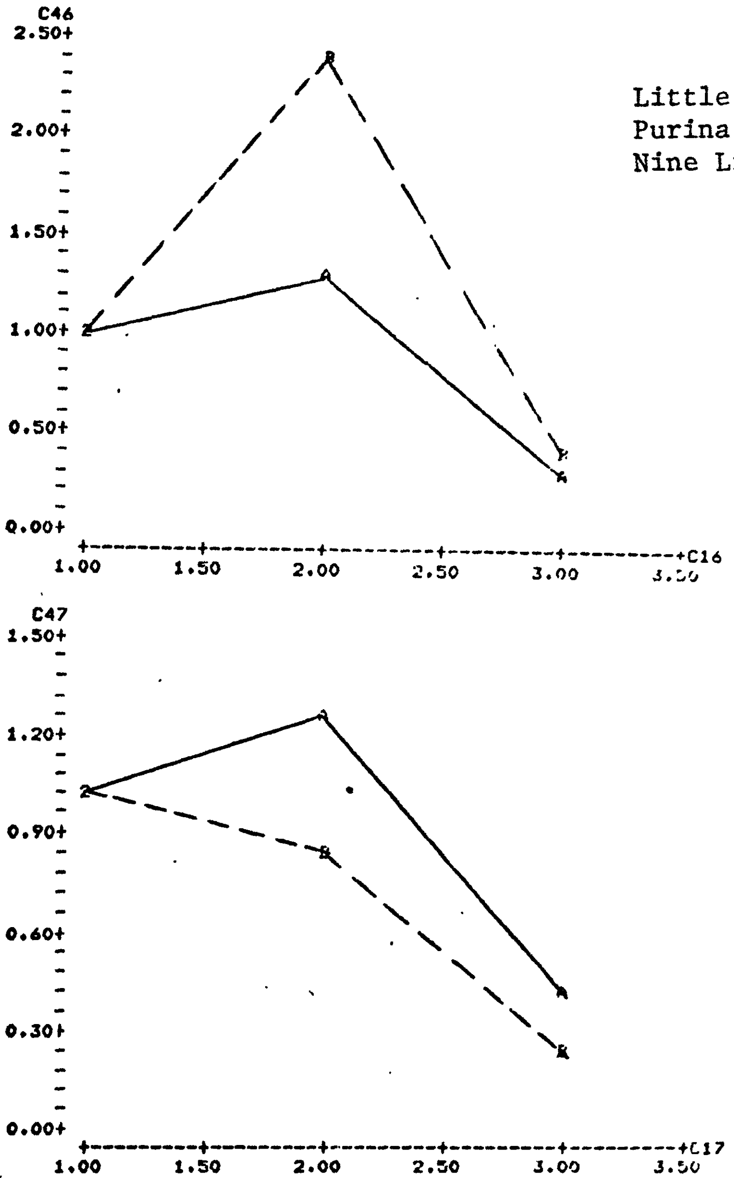
Dial
Zest
Safeguard

Order


Shindex
C43
1.00
0.80
0.60
0.40
0.20
0.00
C13
1.00
1.50
2.00
2.50
3.00
1.05
0.90
0.75
0.60
0.45
0.30
C14
1.00
1.20
1.40
1.60
1.80
2.00
Order
Lite
Schlitz
A/B
Cup of Soup
Soup Time


Shindex
C45
1.60
1.20
0.80
0.40
0.00
1.00
1.50
2.00
2.50
3.00
3.50
C15
True
Doral
Vantage
Order

