ON SOME ANALYTICAL APPROACHES TO THE STUDY OF CONSUMER BRAND-SWITCHING BEHAVIOR

by

PIERRE MARIE WINDAL

D.E.S.C., Lille, France, 1972
M.B.A., The University of British Columbia, 1974.

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in the

FACULTY OF GRADUATE STUDIES

Faculty of Commerce and Administration

We accept this thesis as conforming
to the required standard

The University of British Columbia
December, 1977

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of COMMERCE

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date 16/12/1977

ABSTRACT

The purpose of this research is to analyse, discuss and extend the analytical methodology associated with the study of consumer brand-switching behavior. As such, it attempts to add to the existing understanding of the structure of the consumer brand choice process.

Rational human behavior may be viewed as a succession of choices made among more or less well defined alternatives. The problem we analyse in this study is how to predict these choices when the alternatives are fixed in advance. The alternatives considered in this study are low-cost, frequently purchased, brand identified consumer products. The unit of analysis is the individual consumer.

Stochastic models of brand choice are developed and used as constructs for organizing and interpreting brand choice data. These models are subsequently used to test specific hypotheses about brand loyalty (the tendancy for consumers to hold a favorable attitude toward - and concentrate their purchases on - a particular brand) and brand-switching (the tendancy for consumers to purchase more than one brand over a period of

time). In this respect, this dissertation follows the framework of earlier brand choice studies.

In many dimensions, however, this research is significantly different from most stochastic models of brand-switching behavior developed in the past. First, this research deals essentially with multi-brand switching behavior as opposed to mere brand loyalty. By collapsing the market into an artificial two-brand market (to achieve mathematical tractability), earlier researchers were forced to concentrate on repeat purchase behavior only. All the information about brand switching activity was lost in the aggregation process. In today's differentiated markets, the competition has to be monitored on a brand-by-brand basis, and this is best achieved through the use of models that deal explicitly with multibrand switching, such as the one developed in this study.

Second, this research views consumer brand choice behavior as both a cognitive and a stochastic process. A multi-dimensionally scaled configuration is used as a specification of consumers' cognitive structures. Perceptual distances derived from this configuration are then related to brand choice and brand-switching probabilities through a model that takes into account the constraints imposed on the

various probabilities.

The empirical results demonstrate that perceptions, preferences and cognitive structures are indeed significant determinants of consumer brand-switching behavior.

TABLE OF CONTENTS

Page

LIST OF TABLES

PAGE

LIST OF FIGURES.

ACKNOWLEDGMENTS

It is a pleasure to express my gratitude to all the people and organizations who directly or indirectly made it possible for me to eventually write these words.

My best thanks go to Professor Doyle L. Weiss. As my dissertation chairman, he is to be credited for all those "right" things that happened at the "right" time. His unfailing optimism, technical expertise and broad vision contributed much to this particular piece of research as well as to my understanding of research in general. As a friend, he helped me step away from some of the many falling rocks that have been aimed at generations of doctoral students' heads.

My thesis' committee members made many valuable suggestions towards the overall improvement of this study. Professor J. Claxton oversaw its marketing content while Professors M. Puterman, M. Schulzer and W. Ziemba made sure the mathematics were honest.

I am also greatly indebted to Professor F.M. Bass, of Purdue University, who graciously provided me with the

soft drink experimental data upon which most of the empirical results presented in this study are based.

I certainly wish to acknowledge here the generosity of the following organizations that allowed me to start and complete my doctoral studies. They include:

The Canada Council
The Killam Foundation
The University of British Columbia
The Oppenheimer Bros. & Company
The Provincial Government of British Columbia

To all of them, I express my gratitude.

Several individuals deserve a special mention for bringing fun and excitement into my student life. Since they are too numerous to be cited here, I will just single out one of them, Martin Kusy, whose ability to enjoy life and willingness to share it was definitely an asset for all his friends.

Finally, I owe a special debt of thanks to my wife, Dominique, who studied along with me at U.B.C. and who probably knows more about this dissertation than anybody else. Her presence, and that of my two children, Timothée and Stéphanie (who by the way dearly resent the abrupt cut-off of their daily ration of junk computer

print-outs which the completion of this research has caused) are gratefully acknowledged.

TO DOMINIQUE

Foreword

"I perceive now that the real charm of the intellectual life - the life devoted to erudition, to scientific research,to philosophy, to aesthetics, to criticism - is its easiness. It's the substitution of simple intellectual schemata for the complexities of reality, of still and formal death for the bewildering movements of life."

Aldous Huxley, Point Counterpoint.

" J'en parlerai à mon cheval "

Anonyme.

INTRODUCTION

The purpose of this research is to analyse, discuss and extend the analytical methodology associated with the study of brand-switching data. As such, it attempts to add to the existing understanding of the structure of the consumer choice process.

Rational human behavior may be viewed as a succession of choices made among more or less well defined alternatives. The problem we analyse in this paper is how to predict these choices when the alternatives are fixed in advance. The alternatives considered in this study are low-cost, frequently purchased, brand-identified consumer products. The unit of analysis is the individual consumer. The data are made up of the purchase history of each individual over time. For any two successive purchase occasions, consumers are said to make a repeat-purchase if the same brand was purchased on both occasions. Similarly, consumers are said to switch brands if two different brands were purchased on the two purchase occasions. Brand-switching data are the collections of such purchase histories considering all possible brands and purchase occasions.

The objective of this research is to develop stochastic brand-switching models to be used as constructs for organizing and interpreting brand-switching data. In this respect, this dissertation follows the framework of brand choice analyses originated by a series of articles written by Brown [1952] and Cunningham [1956] which set the pace of the subsequent consumer choice stochastic modeling activity of the last decade. This activity was marked by the works of Kuehn [1962 - 65], Haines [1964], Massy [1965 - 66 - 67], Montgomery [1966 - 7 - 9], Morrison [1965 - 6], Ehrenberg [1959 - 65], Jones [1970 - 1 - 3]), Herniter [1973] and Bass [1974 - 6].

In many dimensions, however, this research is significantly different from most stochastic models of brand-switching developed in the past. First, in contrast to the approach of testing a-priori theories of consumer choice behavior, the method in this study is that of description leading to generalization (see e.g., Simon [1968] and Ehrenberg [1972]). To the extent that the descriptions generalize across circumstances and product classes, knowledge is gained about the existence and structure of underlying choice relationships.

Chapter II provides an example of such an exercice in data analysis. In that chapter, a set of consumer brand-switching data are submitted to a purely statistical analysis in an attempt to uncover possible "regularities" in the data. These "regularities" are then exploited in chapter III to construct a brand choice model based on the results of the preliminary statistical analysis.

Second, this research deals essentially with multi-brand switching behavior as opposed to mere brand loyalty. Herniter [1973] and Bass [1974] notwithstanding, all of the authors mentioned above developed models of brand choice behavior rather than models of brand-switching behavior. By collapsing the market into an artificial two-brand market (to achieve mathematical tractability), they were forced to concentrate on repeat purchase behavior only. All the information about brand-switching activity was lost in the aggregation process. Purchases of brands other than the brand under study were treated as being purchases of "The Competitor's" brand without consideration for the varying degree of competition between the various brands. In today's differentiated markets, the brand manager can no longer view his competitors' brands as being equally threatening to his particular brand. The competition has to be monitored on a brand-by-brand basis, and this is

best achieved through the use of models that deal explicitly with multi-brand switching such as the one developed in this study.

Last, while most existing brand choice models treat consumer choice behavior as being completely stochastic or entirely deterministic, this research views it as both a stochastic and a cognitive process. Stochastic, since brand selection on a given trial cannot be predicted precisely; and cognitive because the steady state choice probabilities observed over a sequence of choices reveal a choice pattern consistent with the consumer perceptions, preferences and beliefs toward a particular set of brands. To acknowledge both the stochastic and deterministic features of the brand-switching phenomenon, this study introduces a class of models which implies aggregate brand-switching and repeat purchase probabilities. In addition, it also directly incorporates the impact of a-priori relevant exogenous variables into its structure. This approach, while preserving the important features associated with stochastic brand choice models from the past, allows researchers (at least those with more faith in human rationality) to include in their models variables of behavioral and managerial significance.

The research reported here is presented in seven chapters. Chapter I relates this research to the relevant literature and illustrates the multidisciplinary nature of brand-switching modeling activity. The next two chapters are devoted to the analysis of consumer brand loyalty (as opposed to consumer multi-brand switching behavior). In chapter II, we perform a statistical analysis of consumer brand choice data in an attempt to discover possible "regularities" in the data. In chapter III, we construct a stochastic model of brand choice based on the "regularities" uncovered by the statistical analysis of the preceding chapter.

The major analytical work on which this dissertation rests is contained in chapters IV to VI. Chapter IV lays the theoretical base which guides the development and empirical investigation of the brand-switching models offered in the following two chapters. Chapter IV is included to familiarize the reader with the probabilistic concepts and the model-building strategy that underlie the development of the operational formulations presented in the empirical chapters V and VI.

In chapter V, a joint space theory of brand-choice is offered to analyse consumer brand-switching behavior. The central concept underlying this theory is

that of cognitive consistency. The chief hypothesis holds that consumers strive to maintain an equilibrium between their perceptions and preferences of the brands, on the one hand, and their actual brand choice on the other hand. This hypothesis is empirically tested with the help of the mathematical tools developed in chapter IV.

Chapter VI discusses a procedure for building a perceptual map of a market based on actual choice data (brand-switching probabilities) rather than the more widely used (see e.g., Green [1975] and Bouroche [1977]) preference and similarity data.

Finally, chapter VII contains a summary and an evaluation of the dissertation findings.

CHAPTER I

THE EFFORT OF THE PAST: BRAND CHOICE STUDIES

An overview of analytic approaches employed in the study of consumer choice behavior necessarily reflects the diversity of philosophies of the authors of the large and growing body of literature in this field. Since the focus here is on analytical approaches, the literature cited will be selective. While the number of published articles, combined with the diversity of approaches to the study of brand choice render any classification attempt of the literature somewhat self-defeating, some kind of classification is necessary. To this end, the following scheme has been retained for its simplicity. Past research will be segmented according to its basic philosophy, i.e. whether it views consumer brand choice as a deterministic or a stochastic process. Deterministic approaches will be further differentiated on the basis of their position along the brand-specific versus person-specific continuum.

Brand choice can be symbolically represented by the following relation:

$$P_{ik} = f_{ik} (X_k, Y_i)$$

where

P_{ik} = preference of individual k towards brand i or probability that individual k chooses brand i

X_k = a vector of household-related variables

Y_i = a vector of brand-related variables

f_{ik} = some mathematical function.

The two most crucial problems facing the marketing researchers are to specify:

i) the variables to be included in the X and Y vectors

ii) the form of f_{ik}.

The following discussion will be wholly concerned with the first issue. As will be seen later, most theories of consumer brand choice behavior can be classified according to their handling of this fundamental problem.

1.1 Deterministic Approaches to Brand Choice Behavior

1.1.1 Socio-economic and personality studies of brand choice

The most fundamental question that can be asked about consumer choice behavior is whether that behavior is at least partially stochastic or whether there exists causes and explanation for all such behavior.

Many behavioral scientists since Freud have believed that there exists an explanation for all human behavior even if the explanation must be sought in the unconscious. This belief has shaped much of the early research concerned with finding the so-called brand choice correlates and the demographic profile is probably the most familiar result of this effort. Examples of attempts to link the various components (e.g., age, income, education) of the demographic profile to consumer brand choice behavior are given by the studies of Frank et al. [1965,7,8,9]. In general, demographic profile analysis has had limited success in explaining individual choice behavior. As a result, researchers have turned to other explanations of brand choice behavior based on non-demographic consumer characteristics such as socio-economic and personality factors.

As with demographic profile analysis, the new stream of research met with mixed success. Some studies (see e.g., Day [1969] and Carman [1970]) did find some relationships between brand loyalty and certain consumer characteristics. For instance, Day [1969] found the brand loyal consumer to be very conscious of the need to economize when buying, confident in her judgment, older and living in a smaller than average household. Other studies

found that high brand loyal households apparently have a profile of personality and socio-economic characteristics that is virtually identical of that of households exhibiting a lower degree of brand loyalty (see e.g., Frank et al. [1969]). Because of their often contradictory results, it is fair to say that these and other deterministic studies of individual consumer choice behavior have consistently failed to explain a substantial portion of the variance in the dependent variable.

The inconclusive or even contradictory results of this research are due in part to the absence of a widely accepted research tradition. This makes comparisons between studies difficult as different researchers use different definitions, concepts and methodologies. The failure of brand choice correlates to explain choice behavior has led researchers to investigate the relationships between brand choice and certain market characteristics, such as the availibility of brands, price fluctuations and dealing activity (see e.g., Massy & Frank [1965], Farley & Ring [1970]). In this research the focus shifts from the X_k variables (person specific) to the Y_i variables (brand specific). This approach has been welcomed by the business community because it relates consumer decision making to variables that are controllable by the firm.

The importance of these variable types was demonstrated by Farley's study [1964] which found brand loyalty to be influenced by price, distribution and promotional activities. Based on this finding, Farley concluded that much of the apparent differences in brand choice behavior across product classes could be explained on the basis of structural variables describing the market in which the product is sold. Farley's results were subsequently strongly challenged by Massy and Frank [1965] and Anderson [1966] so that once again no invariant generalizations can be made.

1.1.2 Multi-Attribute Attitude Theory

A large number of recently published articles in the marketing literature have extended the attitude theory concepts developed in social psychology to the study of brand preference and brand choice (see e.g., Lehmann [1971], Pessemier et al. [1971,2,2a], Bass et al. [1972] and Wilkie & Pessemier [1973]). There is no intent here to discuss the details of attitude theory studies in marketing[1]. It will be useful, however, to mention briefly the underlying nature and basic structure of those studies.

1 For an excellent survey of psychological theories of consumer choice, see Hansen [1976].

The central proposition in attitude theory is that attitudes are composed of beliefs about the attributes of objects and the evaluative aspects of these beliefs. Thus, for example, if one were measuring attitudes towards different brands of toothpaste, one would determine the beliefs which consumers have about the extent to which each of the several brands possess properties such as decay prevention, teeth whitening, taste and breath control. One would also determine the importance which consumers attach to each property for brand choice purposes.