Shindex

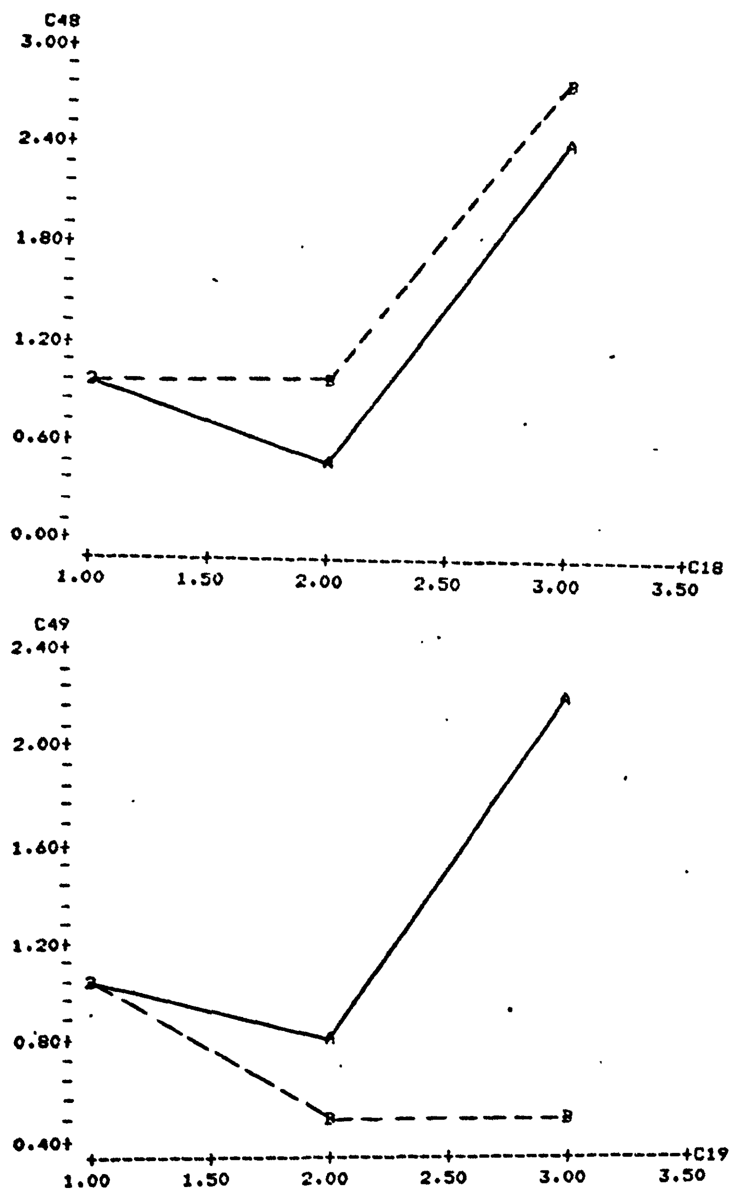


Little Friskies
Purina Cat Chow
Nine Lives

Wizard
Airwick
Renuzit

Order

Shindex

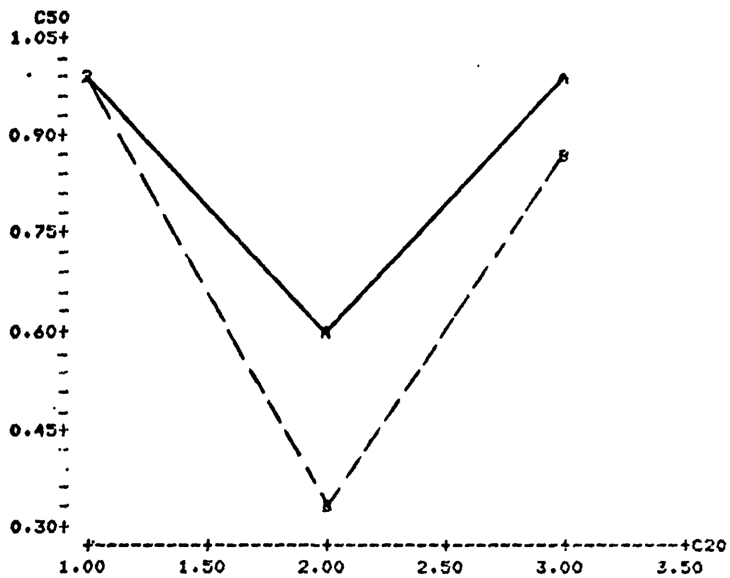


Tums
Digel
Rolaids

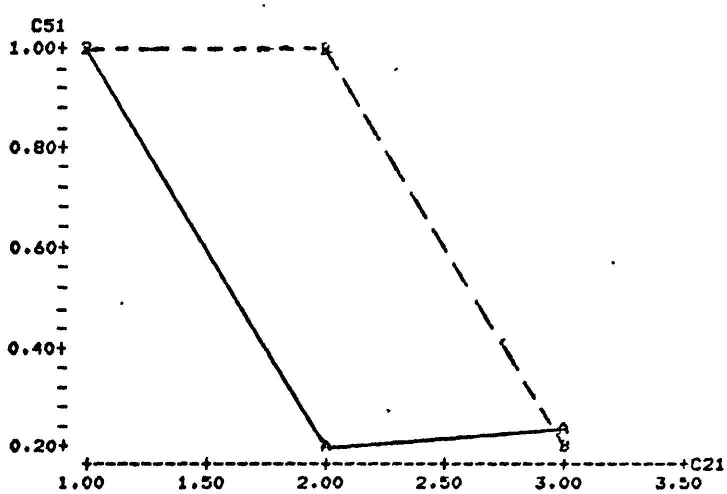
Corricidin
Dristan
Contac

Order

Shindex



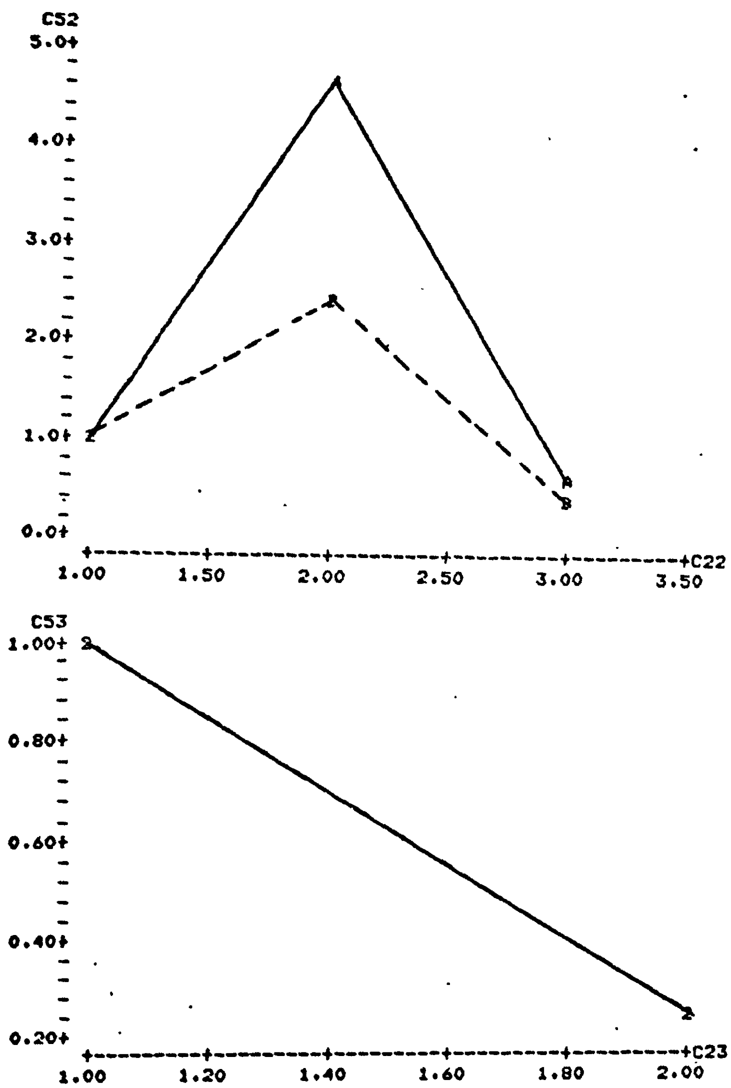
Tampax
Kotex
Playtex



Stayfree
Kotex
New Freedom

Order

Shindex

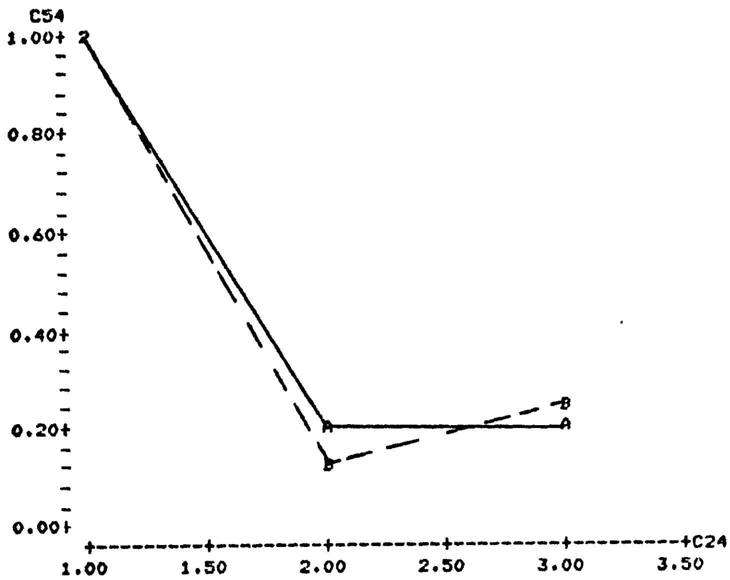


Old English
Pledge
Endust

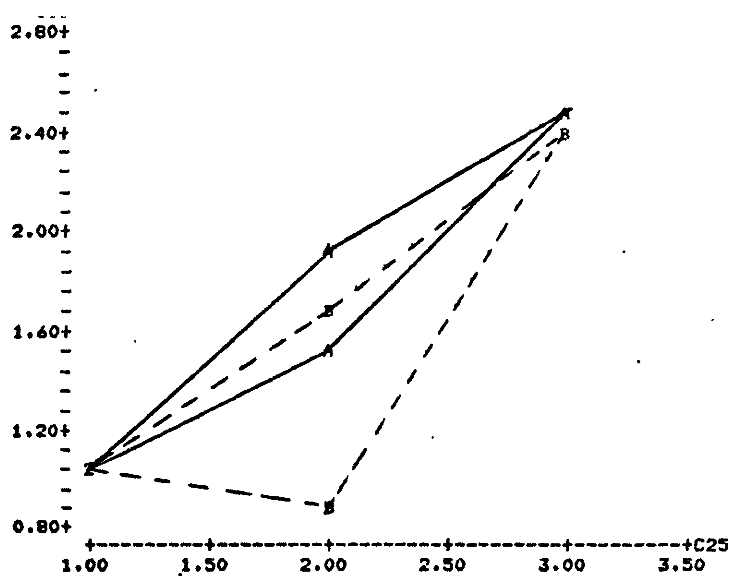
A-1
Heinz 57

Order

Shindex



Dr. Schols
Absorbine Jr.
NP-27



Sleep-eze
Nytol
Sominex

Order

APPENDIX C

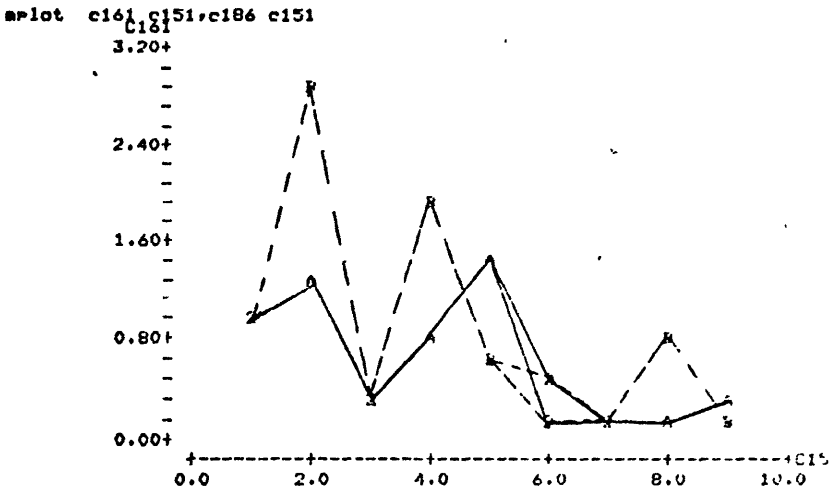
PLOTS OF PREDICTED VS. ACTUAL MARKET SHARES

PREDICTIVE DATA SET ALL BRANDS

- - - = Predicted Market Share

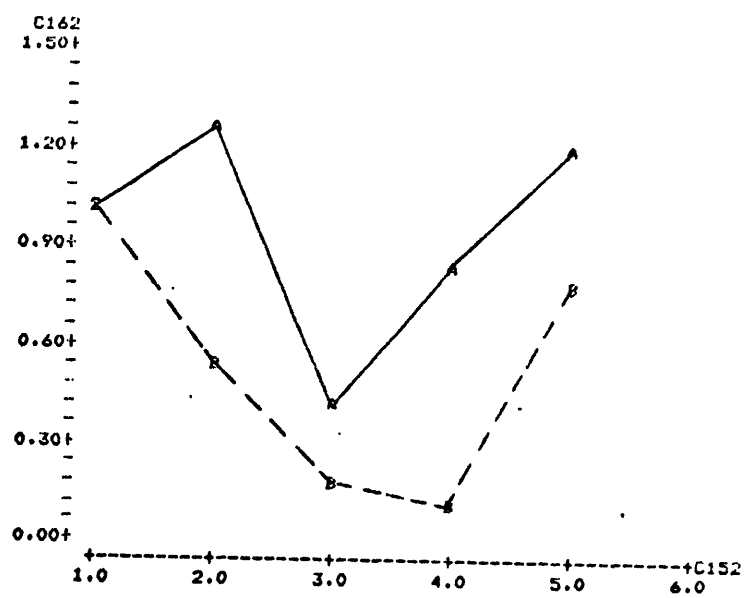
_____ = Actual Market Share

Shindex



Little Friskies
Purina Cat Chow
9-Lives
Purina Special
Dinners
Meow Mix
Chef's Blend
Country Blend
Good Mews
Ocean Blend
Crave