The rationale for using multi-attribute attitude (MAA) models to study consumer decision making can be traced to the work of Rosenberg [1960], Fishbein [1965] and Lancaster [1966]. The first two offered a psychological interpretation of the MAA approach while the latter provided an economic foundation with his suggestion that "people buy not products but bundles of attributes that meet their needs".

In general, it is established that the MAA models are good predictors of overall evaluation or attitude, whereas their ability to predict brand choice behavior is more varied (see e.g., Wilkie & Pessemier [1973]). The data shows that even when stated preferences are unchanging

consumer brand choice does change (see e.g., Bass [1974]). Thus, stated preference is not a good predictor of choice for a single choice occasion and most importantly a basic premise of much behavioral science theory is brought into question. This conclusion has led several researchers to develop and test stochastic models of consumer brand choice to which we now turn.

1.2 Stochastic Models of Consumer Choice Behavior

1.2.1 General discussion

When hope of understanding the motivations and subsequent actions of consumers is gone, researchers turn to stochastic models of purchase behavior. The brand-switching literature of the past decade is rich with such fruits of theoretical despair. Typically, stochastic models take the probabilistic nature of consumer brand choice as given and make little or no attempt to model the underlying mechanisms of individual brand choice behavior. As a result, they predict but do not explain the differing brand purchase probabilities which exist between individuals. This is a radical departure from traditional deterministic theories that purport to explain an individual's brand preferences in terms of demographic or psychologic characteristics of the individual, or in terms

of his beliefs and/or attitudes concerning the attributes of available brands and some sought-after "ideal" level of these attributes.

The underlying rationale behind the previously discussed "traditional" theories is that behavior is deterministic. Thus, if object A is preferred to object B, then, other things being equal, object A will be chosen, and unless preferences change, choice will not change. Any other pattern would suggest irrationality. Proponents of the stochastic school have challenged this premise. They argue that there exists a substantial stochastic component in brand choice behavior, and as a result it is no more possible to provide an explanation for that component than it is to provide an explanation for the outcome of the toss of a coin.

Massy et al. [1970] have provided a detailed review of the issues and the structure of stochastic models developed before 1970. Their classification of the various models, reproduced in table I.1, has been updated to incorporate the recent development in the field. The table classifies the models according to four characteristics:

(i) time effects: whether the model deals with discrete time (purchase occasion) or continuous time;

(ii) purchase event feedback: whether the model assumes that the act of purchasing and using a product has a direct effect on the household's subsequent purchase probabilities;

(iii) population heterogeneity: whether the model allows for the fact that consumers differ from one another in many ways (in terms of demographic and socio-economic factors, awareness, attitudes etc...);

(iv) number of alternatives considered: whether the model can handle multi-brand markets or collapses the market into an artificial two-brand market.

As can be seen from Table I.1, the early Markov and Linear Learning models incorporate assumptions about the effects of purchase-event feedback on brand choice, but do not make provisions for time effects and population heterogeneity. The assumption of population homogeneity has later been relaxed by Morrison [1965a,b] and Massy [1965] who developed "compound" versions of the Bernouilli, Markov and Linear Learning models respectively. The term "compound" denotes the fact that an explicit provision for a probability distribution of relevant parameter values is included in the model.

TABLE I.1

CLASSIFICATION OF STOCHASTIC BRAND CHOICE MODELS

CHARACTERISTICS	No Time Effects		Time Effects	
Two-alternative models	No purchase event feedback	Purchase event feedback	No purchase event feedback	Purchase event feedback
Homogeneous population	Bernouilli	Markov Lipstein [1959] Linear Learning Kuehn [1958]	none	Semi-Markov Howard [1963] Variable Markov Telser [1963] Dynamic Markov Lipstein [1965]
Heterogeneous population	Household-Bernouilli Frank [1962] Compound Bernouilli Morrison [1965a,b] Aaker's New Trier [1971]	Household-Markov Massy [1966] Composite model Jones [1973]	Response Uncertainty Coleman [1964] Dynamic Inference Howard [1965] Probability Diff. Montgomery [1966] Semi-Markov Herniter [1971]	Household variable Markov Duhamel [1966] Variable Learning Kuehn et al. [1967] Learning-Diff. Jones [1969-71]
Multialternative models Heterogeneous population	Hendry model [1966] Entropy model Herniter [1973] Bass PNBD [1974-6]		Markovian Policy model Leeflang [1974]	

Basic contributions to two-alternative brand choice models combining time effects and population heterogeneity were made by Coleman [1964], Howard [1965] and Montgomery [1966]. In the work of Coleman, households' purchase probabilities are initially distributed over members of the population, and are then changed according to a time trend function. In contrast, Howard's work assumes a household to draw its purchase probability from a distribution not once, as in the Coleman model, but again and again at randomly distributed points in time. Montgomery's Probability Diffusion model [1966] is an extension of Coleman's model. Herniter's semi-Markov model [1971] extends Howard's and Mongomery's work and accounts for both purchase timing and brand selection. A first order Markov process is used by Herniter to describe brand selection and Erlang density functions are used to describe time between purchases.

Models including time effects and purchase event feedback but not population heterogeneity have also been developed (see e.g., Howard [1969], Telser [1963] and Lipstein [1965]. Relatively little has been published on brand choice models including all four of the characteristics shown in Table I.1. Exceptions are Duhamel [1966] who estimated Telser's Variable Markov model using

household data, and Jones [1969-70] who extended Montgomery's Probility Diffusion model to include learning characteristics.

The recent trend is toward building "composite" models which extend the notion of consumer heterogeneity to allow for the fact that different consumers may obey different mechanisms of behavior. Previously proposed brand choice models have always assumed that each consumer in the population obeys the same underlying mechanism. In this context, Jones [1973] has discussed a model which does not assume a single behavioral mechanism for the entire population, but is in fact a composite model which allows each consumer to obey one of three mechanisms: Bernouilli, Markov and Linear Learning. Jones provides for two types of heterogeneity in the composite model. The first type involves differences between consumers who obey the same behavioral mechanism (parameter heterogeneity) while the second type distinguishes the composite model from its components (model heterogeneity). This work, while significant, has not been empirically tested.

An approach similar to Jones has been suggested by Blattberg and Sen [1974-76] who described a Bayesian discrimination procedure which determines for each consumer the stochastic model of brand choice best supported

by his past purchasing behavior. Blattberg and Sen claim that a market segmentation strategy based on the consumer behavior mechanism provides better information to the marketing decision maker. Their work marks the re-emphasis of stochastic based research away from model-fitting and back to the more interesting diagnostic potential, i.e. the extent to which the research can contribute to a better understanding of brand choice behavior.

Meanwhile, zero order models of brand choice are again become popular, especially in the context of brand-switching analysis. Herniter [1973] and Bass [1974] have employed zero-order models in a heterogeneous population for the study of brand-switching on adjacent purchase occasions. The distinguishing feature of these models rests on their ability to deal with multi-brand markets. Previously, all early brand choice models (such as the ones in the upper half of Table I.1) had to combine the market into a two-brand market, i.e. the "favorite" brand plus an "all-other" brand.

Herniter's Entropy model [1973] deserves our attention if only because it is the first multi-brand switching model to be published in the literature[2].

2 Actually, the Hendry model antedates the Entropy model, but details of the model's derivation are not available.

Although the Entropy model is based on sound mathematical and behavioral assumptions, its complexity becomes unwieldy as the number of brands goes past four. Moreover, the Entropy model is a normative model and produces the brand-switching matrix as a function of market shares only. If there does not exist a one-to-one correspondence between a set of market shares and the associated brand-switching matrix, the model may not consistently yield good predictions. Herniter, however, has supplied some empirical evidence that supports the model in one such case - that of an equilibrium market.

The major weakness of the maximum entropy approach to the estimation of brand-switching probabilities, as used by Herniter, is its inflexibility. The entropy estimates of brand-switching depend only on the market shares and are therefore, independent of product category. This has led Bass [1974] to develop a model which can accomodate a much richer range of brand-switching situations.

1.3 In Summary

After this brief "tour d'horizon" of the various approaches to stochastic and deterministic brand choice modeling, several comments are in order. First, most stochastic brand choice models are at the aggregate

level, even though they often employ a parameter to represent heterogeneity across the population. Second, since stochastic models rarely contain marketing variables (for some exceptions, see e.g., Haines [1969], Kuehn and Rohloff [1967]; Lilien [1974] and Nakanishi [1973]) using them at the aggregate level has offered marketing managers little insight into their problems. Third, both approaches to consumer brand choice modeling have weaknesses and advantages. The deterministic approaches (demographics, psychographics, and more recently, the multi-attribute attitude theory) are rich in their implications for marketing strategy but usually lack strong empirical support. With stochastic brand choice models, exactly the opposite circumstances occur. The models exhibit strong empirical support but offer little in the way of marketing action and/or policy implications. In contrast, the model building approach presented in chapter IV provides a mean of incorporating the influence of both behavioral and marketing variables, and represents a viable alternative to the existing brand-switching models. In preparation for this alternative, the next chapter is wholly concerned with a statistical analysis of consumer brand choice data.

CHAPTER II

PRELIMINARY STATISTICAL ANALYSIS OF CONSUMER BRAND-SWITCHING DATA

In contrast to the approach of testing a priori theories of consumer choice behavior, the method in this chapter is that of description leading to generalization. To the extent that the descriptions generalize across circumstances and product classes, knowledge is gained about the existence of relationhips if not about the reasons for the observation. Following Ehrenberg (1972), we shall search for "regularities" in consumer brand-switching data. These regularities, to the extent that they extend beyond the data from which they were generated can lead to the construction of explanatory theories from which the generalizations can be derived.

Our methodology consists of two steps:

1. In this chapter, we shall perform a purely statistical analysis of consumer brand-switching data in an attempt to uncover possible "regularities" in the data.

2. In the next chapter, we shall construct stochastic models of brand choice based on the "regularities" uncovered by the preceding statistical analysis.

To this end, we have organized the following discussion around eight major sections:

- Description of consumer brand-switching data.
- Choice of a statistical technique.
- The log-linear model.
- Model fitting and hypothesis testing procedures
- Presentation of the empirical results
- Comparison of brand-switching behavior across segments.
- Summary of the results and conclusion.

2.1 Description of consumer brand-switching data.

Consumer brand-switching data denote the collection of individuals' purchase histories aggregated over all possible brands and purchase occasions. Typically, the purchase behavior of a sample of N individuals is observed over a given time period T.

For a given product class, each individual's purchase history can be represented by a string of 1's and 0's. A "1" represents a purchase of the brand under

consideration while a "O" represents a purchase of any other brand[1].

Assume that data are available for five consecutive purchases for each of the N individuals in the panel. The N individuals can be segmented by these five purchases into 2^5 or 32 mutually exclusive and collectively exhaustive categories. The panel data can now be represented in terms of N_{ijklm}, the number of individuals who purchased brand i, j, k, l and m respectively on five successive purchase occasions, where i, j, k, l, m = 1 or 0. The corresponding proportion P_{ijklm} is just the ratio of N_{ijklm} to the number of individuals in the panel i.e.:

$$P_{ijklm} = N_{ijklm}/N. \tag{2.1}$$

We will also pay particular attention to the proportion of individuals who purchased brand m on the fifth trial given that they purchased brand i, j, k and l on the first four purchase occasions, where again i, j, k, l, m = 1 or 0. This proportion, which we shall write $P_{m/ijkl}$ is by definition equal to:

---

1 In this chapter, we will limit ourselves to a two-brand market: brand 1, the brand under study and brand 0, an all other category. The brand under study is often the individual's favorite brand. Models for multi-brand markets will be introduced in chapters IV.

$$P_{m/ijkl} = \frac{P_{ijklm}}{P_{ijkl0} + P_{ijkl1}} \tag{2.2}$$

The P_{ijklm}'s represent a five-way contingency table. This multidimentional contingency table describes the joint distribution of five qualitative dichotomous variables. Qualitative, since the variables are nominal, and dichotomous because we are dealing with a two-brand market. These five binary variables represent the purchase outcome observed for the corresponding five purchase occasions. That is, variable 1 stands for the purchase outcome at purchase occasion 1. This variable takes on a value of 1 or zero depending on which of the two brands was purchased. The remaining variables are similarly defined.

For illustration purposes, Table II.9 displays a five-way contingency table related to coffee consumption of a sample of 538 families from the Chicago Tribune Consumer Panel[2]. As can be noted from the table, 37.06% (1968/5,310) of the individuals purchased brand 1 on all five purchase occasions. Of those who bought brand 1 on the first four purchase occasions, 88.90% (1968/(1968 + 245)) purchased it again on the fifth occasion.

---

2 These and other related data will be described in a later section.

Example of the five-way contingency table: Brand-switching data from the Chicago Tribune Consumer Panel data.#

Purchase outcome at occasion																	
t-4					1								0				
t-3			1				0				1				0		
t-2		1		0		1		0		1		0		1		0	
t-1	1		0		1		0		1		0		1		0		1
t 1	1968	203	196 *	90	222	67	75	64	276	79	84	56	128	62	105	108	
t 0	245	132	84	94	73	72	49	110	93	55	62	90	59	68	85	156	

Sample size: 5310

* Read : 196 individuals purchased brands 1,1,0,1 and 1 at purchase occasions t-4,t-3, t-2,t-1 and t respectively.

\# These data are described at length in section 2.5

TABLE II.9

The corresponding proportion for those who bought brand 0 on the first four purchase occasions shrinks to 40.9%.

As a rule, we observe that the proportion of families who bought brand 1 given their past purchase history ($P_{1/ijkl}$) increases with both the number of past purchases of brand 1 and the recency of these purchases. For instance, we observe that $P_{1/1111} > P_{1/0000}$ and that $P_{1/0001} > P_{1/1000}$, an indication that both the number and position of the 1's in the purchase string affect the observed proportions.

It is the purpose of this chapter to formally estimate the influence, if any, of past purchase outcomes on future purchase decisions from brand-switching data such as the ones just described. To this end, we shall first need to decide on an appropriate statistical technique. Hypothesis testing, estimation procedures and the empirical data will then be presented and the results discussed.