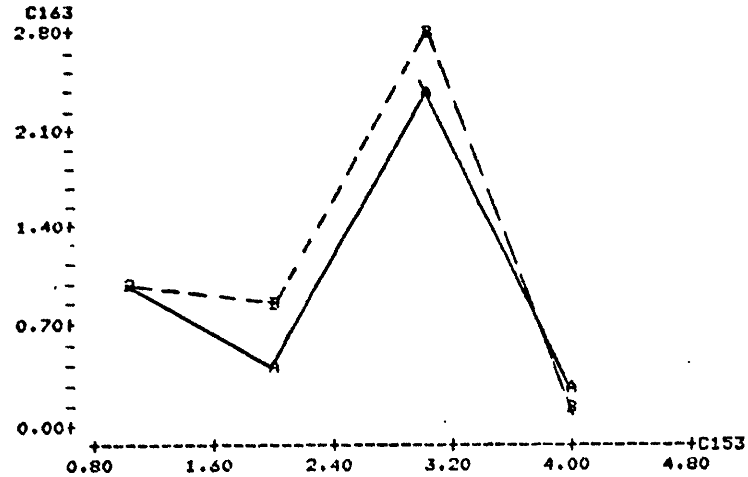
mplot c162 c152,c187 c152



Wizard
Airwick
Renuzit
Glade Solid
Twice as Fresh

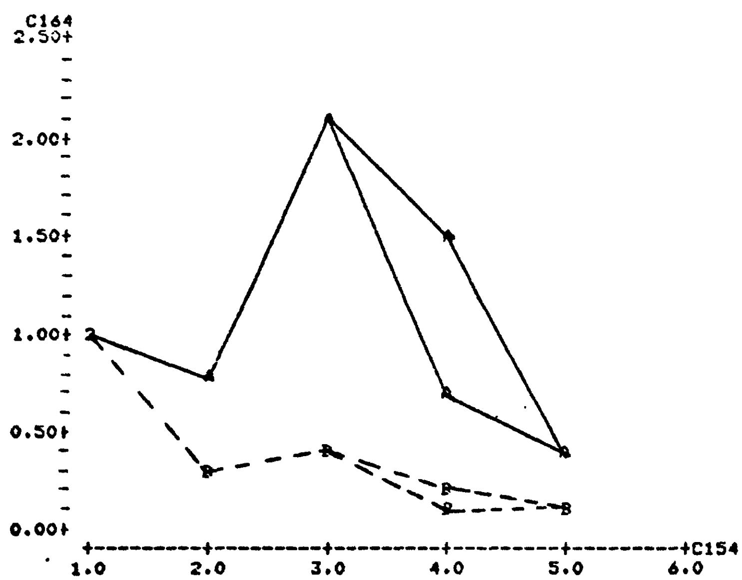
Order

Shindex



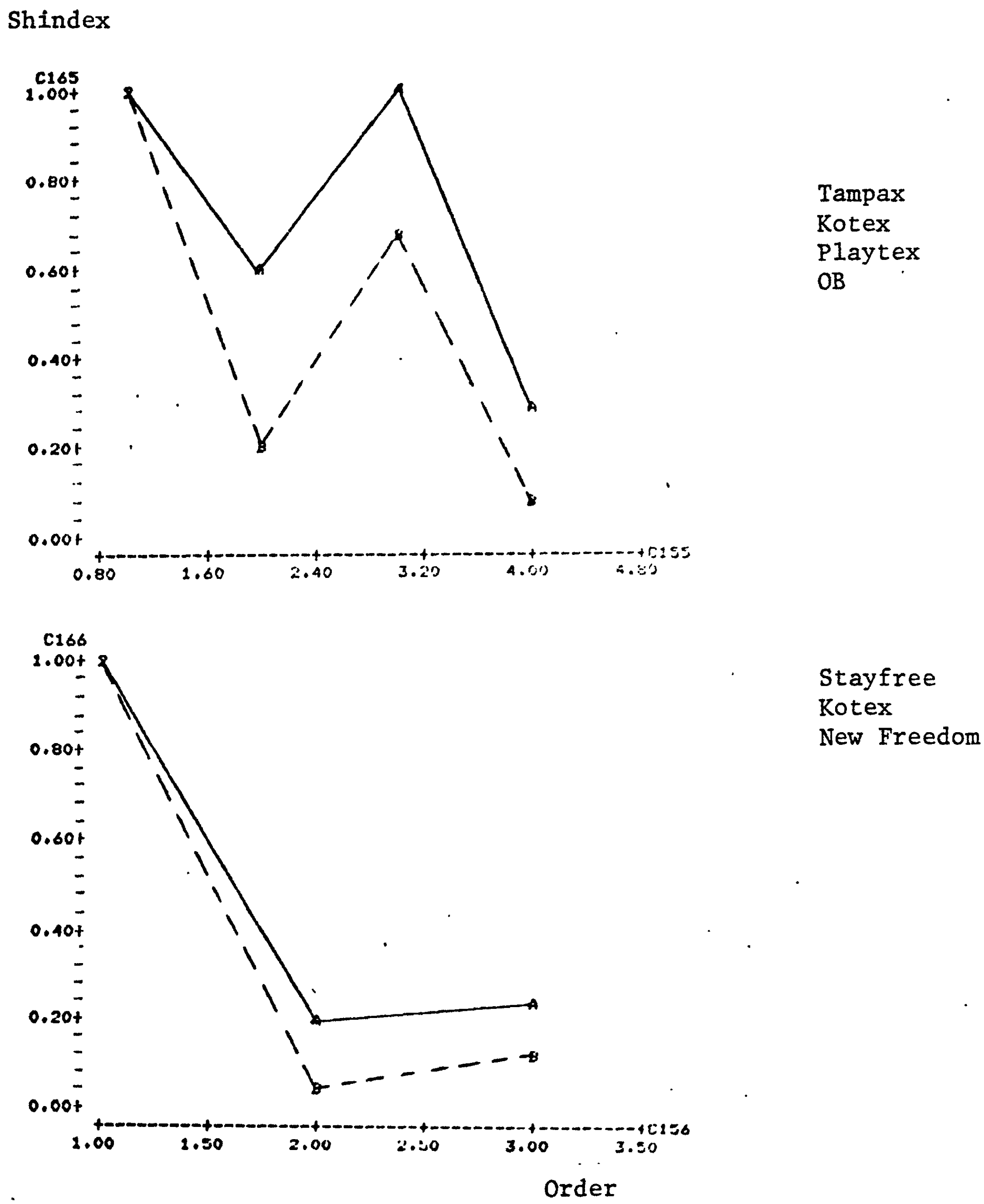
Tums
Digel
Rolaids
Alka 2

aplot c164 c184,c189 c154

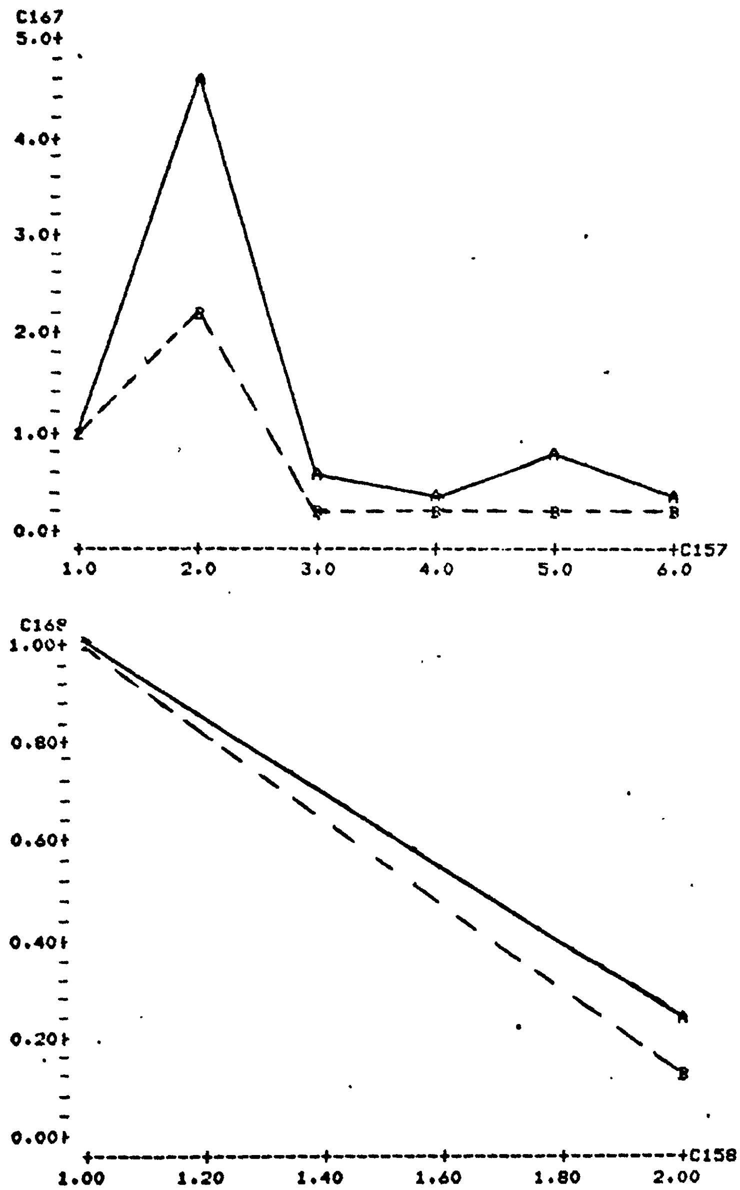


Corricidin
Dristan
Contac
Cotylenol
Nyquil
Daycare

Order


Shindex
C165
1.00
0.80
0.60
0.40
0.20
0.00
0.80
1.60
2.40
3.20
4.00
4.80
C155
Tampax
Kotex
Playtex
OB
C166
1.00
0.80
0.60
0.40
0.20
0.00
1.00
1.50
2.00
2.50
3.00
3.50
C156
Stayfree
Kotex
New Freedom
Order