2.2 Choice of a statistical technique.

The restricted range of the criterion (proportion) and the qualitative nature of the predictors violate crucial assumptions of most traditional statistical techniques. Because the criterion variable is a proportion,

application of dummy-variable regression, Automatic Interaction Detector or Anova would entail three limitations:

i) the criterion variable (proportion buying brand 1 given past purchase history) is not normally distributed, since it is sum-constrained;

ii) the criterion variable does not show constant variance (homoscedasticity) across variations in the predictors (dummy-coded variables in this case);

iii) the model's prediction could fall outside the range at 0 to 1.

One way out of these difficulties is to apply a transformation to the proportion data, whose value are constrained to range from 0 to 1, and then to use an estimation procedure that takes unequal variance error into account. The logistic transformation is one such transformation and the logit model provides a useful and natural representation of the data (see e.g., Berkson [1944], Bishop [1969]). So is the log-linear model of multidimensional contingency table which we shall use in this study to detect meaningful patterns in the brand-

switching data to be analysed[3].

2.3 The log-linear model.

Following Birch [1969] and Goodman ([1968], [1969], [1970], [1971], [1972]), we can express the logarithm of the observed proportions P_{ijklm} into main effects (functions of a single subscript) and interaction effects (function of two or more subscripts). The latter are, in turn, distinguished as bivariate interactions, trivariate interactions and so forth. The (full) log-linear model by analogy with Anova models is defined as:

$$
\begin{aligned}
(2.3)\quad \log(P_{ijklm}) = {} & \theta + \lambda_i^B + \lambda_j^C + \lambda_k^D + \lambda_l^E + \lambda_m^F \\
& + \lambda_{ij}^{BC} + \lambda_{ik}^{BD} + \lambda_{il}^{BE} + \lambda_{im}^{BF} + \lambda_{jk}^{CD} + \lambda_{jl}^{CE} + \lambda_{jm}^{CF} + \lambda_{kl}^{DE} \\
& + \lambda_{km}^{DF} + \lambda_{lm}^{EF} + \lambda_{ijk}^{BCD} + \lambda_{ijl}^{BCE} + \lambda_{ijm}^{BCF} + \lambda_{ikl}^{BDE} + \lambda_{ikm}^{BDF} \\
& + \lambda_{ilm}^{BEF} + \lambda_{jkl}^{CDE} + \lambda_{jkm}^{CDF} + \lambda_{klm}^{DEF} + \lambda_{ilm}^{CEF} + \lambda_{ijkl}^{BCDE} + \lambda_{ijkm}^{BCDF} \\
& + \lambda_{jklm}^{CDEF} + \lambda_{iklm}^{BDEF} + \lambda_{ijlm}^{BCEF} + \lambda_{ijklm}^{BCDEF}
\end{aligned}
$$

---

3 The logit model is a particular case of the more general log-linear model. In the marketing context, Kuehn [1958] was the first to realize the diagnostic potential of the factorial analysis of variance technique to isolate the effects of purchase decisions on each trial in the past upon the probability of purchasing the favorite brand on the next trial. Of course, the present situation differs from the analysis of variance because we are dealing with a sample from a cross-classification rather than with independent observations from populations that are normally distributed and homoscedastic.

where for uniqueness[4] the λ's satisfy the usual kinds of conditions:

$$(2.4) \quad \sum_i \lambda_i^B = 0, \ldots\ldots, \sum_i \lambda_{ij}^{BC} = 0, \ldots\ldots, \sum_i \lambda_{ijk}^{BCD} =$$

$$\sum_j \lambda_{ijk}^{BCD} = \sum_k \lambda_{ijk}^{BCD} = 0, \ldots\ldots, \sum_i \lambda_{ijklm}^{BCDEF} =$$

$$\sum_m \lambda_{ijklm}^{BCDEF} = 0.$$

The λ's represent the possible "effects" of the five variables (purchase occasions) on P_{ijklm}: the main effects are $\lambda_i^B, \ldots, \lambda_m^F$; the remaining λ's represent interaction effects. The upperscript on λ stand for purchase occasions (B, C, D, E, F in order of increasing recency) while the subscripts denote brand choices. The number of superscripts (or subscripts) on a particular λ indicates the order of the interaction: one for main effects, two for bivariate interactions, etc... For instance, the parameter λ_{ikl}^{BDE} quantifies the effect on P_{ijklm} of a purchase of brand i at occasion t - 4, k at occasion t - 2, and l at occasion t - 1. The entire set of λ's must be estimated from the observed proportions.

---

4 See e.g., Goodman (1968).

In this form, the log-linear model is known as a saturated model because there is a parameter for each data point.

Making use of the set of constraints expressed in (2.2), the log-linear model can be rewritten as:

$$
\begin{aligned}
(2.5)\quad \text{Log}\,(P_{ijklm}) = {} & \theta + \lambda_1\delta_i + \lambda_2\delta_j + \lambda_3\delta_k + \lambda_4\delta_l + \lambda_5\delta_m \\
& + \lambda_{12}\delta_{ij} + \lambda_{13}\delta_{ik} + \lambda_{14}\delta_{il} + \lambda_{15}\delta_{im} + \lambda_{23}\delta_{jk} + \lambda_{24}\delta_{jl} \\
& + \lambda_{25}\delta_{jm} + \lambda_{34}\delta_{kl} + \lambda_{35}\delta_{km} + \lambda_{45}\delta_{lm} + \lambda_{123}\delta_{ijk} + \cdots \\
& + \lambda_{235}\delta_{jlm} + \lambda_{1234}\delta_{ijkl} + \cdots\cdots + \lambda_{1245}\delta_{ijlm} \\
& + \lambda_{12345}\delta_{ijklm}
\end{aligned}
$$

where

$$\delta_i = \begin{matrix} 1 \\ -1 \end{matrix} \text{ if } \begin{matrix} i = 1 \\ 0/w \end{matrix}$$

$$\delta_{ij} = \begin{matrix} 1 \\ -1 \end{matrix} \text{ if } \begin{matrix} i+j \\ 0/w \end{matrix} \text{ is even or zero}$$

$$\delta_{ijk} = \begin{matrix} 1 \\ -1 \end{matrix} \text{ if } \begin{matrix} i+j+k \\ 0/w \end{matrix} \text{ is odd}$$

$$\delta_{ijkl} = \begin{matrix} 1 \\ -1 \end{matrix} \text{ if } \begin{matrix} k+j+k+l \\ 0/w \end{matrix} \text{ is even or zero}$$

$$\delta_{ijklm} = \begin{matrix} 1 \\ -1 \end{matrix} \text{ if } \begin{matrix} i+j+k+l+m \\ 0/w \end{matrix} \text{ is odd}$$

If we further require equality between predicted and observed marginals, equation (2.5) reduces to:

(2.6) $$P_{ijklm} = a_i b_j c_k d_l e_m \, M_{it-4} \, M_{jt-3} \, M_{kt-2} \, M_{it-1} \, M_{mt} \, A_{ijklm}(\tilde{\theta})$$

where

$$a_i = \{ \sum_{j,k,l,m,} b_j c_k d_l e_m \, M_{jt-3} \, M_{kt-2} \, M_{lt-1} \, M_{mt} \, A_{ijklm}(\tilde{\theta}) \}^{-1}$$

$$b_j = \{ \sum_{i,k,l,m} a_i c_k d_l e_m \, M_{it-4} \, M_{kt-2} \, M_{lt-1} \, M_{mt} \, A_{ijklm}(\tilde{\theta}) \}^{-1}$$