Shindex

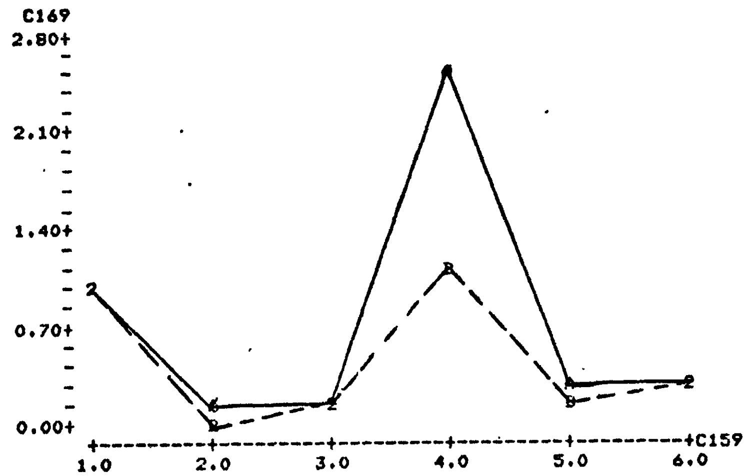


Old English
Pledge
Endust
Favor
Behold
Scott's Liquid Gold

A-1
Heinz 57

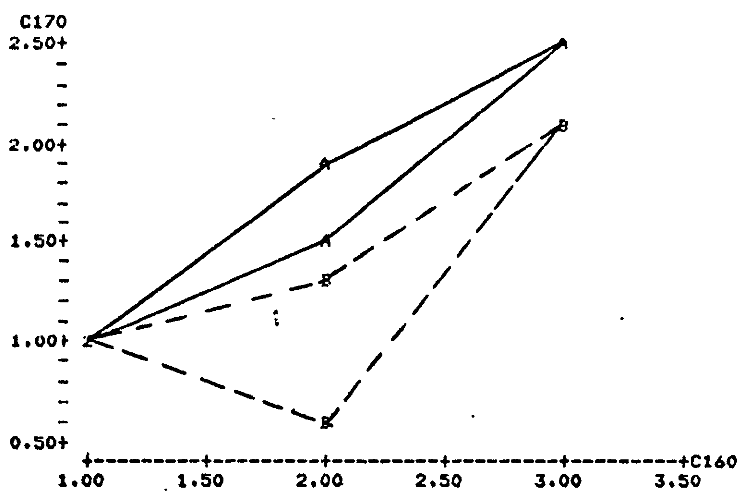
Order

Shindex



Dr. Schols
Absorbine Jr.
NP-27
Desenex
Tinactin
Aftate

mplot c170 c160,c195 c160



Sleep-eze
Nytol
Sominex
Excedrin PM

Order

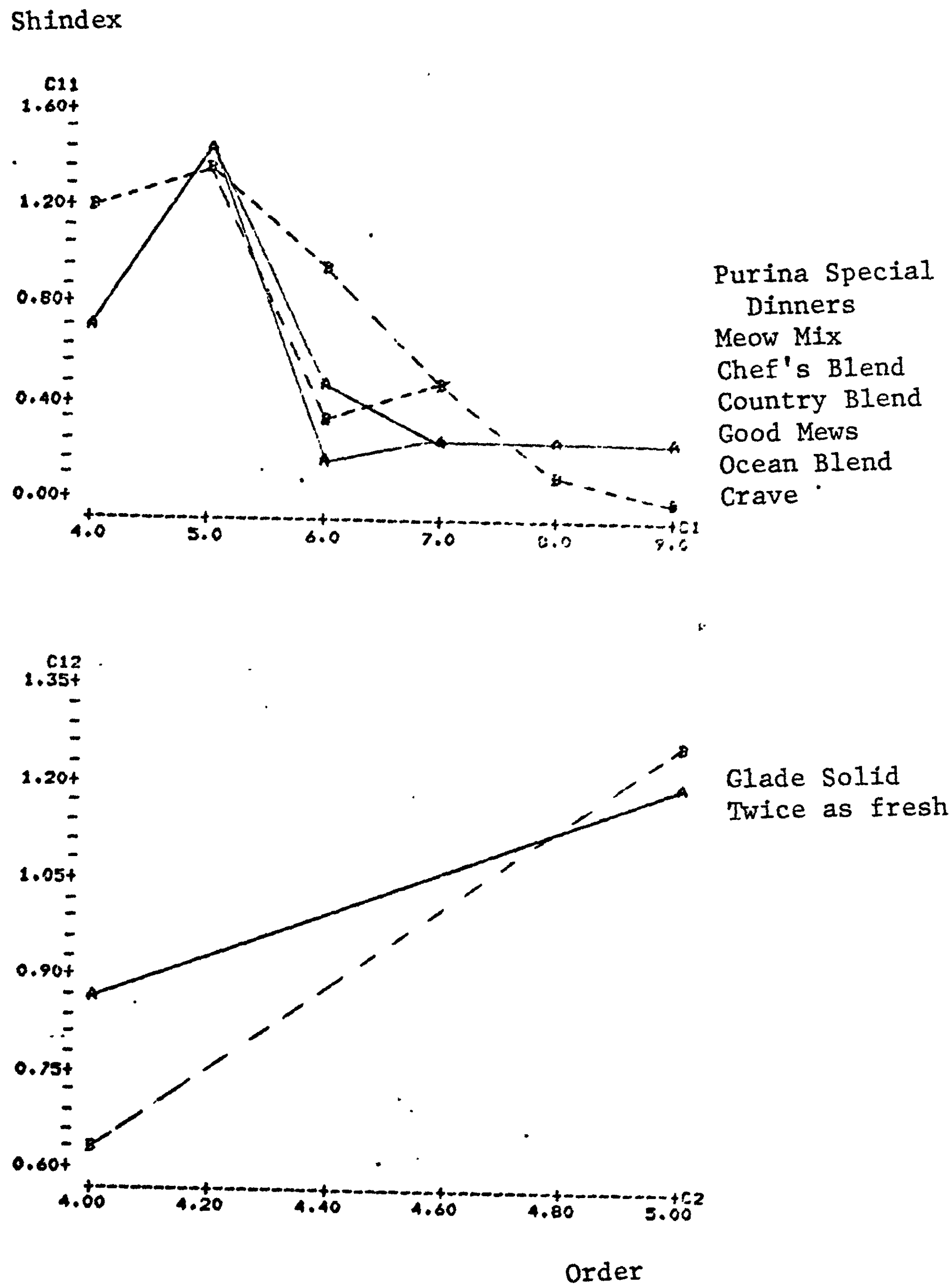
APPENDIX D

PLOTS OF FITTED VS. ACTUAL MARKET SHARES

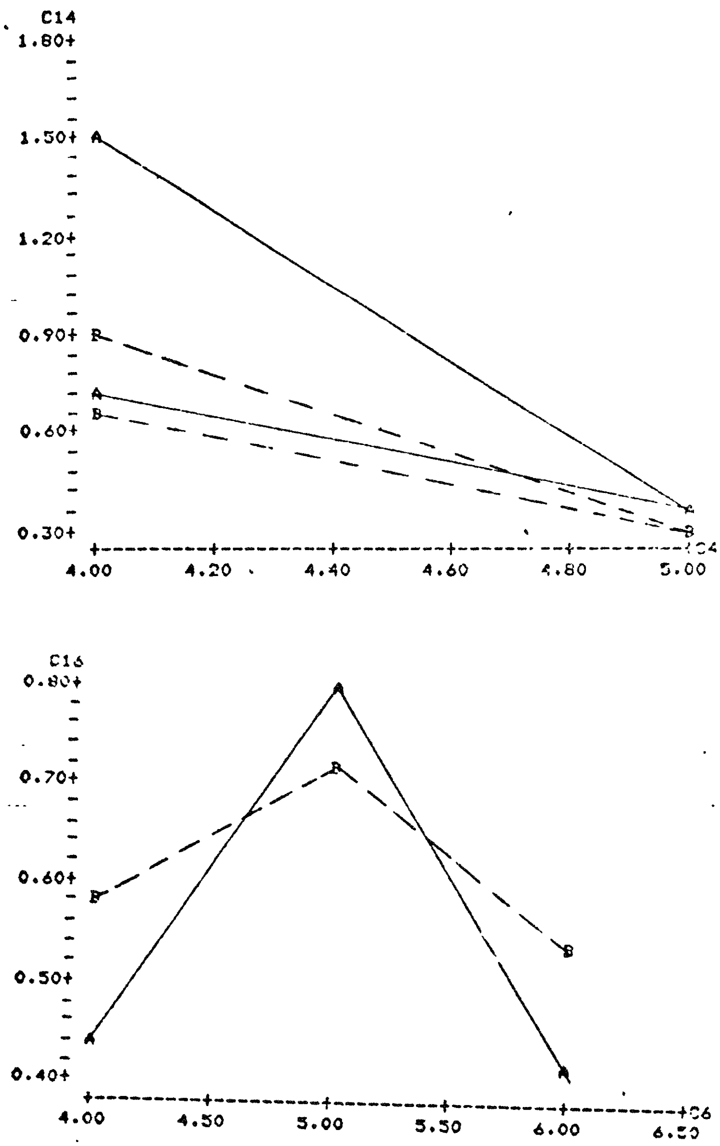
FITTED BRANDS 4^{th} THRU N^{th}

- - - = Predicted Market Share

_____ = Actual Market Share


Shindex
C11
1.60
1.20
0.80
0.40
0.00
4.0
5.0
6.0
7.0
8.0
9.0
C1
Purina Special Dinners
Meow Mix
Chef's Blend
Country Blend
Good Mews
Ocean Blend
Crave
C12
1.35
1.20
1.05
0.90
0.75
0.60
4.00
4.20
4.40
4.60
4.80
5.00
C2
Glade Solid
Twice as fresh
Order

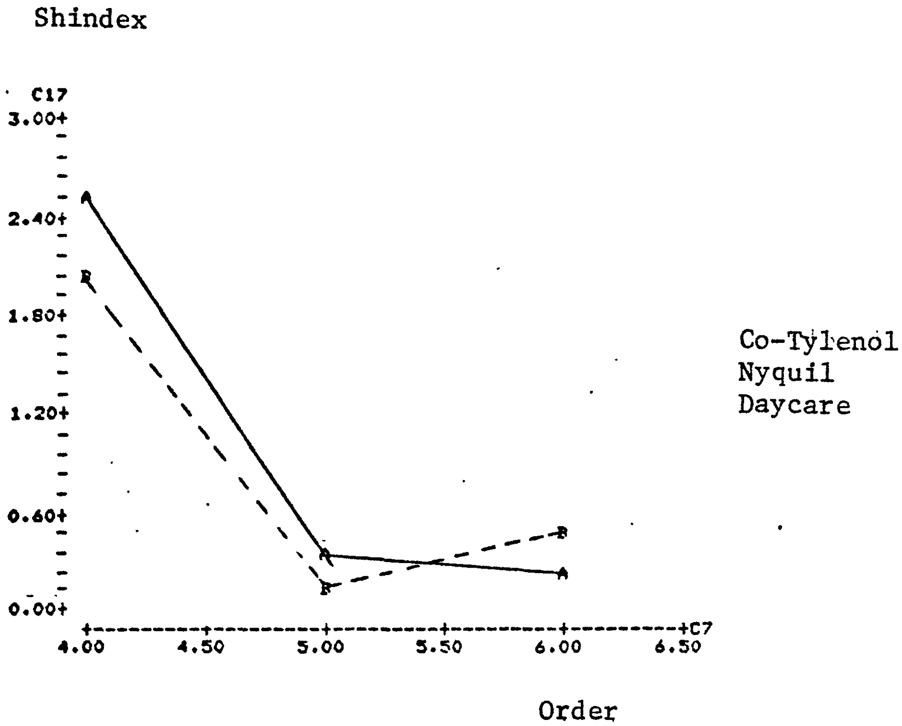
Shindex



Favor
Behold
Scott's
Liquid Gold

Desenex
Tinactin
Aftate

Order



Shindex
C17
3.00
2.40
1.80
1.20
0.60
0.00
4.00
4.50
5.00
5.50
6.00
6.50
C7
Order
Co-Tylenol
Nyquil
Daycare

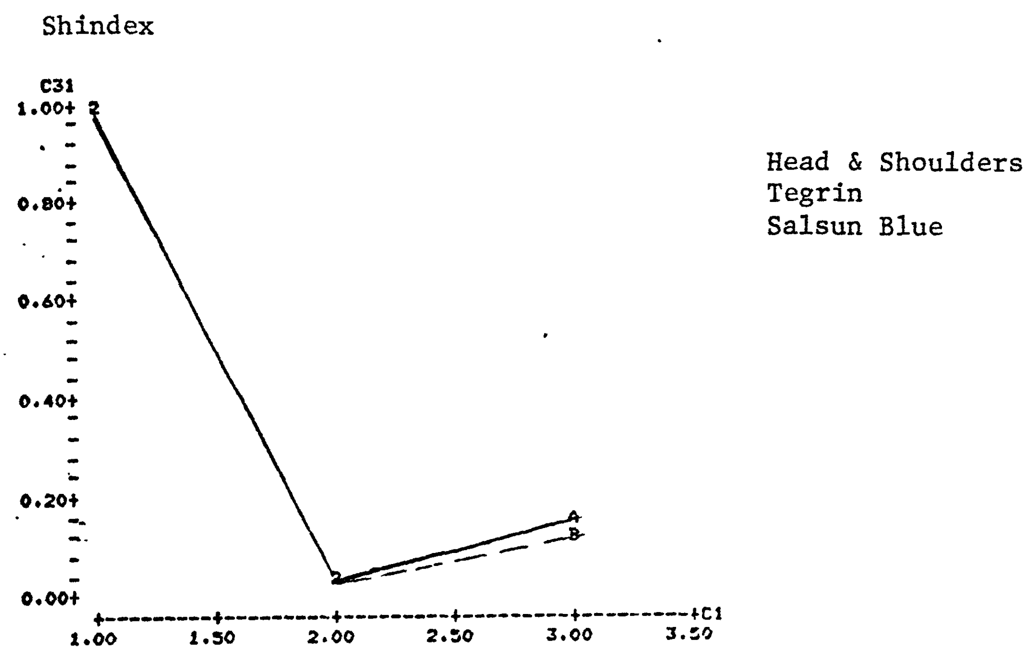
APPENDIX E

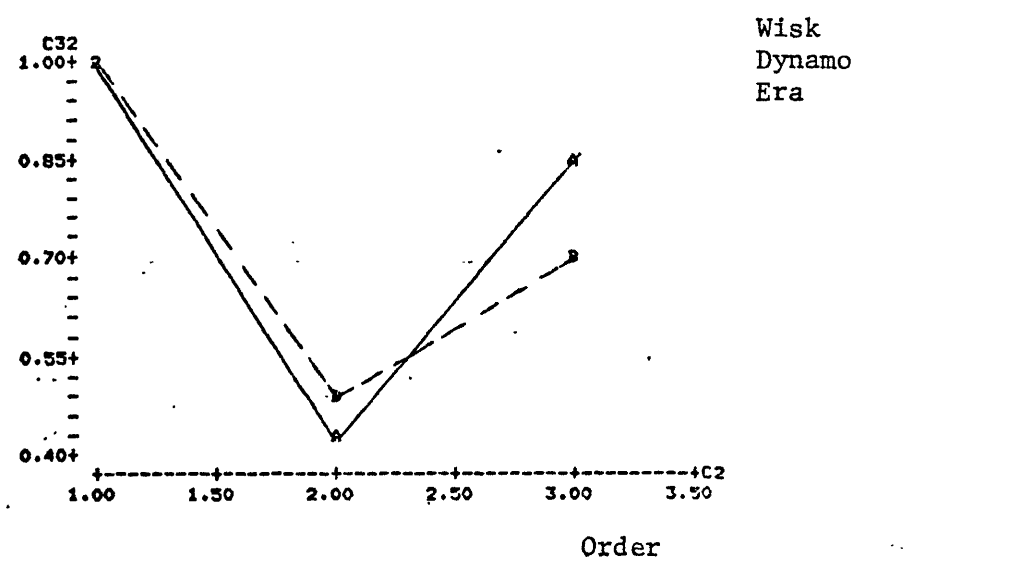
PLOTS OF FITTED VS. ACTUAL MARKET SHARES

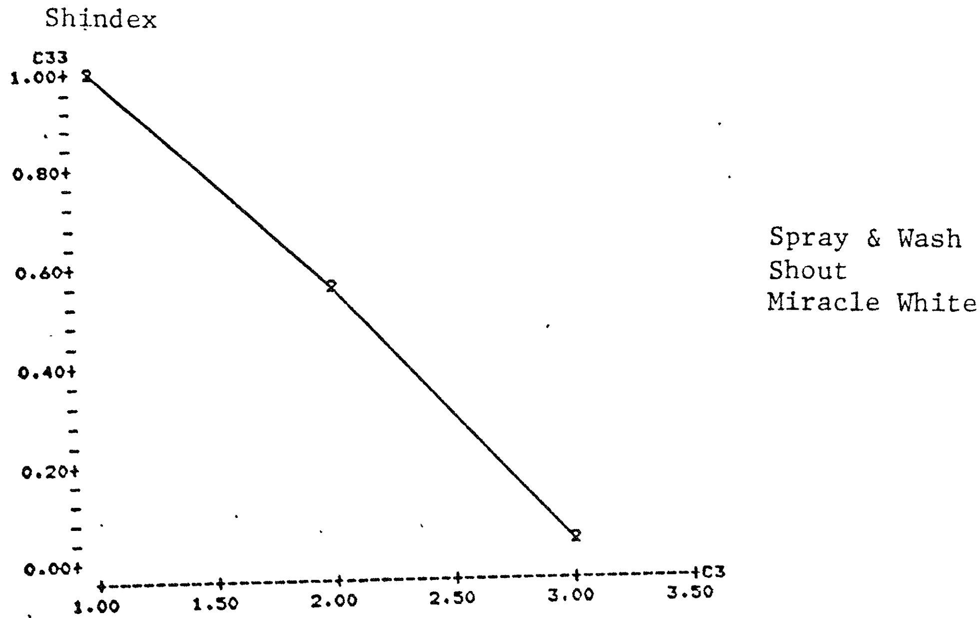
FITTED ESTIMATION SAMPLE 3-BRAND MAXIMUM

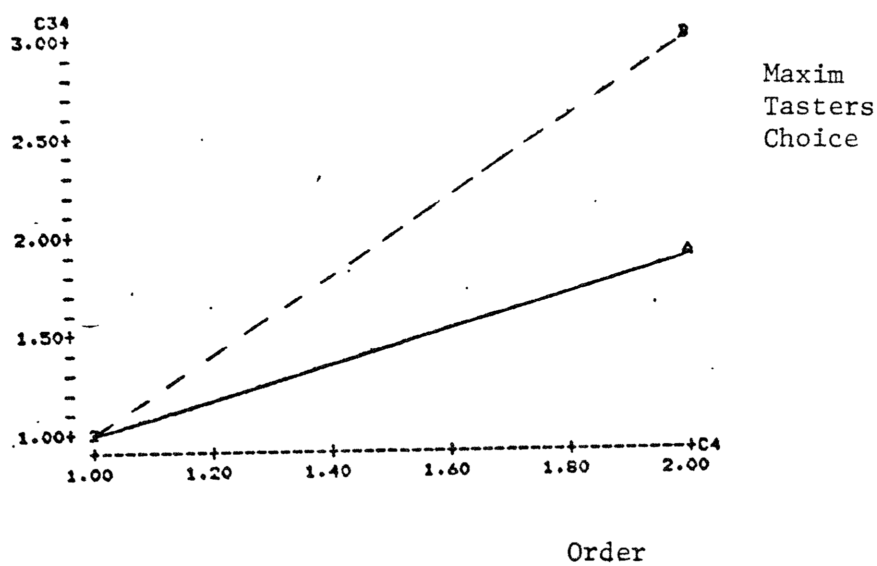
- - - = Predicted Market Share

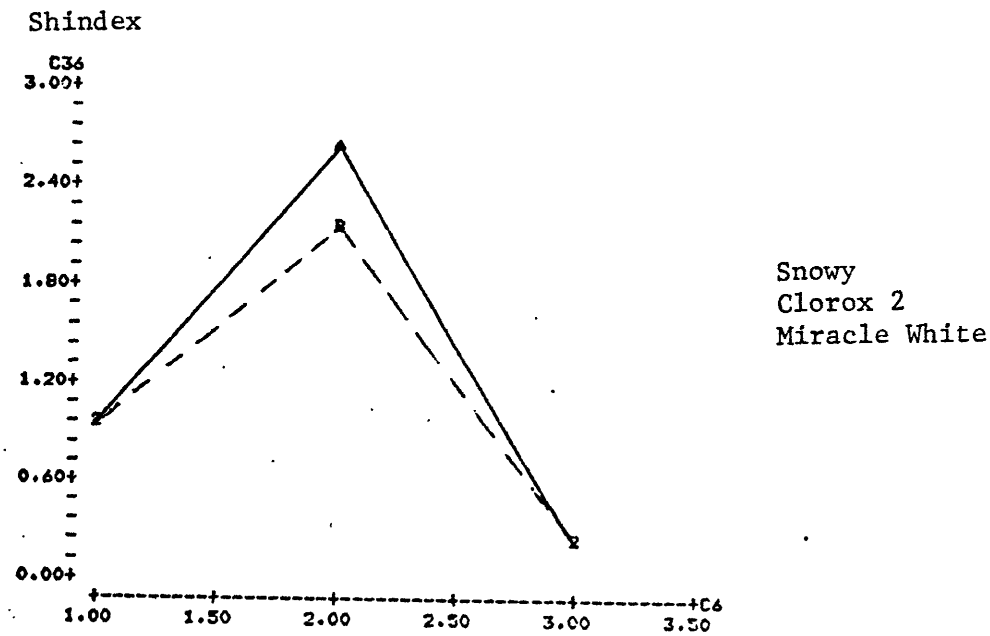
_____ = Actual Market Share

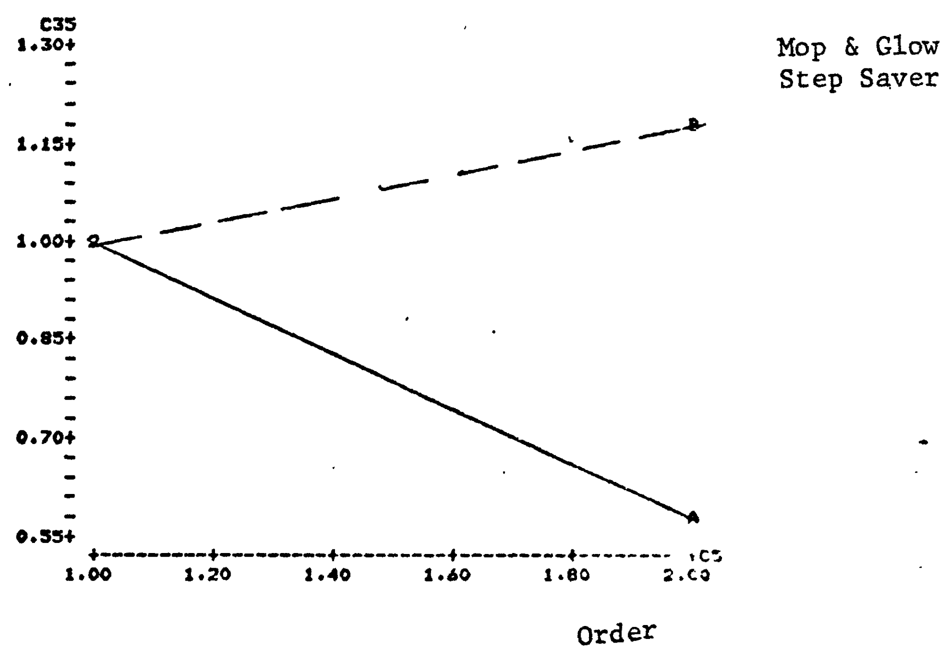

Shindex
C31
1.00
0.80
0.60
0.40
0.20
0.00
1.00
1.50
2.00
2.50
3.00
3.50
C1
Head & Shoulders
Tegrin
Salsun Blue

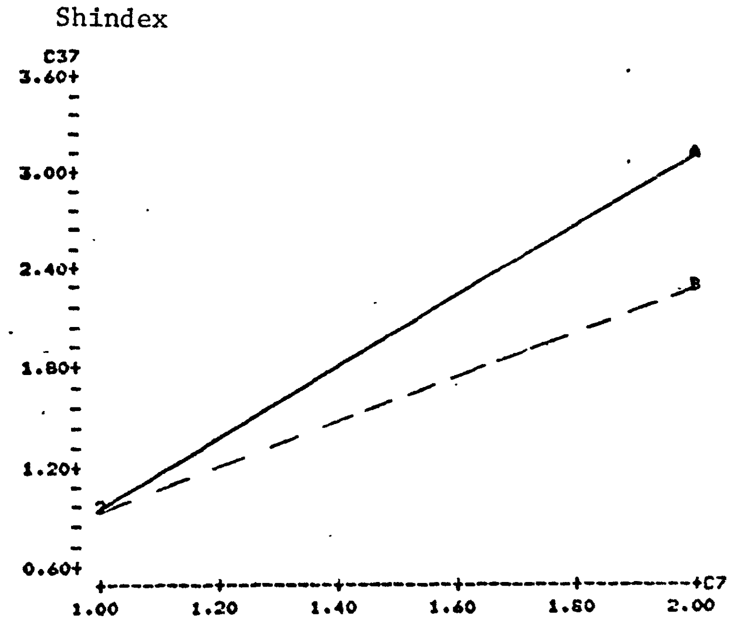

C32
1.00
0.85
0.70
0.55
0.40
1.00
1.50
2.00
2.50
3.00
3.50
C2
Wisk
Dynamo
Era
Order