(2.7) $$c_k = \{ \sum_{i,j,l,m} a_i b_j d_l e_m \, M_{it-4} \, M_{jt-3} \, M_{lt-1} \, M_{mt} \, A_{ijklm}(\tilde{\theta}) \}^{-1}$$

$$d_l = \{ \sum_{i,j,k,m} a_i b_j c_k e_m \, M_{it-4} \, M_{jt-3} \, M_{kt-2} \, M_{mt} \, A_{ijklm}(\theta) \}^{-1}$$

$$e_m = \{ \sum_{i,j,k,l} a_i b_j c_k d_l \, M_{it-4} \, M_{jt-3} \, M_{kt-2} \, A_{ijklm}(\theta) \}^{-1}$$

where

(2.8) $$M_{it-4} = \sum_{j,k,l,m} P_{ijklm}$$

$$M_{jt-3} = \sum_{i,k,l,m} P_{ijklm}$$

$$M_{kt-2} = \sum_{i,j,l,m} P_{ijklm}$$

$$M_{lt-1} = \sum_{k,j,k,m} P_{ijklm}$$

$$M_{mt} = \sum_{i,j,k,l} P_{ijklm}$$

and

$$\begin{aligned} \text{Log } (A_{ijklm}(\tilde{\theta})) = & \lambda_{12}\delta_{ij} + \lambda_{13}\delta_{ik} + \lambda_{14}\delta_{il} + \lambda_{15}\delta_{im} \\ & + \lambda_{23}\delta_{jk} + \lambda_{24}\delta_{jl} + \lambda_{25}\delta_{jm} + \lambda_{34}\delta_{kl} \\ & + \lambda_{35}\delta_{km} + \lambda_{45}\delta_{lm} + \lambda_{123}\delta_{ijk} + \ldots\ldots \\ & + \lambda_{235}\delta_{jlm} + \lambda_{1234}\delta_{ijkl} + \ldots\ldots\ldots \\ & + \lambda_{1245}\delta_{ijlm} + \lambda_{12345}\delta_{ijklm} \end{aligned}$$

The M's in equation (2.6) are the so-called marginals of the joint distribution P_{ijklm}. These marginals represent the proportion of consumers who purchased a given brand on a particular purchase occasion. The a's, b's, c's, d's and e's are just normalization constants that force equality between predicted and observed

marginals. Note that the parameter θ and the parameters for the main effects (λ_1 through λ_5) are no longer present in the final form of the log-linear model. Their effect on P_{ijklm} have been taken care of by the normalization constants.

It may be appropriate at this stage to remind the reader of the purpose of this chapter, lest he be overwhelmed by the messy notation of the log-linear model. The purpose of this chapter is to search for "regularities" in a set of brand-switching data to be described later. In particular, we would like to be able to quantify the nature and extent of the influence of past purchase outcomes on future ones.

Given the restricted range of the criterion and qualitative nature of the predictors, we were led to propose the log-linear model of multi-dimensional contingency table as the most appropriate model for the purpose at hand. The log-linear model first involves taking the natural logarithm of each P_{ijklm} and expressing it as a function of main and interaction effects. In doing so, we hope to detect meaningful patterns among the interactions that would either confirm or infirm whatever hypotheses marketing researchers generally hold with

respect to consumer brand-switching behavior[5]. Building on the results of this purely statistical analysis, we will then construct stochastic models of brand-switching behavior and test them with consumer panel data. We now return to the log-linear model to discuss alternative model fitting and hypothesis testing procedures.

2.4 Model Fitting and Hypothesis Testing Procedures.

Formulae (2.6 - 2.9) describe the "saturated" model in which all possible "effects" are included. In the saturated model, there are as many parameters as there are data points, and the data will be fitted perfectly. Hence, fitting the saturated model is of little value in itself. Rather, its value lies in pointing out possible clues to unsaturated models, where some of the λ's are set equal to zero, or equal to one another.

Since the five variables which constitute the five-way contingency table represent the purchase outcomes of individuals over five successive purchase occasions, they are closely related. This dependence should translate itself into a simpler structure in terms of inter-

---

5 Such as Kuehn's [1958] learning hypothesis or Frank's [1960] zero-order hypothesis.

action effects. That is, some interaction effects should either vanish or exhibit some clear pattern with one another.

Also, brand-switching data have already been analysed with the help of (stochastic) models that were parsimonious in terms of number of parameters. Kuehn's Linear Learning model [1958] and to a lesser extent, Morrison's Brand Loyal model [1970] appear to fit a variety of brand-switching data with fairly good accuracy, thus reinforcing our a-priori belief that it should be possible to "explain" the data to a satisfactory degree without including in the log-linear model an overwhelming number of parameters. In the remainder of this section, we shall briefly outline some estimation and testing procedures to estimate and select various "unsaturated" models.

2.4.1 Estimation procedure.

Two different sets of parameters need to be estimated:

i) the interaction parameters: the λ's;

ii) the normalization constants: $\{a_i\}, \ldots\ldots, \{e_m\}$.

Since the normalization constants are functions of the λ's [see (2.6) - (2.8)], all we need to do is to estimate the latter from which the former will follow.

However, given the non-linearity of the relationships between the normalization constants, iterative procedures must be resorted to.

The parameters were estimated by minimizing the usual chi-square goodness of fit statistic, which measures the agreement between the observed proportions P_{ijklm} in the contingency table and the corresponding estimate $\overline{P}_{ijklm}$ under a given hypothesis:

$$(2.10) \qquad X^2 = N \sum_{i,j,k,l,m} (P_{ijklm} - \overline{P}_{ijklm})^2 / \overline{P}_{ijklm}$$

Where N is the sample size upon which the proportions P_{ijklm} are based. For the saturated model (2.6), these minimum chi-square estimates are also maximum likelihood estimates, since the fit is perfect. For an alternative maximum likelihood procedure, see Goodman[6] [1971].

---

[6] For testing purposes, Goodman [1970] advocates the use of yet another chi-square statistic derived from the likelihood ratio criterion, and denoted by X^2_{LR} in this study. This statistic is defined as:

$$(2.11) \qquad X^2_{LR} = 2 \sum_{i,j,k,l,m.} N_{ijklm} \log (P_{ijklm}/\overline{P}_{ijklm})$$

where $\overline{P}_{ijklm}$ is the model predicted proportion. For convenience, the value of both chi-square statistics are reported in the empirical section.

2.4.2 Unsaturated model selection.

One (rather ad-hoc) way of selecting unsaturated model is to compute a t - value for each coefficient and to select as an initial unsaturated model those interactions (λ's) with significient t - values. However, we will often be concerned with a set of λ's rather than with a single λ.

The test which we shall use to discriminate between two different models makes use of the fact that one is a constrained version of the other (see Goodman [1970]). This will be true of all the models entertained below. For example, the model in which all the four-way and the five-way interactions are zero is a constrained version of the model in which only the four-way interactions are assumed to vanish. Both models are in turn "nested" in the general saturated model, i.e., they are constrained versions of it.