Shindex
C33
1.00
0.80
0.60
0.40
0.20
0.00
1.00
1.50
2.00
2.50
3.00
3.50
C3
Spray & Wash
Shout
Miracle White

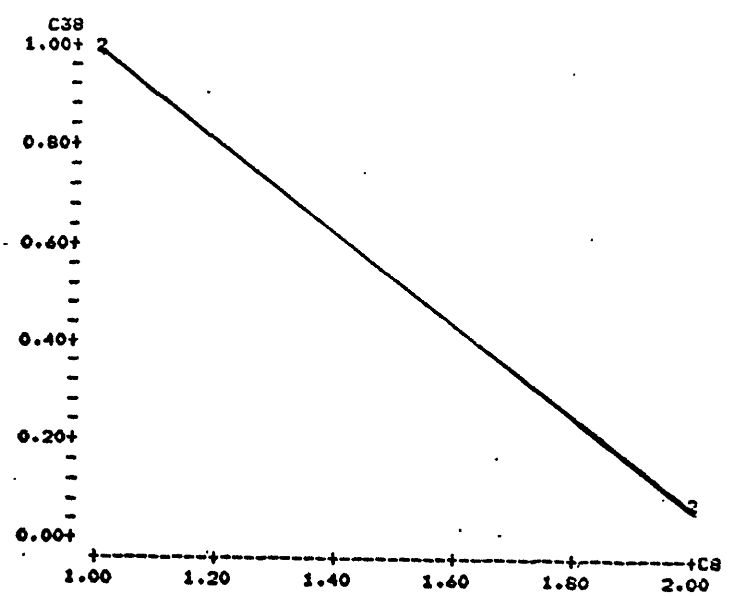

C34
3.00
2.50
2.00
1.50
1.00
1.00
1.20
1.40
1.60
1.80
2.00
C4
Maxim
Tasters
Choice
Order


Shindex
C36
3.00
2.40
1.80
1.20
0.60
0.00
1.00
1.50
2.00
2.50
3.00
3.50
C6
Snowy
Clorox 2
Miracle White

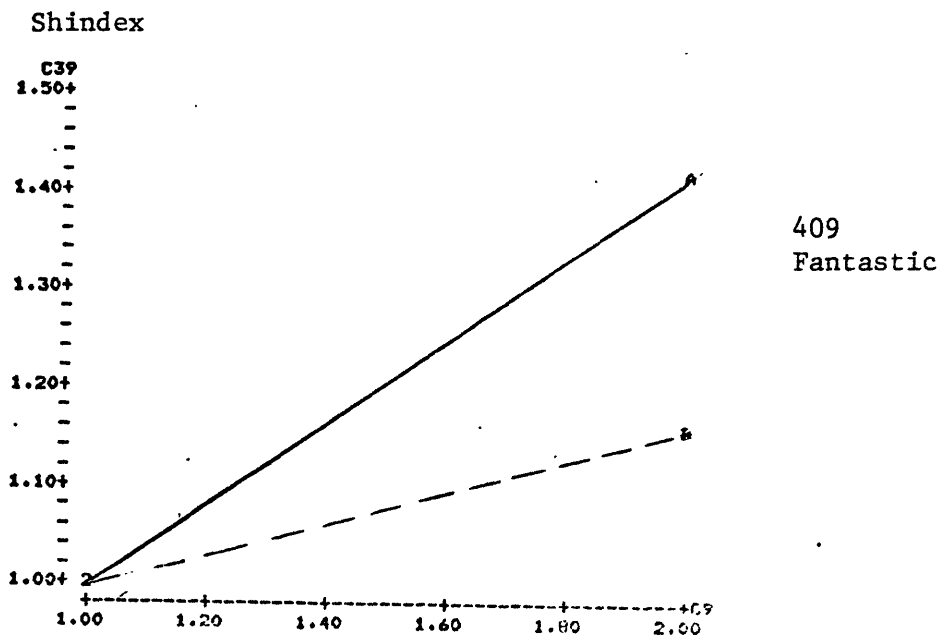

C35
1.30
1.15
1.00
0.85
0.70
0.55
1.00
1.20
1.40
1.60
1.80
2.00
C5
Mop & Glow
Step Saver
Order

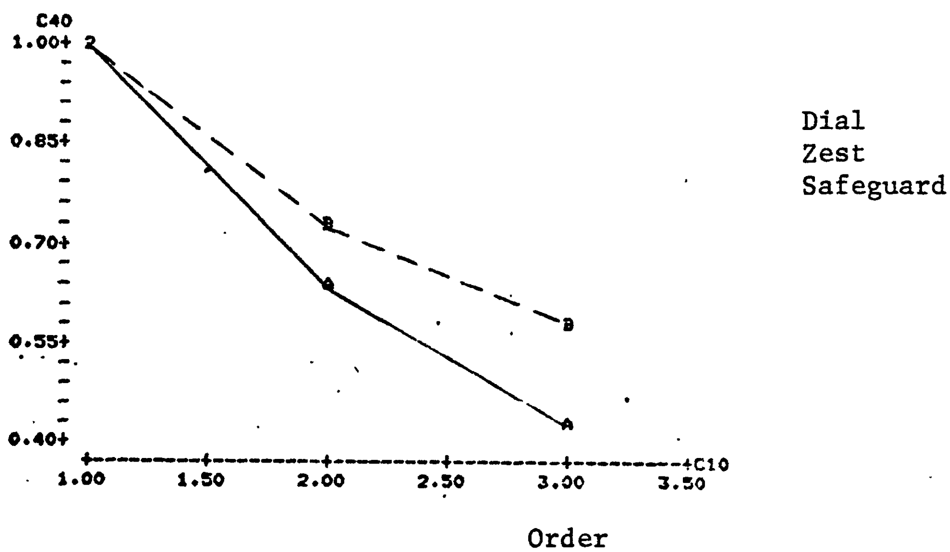


Cling Free
Bounce

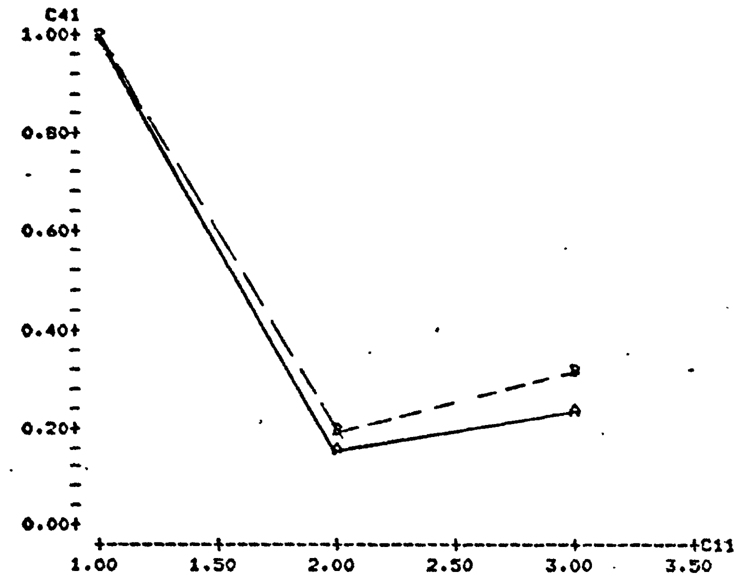


Tylenol
Datril

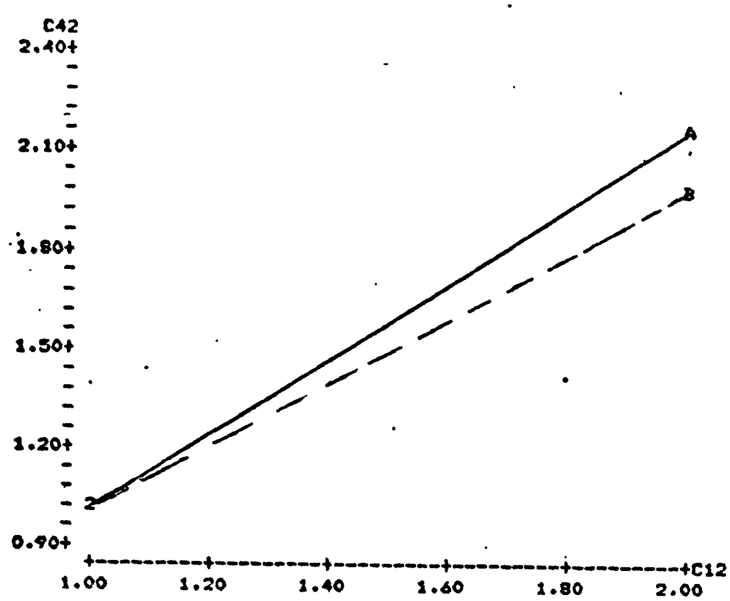

Shindex
C39
1.50
1.40
1.30
1.20
1.10
1.00
1.00
1.20
1.40
1.60
1.80
2.00
C9
409
Fantastic


C40
1.00
0.85
0.70
0.55
0.40
1.00
1.50
2.00
2.50
3.00
3.50
C10
Dial
Zest
Safeguard
Order

Shindex

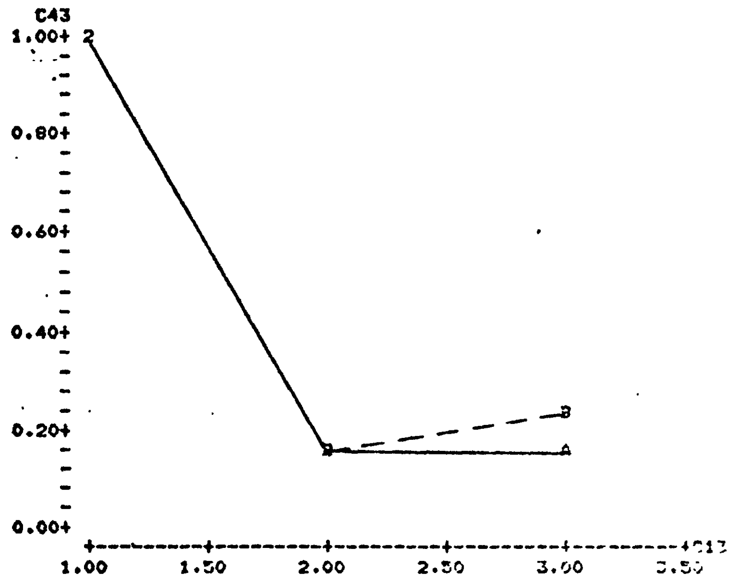


Windex
Ajax
Glass Plus

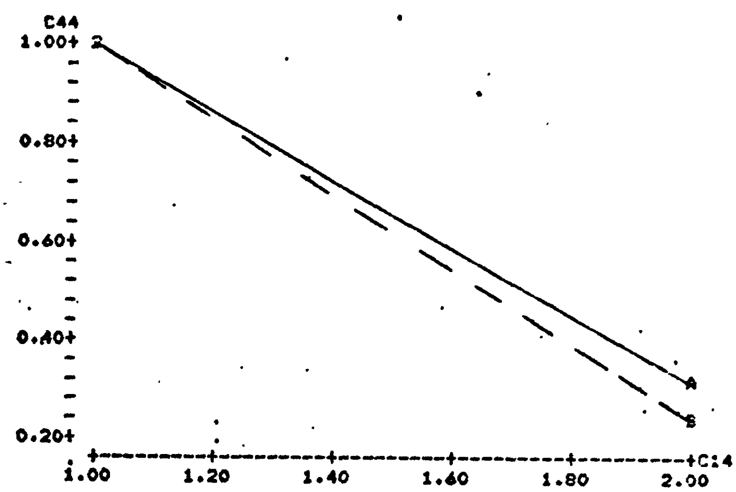


Biz
Axion

Shindex

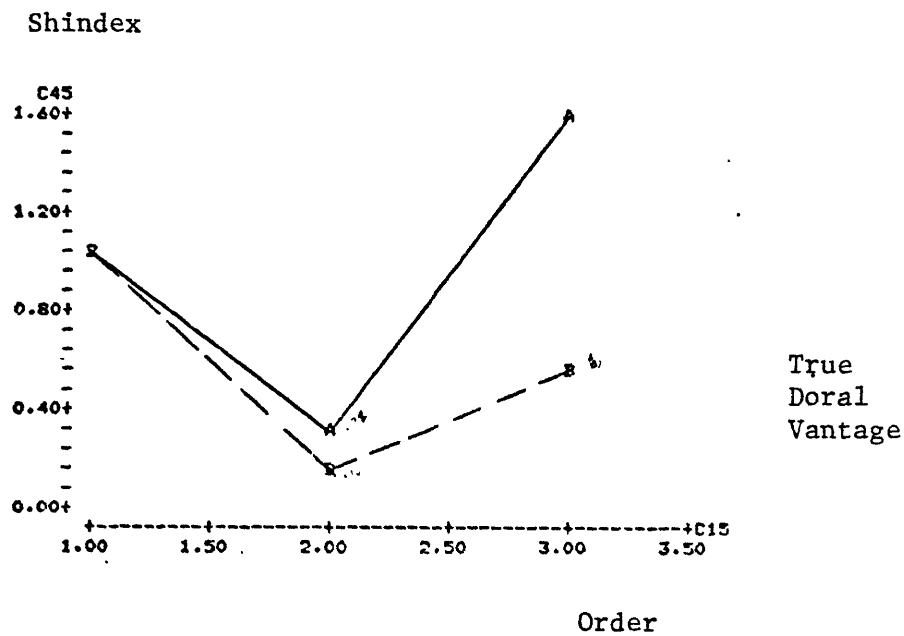


Lite
Schlitz
A/B



Cup-a-Soup
Souptime

Order


Shindex
C45
1.60+
1.20+
0.80+
0.40+
0.00+
+C15
1.00
1.50
2.00
2.50
3.00
3.50
True
Doral
Vantage
Order

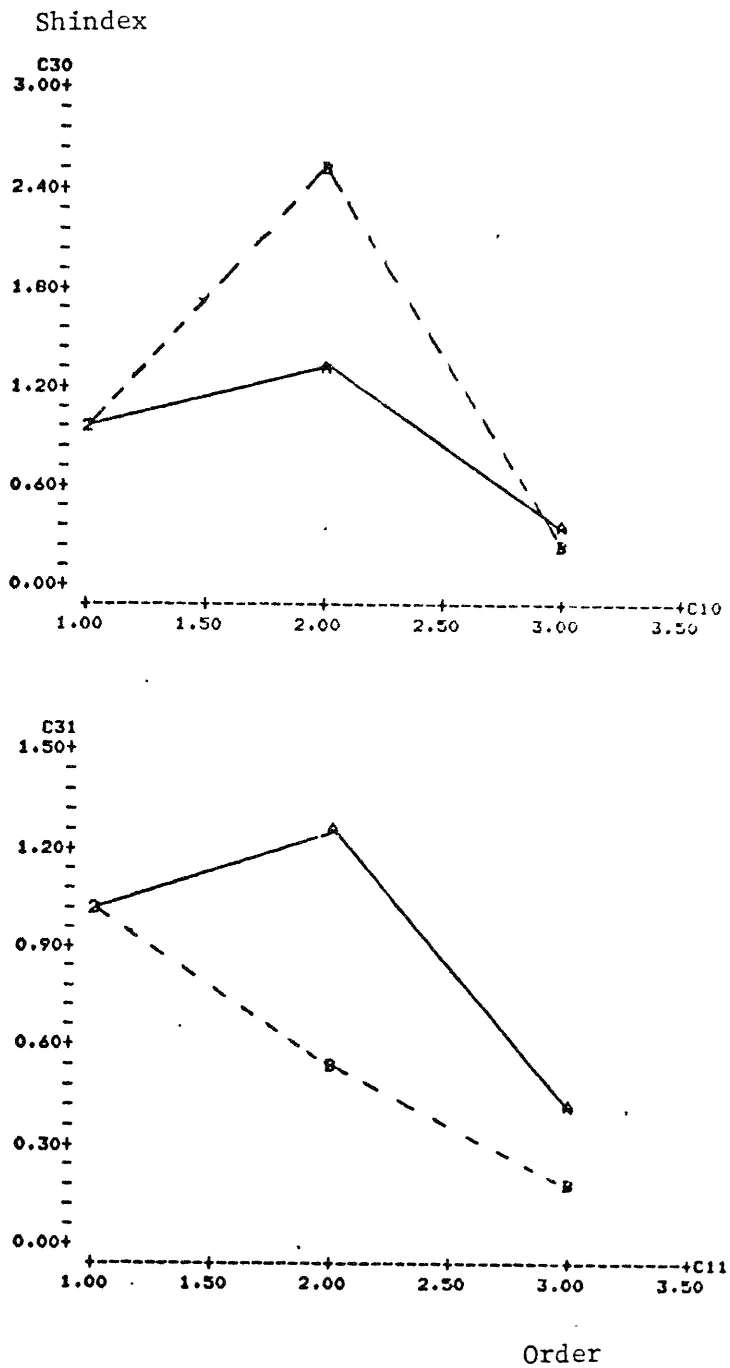
APPENDIX F

PLOTS OF FITTED VS. ACTUAL MARKET SHARES

PREDICTIVE DATA SET 3 BRAND MAXIMUM

\- - - = Predicted Market Share

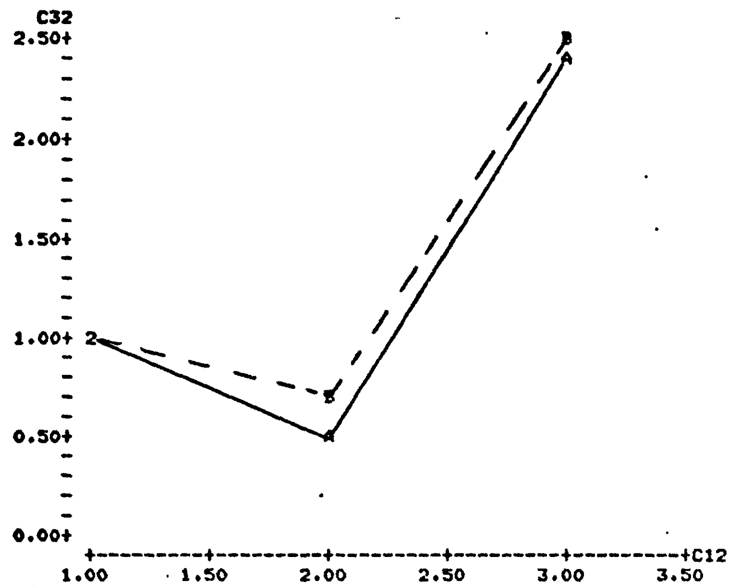
_____ = Actual Market Share



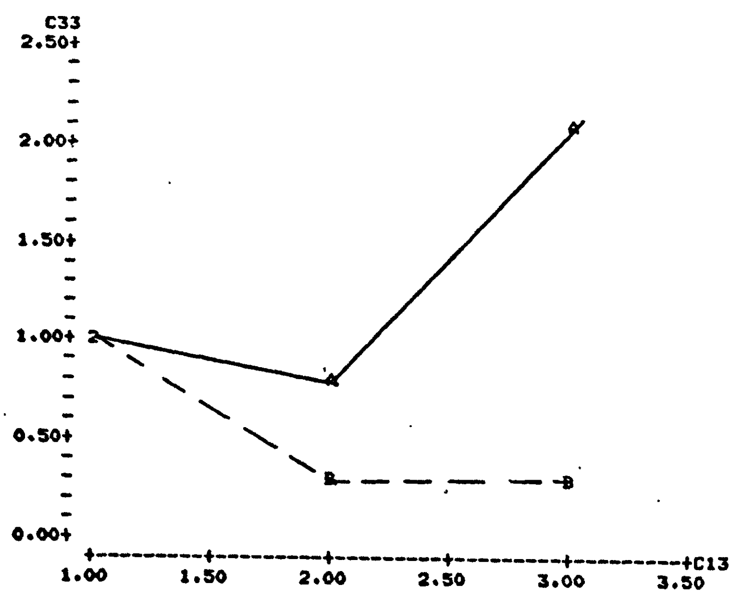
Little Friskies
Purina Cat Chow
Nine-Lives