More generally, let Ω denote the set of λ's that are assumed zero under a given hypothesis H. E.g., if we assume that all of the four-way interactions are zero, then Ω consists of the five parameters appearing in (2.5). For a different model, say H', let Ω' denote the set of λ's that are assumed zero under H'. Consider the case where Ω includes the set in Ω' (in addition to

other sets), and let Ω^* denote the set of λ's that are included in Ω but not in Ω'. Let H^* denote the hypothesis that the λ's in Ω^* are zero, and let H^*/H' denote the hypothesis that H^* is true assuming that H' is true. (Note that H^*/H' is equivalent to the hypothesis that H is true assuming that H' is true). To test H, we calculate X^2 (H) by (2.11). To test H^*/H', we calculate:

$$(2.12)\quad X^2\ (H^*|H') = X^2\ (H) - X^2\ (H') = 2 \sum_{i,j,k,l,m} N_{ijklm} \log (P^{H'}_{ijklm} | P^{H}_{ijklm}),$$

which is the chi-square statistic based upon the likelihood ratio criterion for testing H assuming that H' is true. The asymptotic distribution of (2.12) is the chi-square distribution under $H^*|H'$, with degrees of freedom equal to the number of λ's in Ω^*. For a more complete description of these tests, see e.g., Goodman [1970].

We are now equipped to unravel some of the aspects of consumer brand-switching behavior that lay dormant in the panel data that are described below.

2.5 Description of the empirical data.

In the next sections, the models discussed in this chapter are applied to data obtained form the Chicago Tribune Consumer Panel. The product to be studied here is regular coffee. Coffee was chosen for two reasons. First, the required data were readily available in Massy et al. ([1970] pp. 126 - 128). And second, it is a product category that has been studied previously and thus tentative hypotheses on consumer behavior toward coffee have been formulated. These hypotheses will constitute our prior information regarding the coffee data and will guide our selection of the various versions of the model to be fitted to the data. The data and experimental design are best described in Massy et al.'s ([1970], pp. 128 - 130) own words:

> "The purchase decisions cover the period January 1956 through February 1949. There were 531 families who met the criterion of having at least 30 purchases of regular coffee during that three-year period. This criterion was set somewhat arbitrarily. However, this was about the minimum number of purchases needed to run the tests, and a higher cutoff would have adversely affected the total sample size.
>
> Each family's purchase history was reduced to a 0-1 process, where a 1 indicated the purchase of that family's favorite brand of coffee; a 0 indicated the purchase of any other brand of coffee. The 0-1 process was defined with 1 representing the family's favorite brand and 0 representing all other brands instead of a particular brand being defined as a 1 for all families for two reasons:
>
> 1. It was felt that examining families' behavior toward their favorite brand was the better way to investigate the complex phenomenon of "brand loyalty".

2. For any particular brand there exist a few families who devote virtually all their purchases to this brand, and a sizable portion of the sample that never buy the brand. These two groups of consumers make the beta distribution (a unimodal distribution) a very poor candidate for the heterogeneity in the population of consumers with respect to any particular brand. See Section 3.5.2.

The first 30 purchases of each family were recorded, and this sequence was then broken into two periods. Even if a family had more than 30 purchases, only the first 30 were used. This was to give equal weightings to the purchase histories of light as well as heavy buyers. Period 1 consisted of each family's first 15 purchases and period 2 the last 15 purchases. The brand purchased most often in period 1 was designated the favorite brand for that period, and similarly for period 2. Thus, it became possible to compare the behavior of "switchers" (those whose favorite brand in period 1 was different from their favorite brand in period 2) with the "loyal" consumers who kept the same favorite brand.

Each family yielded two observations, each observation containing a past history of length four and the purchase decision that followed that past history. These ten observations per family were:

	Past History of Length Four		Purchase Decision on the Next Trial
Purchases	4 through 7	and	8
	5 through 8	and	9
	6 through 9	and	10
	7 through 10	and	11
	8 through 11	and	12
	19 through 22	and	23
	20 through 23	and	24
	21 through 24	and	25
	22 through 25	and	26
	23 through 26	and	27

Purchases 14 through 18 were not used because the favorite brand may have switched from period 1 to period 2, thus, the 0-1 process could not be defined. Also purchases 1 through 3 and 28 through 30 were not used in order that the purchase decisions used would be roughly in the middle of the sequence that determined the favorite brand.

Strictly speaking, only two past histories should have been taken (e.g., 6 through 9 and 10 and 21 through 24 and 25). The "overlap" of the past histories makes the

past histories within a family dependent. However, when all the 531 x 10 = 5,310 observations are aggregated, this slight dependence is not harmful to the models. Also, two observations per family would not have been sufficient to test the models adequately.

The segments to be investigated and contrasted were defined as follows:

1.	ALL	All 531 families who met the selection criteria.
2.	100 PERCENTERS	The 142 families who bought only one brand of coffee during period 1 or period 2. (128 of them also had only one brand during both period 1 and period 2.)
	EXCEPT 100 PERCENTERS	The remaining 389 families who had one or more zeros.
3.	HEAVY	Those families who purchased more than the median amount (for the sample) of coffee. Consumption was measured in number of trips.
	LIGHT	The other half of the sample which purchased less than the median amount.
4.	LOYAL	The 360 families whose favorite brand in period 1 was the same as the favorite in period 2.
	NONLOYAL	The remaining 171 families who switched favorite brands.

This completes the description of the empirical data used in this chapter. The results of fitting various unsaturated models to these data are presented in the next sections and their implications for consumer brand switching discussed.

2.6 Presentation of the empirical results

The saturated models described by equations (2.6 - 2.9) were fitted to the data for the six segments described by Massy et al. [1970]. The estimation procedure outlined above was followed. Table[7] II.1 gives the value of each $\hat{\lambda}$, its standardized value $(\hat{\lambda}/S_{\hat{\lambda}})$, variance $(S^2_{\hat{\lambda}})$ and the sample size for each segment. The first column identifies the 26 interaction terms. Bivariate interactions are denoted by two subscripts, trivariate interactions by three subscripts and so on. Equation (2.9) together with the first column of Table IV.1 completely identify the parameters. For instance, the first row of the table exhibits for all segments the raw value of $\hat{\lambda}_{12}$ and its standardized value for the parameter $\hat{\lambda}_{12}$ which denotes the influence of the purchase outcomes of the first two purchase occasions on P_{ijklm}. The estimate of λ_{12} was .227 for the ALL segment and .234 for the LOYAL segment.

We shall now look for the presence of interesting patterns among the parameters within each segment and check whether they persist for all segments. In so doing, we shall only consider those patterns that yield

---

7 The tables have been collected at the end of the chapter.

themselves to meaningful interpretation in terms of behavioral significance. A pattern that does not replicate itself for different sets of data or cannot be readily interpreted in terms of consumer habitual brand switching behavior is of limited interest. With this objective in mind, let us examine the results exhibited in Table II.1.