Wizard
Airwick
Renuzit

Shindex



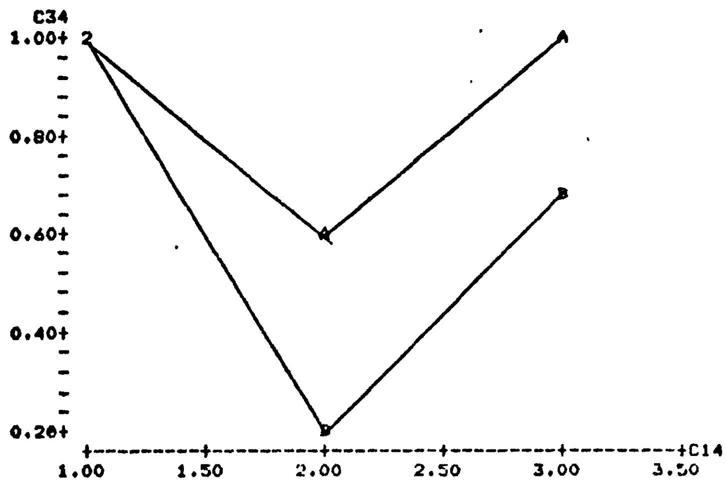
Tums
Digel
Rolaids



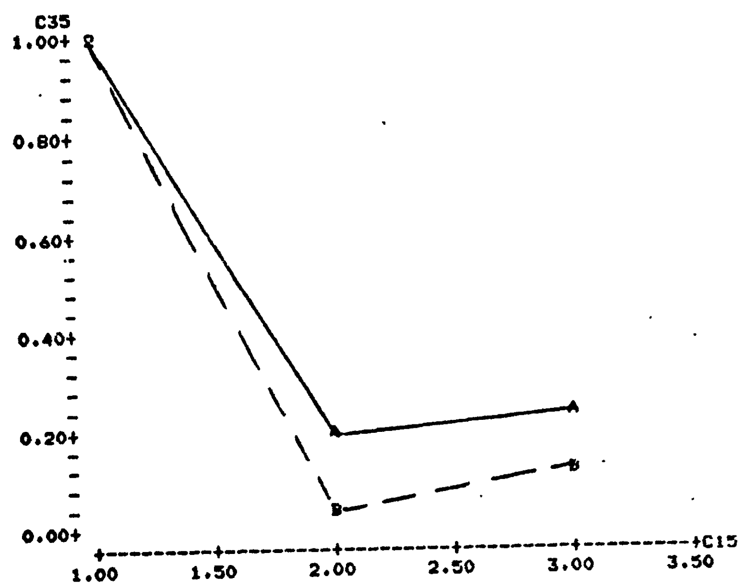
Corricidin
Dristan
Contac

Order

Shindex



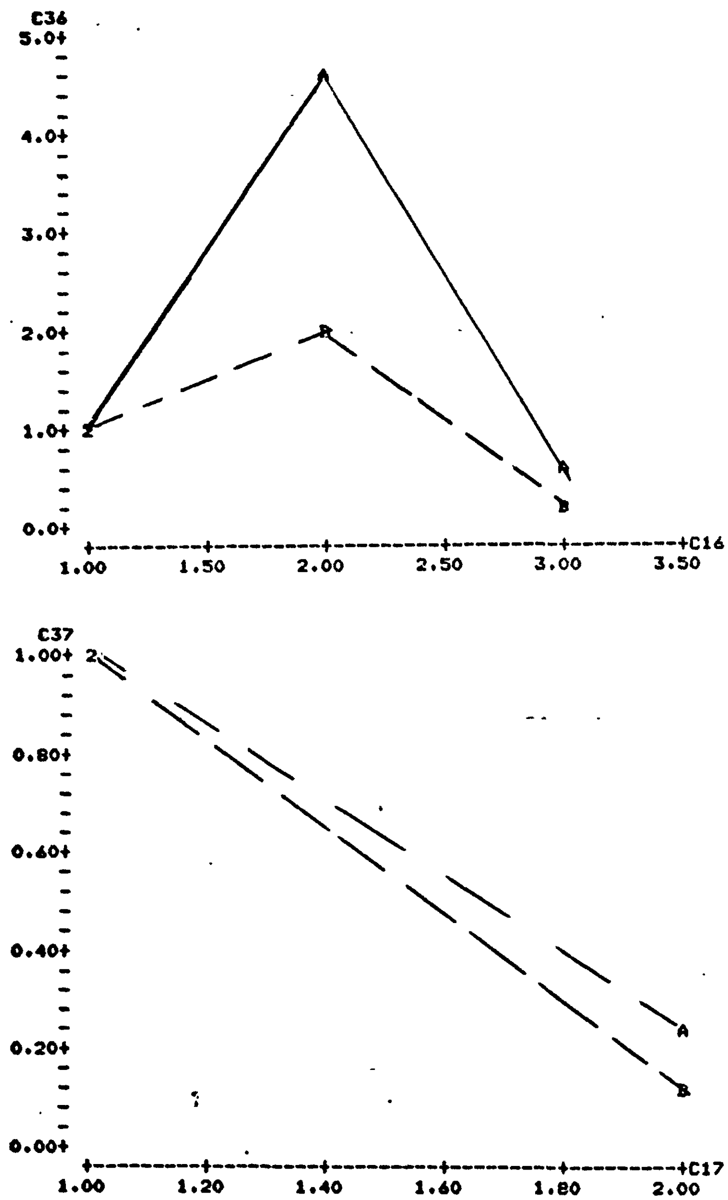
Tampax
Kotex
Playtex



Stayfree
New Freedom

Order

Shindex

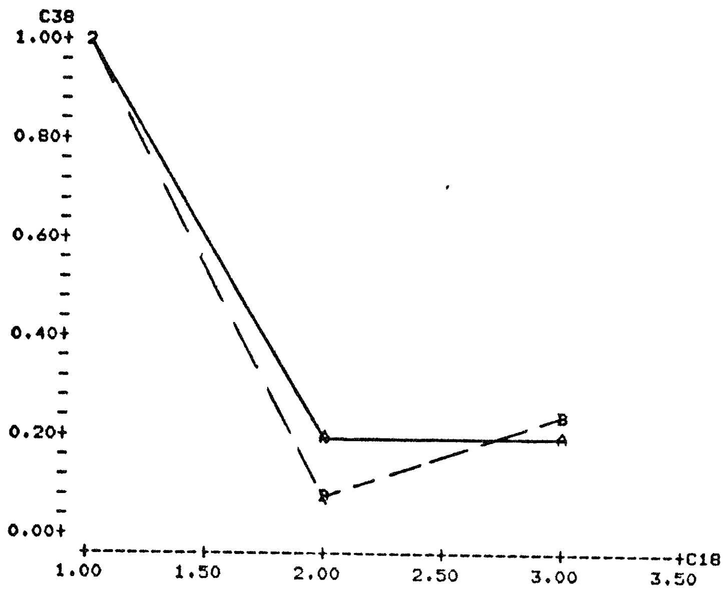


Old English
Pledge
Endust

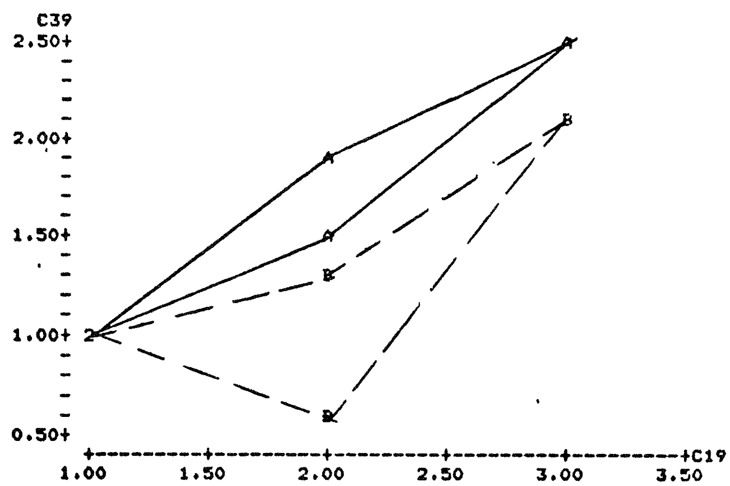
A-1
Heinz 57

Order

Shindex



Dr. Schols
Absorbine Jr.
NP-27



Sleep-eze
Nytol
Sominex

Order

REFERENCES

Bond, R.S. and Lean, D.F., "Consumer Preference, Advertising, and Sales: On the Advantage From Early Entry" Working Paper No. 14, Bureau of Economics, U.S. Federal Trade Commission, Washington, October 1979.

Buzzel, and Farris, "Marketing Costs In Consumer Goods Industries," Report 76-111, Marketing Science Institute, Cambridge, August 1976.

Goldberger, A., Teoria Ekonometric, (Warszawa, Polska: PWE 1972).

Greń, J., Statystyka Matematyczna, Modele i Zadania, (Warszawa, Polska: PWN 1979).

Johnston, J., Econometric Methods, (New York, McGraw-Hill Book Co., 1960).

Lane, W.J., "Product Differentiation in a Market with Endogenous Sequential Entry", Bell Journal of Economics.

Miller, R. and Dean W. Wichern, Intermediate Business Statistics: Analysis of Variance, Regression and Time Series, (Holt, Rinehart & Winston, 1977).

Nagle, T.T., "Does Advertising Really Create a Barrier to Entry?, Working Paper Series, Center for Research in Marketing (Chicago, IL: Graduate School of Business, University of Chicago).

Schmalensee, R., "Product Differentiation Advantages of Pioneering Brands," Working Paper No. (Cambridge, Ma: Sloan School of Management, MIT, August 1980).

Silk, A.J. and G.L. Urban, "Pre-Test Market Evaluation of New Packeged Goods: A Model and Measurment Methodology," Journal of Marketing Research, Vol. 15, No. 2 (May 1978), pp. 171-191.

Urban, G.L., and John R. Hauser, Design and Marketing of New Products, (Englewood Cliffs, NJ: Prentice-Hall, 1980).

REFERENCES (CONT.)

Urban, G.L., P.L. Johnson, R.H. Brudnick, "Market Entry Strategy Formulation: A Hierarchical Modelling and Consumer Measurement Approach," Working Paper No. 1103-80 (Cambridge, MA: Sloan School of Management, MIT, December 1979).

Von Hippel, E., "Appropriability of Innovation Benefit as a Predictor of the Functional Locus of Innovation," Working Paper No. 1084-79 (Cambridge, MA; Sloan School of Management, MIT, June 1979).

Welsch, R.E., E. Kuh, and D. Belsley, Regression Diagnostics, (New York: John Wiley & Sons, 1980).

Welsch, R.E., D.C. Hoaglin, "The Hat Matrix in Regression and ANOVA," The American Statistician, 32, pp. 17-22 and Corrigenda 32, p. 146.