As is often the case in contingency table analysis, the importance of the interaction effects decreases as their order increases. That is, bivariate interactions tend to explain a greater proportion of the variance observed in the P_{ijklm} than do the tri-variate interactions which in turn explain more of the variance than the four-way interactions and so on. While this pattern holds true in general for the coffee data, there are noteworthy exceptions. For example, the five-way interaction proved to exert more influence on P_{ijklm} than all of the four-way and half of the three-way interactions, as evidenced by the absolute values of the λ's, for three segments: ALL, HEAVY and LIGHT.

Considering the absolute values of the λ's one at a time, we see that for all the segments, most of the bivariate interactions are significant at the .05 level or less ($\hat{\lambda}/S_{\hat{\lambda}} \geq Z_{.05} = 1.645$) whereas none of the higher-

order interaction terms is significant at this level except for the LOYAL segment where four such interactions proved to be of significant magnitude. Some caution is required in the use of these λ's as a simple guide in selecting which hypotheses to fit the data, since such guidance is not always foolproof. As will be illustrated later, it would not be correct to make inferences about the statistical significance of a particular set of λ's just on the basis of their individual standardized values. For example, one cannot state that the whole set of three-way interaction terms is not statistically significant even though each one, taken separately, does not reach significance. The standardized value of a particular λ gives us a rough indication of whether this parameter makes a significant contribution toward explaining the variance observed in the P_{ijklm}, given that all of the remaining parameters are included in the model. Thus, a parameter that is not significant in the saturated model may well turn out to be so in some unsaturated models, i.e., models where some of the parameters are assumed to be zero.

We shall first assume that all interactions of a specified order or higher order (e.g., the 5-way, all 4-way and the 5-way, all 3, 4 and 5-way interactions) are

zero[8]. Support for this set of hypotheses can be found in Kuehn [1958] whose factorial analysis of variance design led him to conclude that for his data, all interaction terms were negligible. As mentioned above, the results exhibited in Table II.1 support the hypothesis that all interactions beyond the second-order are negligible when tested one at a time. Does this conclusion still hold when they are all simultaneously assumed to be zero? To answer this question, the previously described chi-square tests were applied to the following models:

Models	Interaction terms set equal to zero
M_{234}	5-way
M_{23}	5-way and 4-way
M_2	5-way, 4-way and 3-way

For example, M_{23} denotes the model in which the 5-way and 4-way interactions are set equal to zero. Thus, for this model, the only interactions that must be

---

8 Goodman [1971] has described stepwise procedures to select unsaturated models that fit the data, using methods that are, in part, somewhat analogous to the usual stepwise procedures in regression analysis. Such procedures will not be followed here as they may not necessary lead to easily interpretable patterns.

estimated from the data are the two and three-way interactions. Tables II.2 and II.3 exhibit the parameter estimates for models M_{234} and M_{23} respectively, together with the statistics required to calculate the chi-square statistic described in (2.11). Table II.7 summarizes the chi-square statistics for all models and all segments, while table II.8 conveniently exhibits the X^2 $(H^*|H')$ values needed to test hypotheses about various nested models for all segments.

A glance at the first three rows of Table II.8 reveals that:

i) For all segments, the hypothesis of an insignificant five-way interaction could not be rejected.

ii) The hypothesis of insignificant four-way interactions could not be rejected but for the LIGHT segment.

iii) The hypothesis of insignificant three-way interactions (given that all four-way and five-way interactions are assumed to be negligible) was rejected for three segments: ALL, HEAVY and LOYAL.

Unfortunately, these results are too non-specific to be of much interest. They do not shed much light onto consumer brand-switching behavior beyond suggesting that

the behavior of LOYAL and HEAVY buyers tend to be more complex[9] than that of NON-LOYAL buyers: it takes more interaction parameters (λ's) into the model for the former group than for the latter to obtain a similar fit. In other words, past purchase outcomes exert a stronger influence on future purchase outcomes for LOYAL and HEAVY buyers than for the NON-LOYAL ones.

There are, however, more interesting patterns among the interactions that tend to support the learning hypothesis[10] of consumer brand switching advocated by Kuehn [1958].

A close look at the numerical estimates of all two-way interactions suggests the following pattern: the interaction parameters represented by (i,j), (j,k), (k,l) and (l,m) are of similar magnitude. This pattern persists for all segments. Similarly, the magnitude of the interaction parameters denoted by the pairs (i,k), (j,l) and (k,m) are also quite similar as are that of the parameters represented by (i,l) and (j,m). For convenience, Table II.4 displays the two-way interactions in a form that illuminates

---

9 When complexity is judged in terms of the number of statistically significant interaction effects.

10 The fundamental concept underlying the learning hypothesis is that of purchase event feedback. That is, the act of purchasing and using a particular brand

the patterns observed among the parameters. The matrix arrangement, reproduced below for the HEAVY segment, makes it clear what is actually happening:

HEAVY Segment

	j	k	l	m
i	.25	.21	.08	.07
j		.22	.18	.10
k			.24	.18
l				.26

Each entry in the above triangular matrix corresponds to a two-way interaction. E.g., the entry corresponding to the second row and third column gives the parameter estimate of the (j,l) interaction (in terms of our earlier notation). As can be seen from the matrix, the diagonal elements are strikingly similar. Taking advantage of this fact, the number of two-way interaction terms to be estimated can be reduced from 10 to 4 by equating the following parameters:

---

is assumed to affect the probability that this brand will be selected again.

$$(2.12) \qquad (i,j) = (j,k) = (k,l) = (l,m)$$
$$(i,k) = (j,l) = (k,m)$$
$$(i,l) = (j,m)$$

where it is understood that the equality (i,l) = (j,m) means that the interaction effects represented by the pairs (i,l) and (j,m) are assumed identical. Before we test the hypothesis expressed in (2.12), let us pause briefly and try to relate the finding to what we already know about brand-switching behavior. The question is: what do the interaction terms that were observed to be of similar magnitude share in common? E.g., what is the common "denominator" between the following four interactions: (i,j), (j,k), (k,l) and (l,m)?

Answer: They all represent the effect of the outcome of two adjacent (successive) purchases! That is, the interaction (i,j) denotes the effect of having purchased brand i at occasion t-4 and brand j at occasion t-3. Similarly, (j,k) denotes the effect of having purchased brand j at occasion t-3 and brand k at occasion t-2 and so on... Thus, the ten two-way interactions can be partitioned in four sets of interactions on the basis of their degree of "adjacentness". By "adjacentness" is meant the extent to which the interaction represents the effect of successive purchase occasions. An interaction will be said to be

1-adjacent if it represents the interaction effect due to the particular outcomes of two successive purchase occasions. The interactions denoted by (i,j), (j,k), (k,l) and (l,m) are all 1-adjacent. Similarly, the interactions denoted by (i,k), (j,l) and (k,m) are 2-adjacent since they represent the interaction effects due to the particular outcomes of purchases that are two occasions apart. We can now summarize our empirical observations by the following hypotheses:

H_1: Interaction effects with the same degree of "adjacentness" are equal (see 2.12).

H_2: The lower the degree of "adjacentness", the greater the interaction effect.

Hypothesis H_2 states that the joint influence of any two past purchases of brand 1 on the probability of buying that brand again varies inversely with the number of purchase occasions that elapsed between those two purchases. This hypothesis is consistent with Kuehn [1958]'s earlier finding that remote purchases exert less of an effect on the probability of buying the brand again than do more recent ones.

Hypothesis H_1 states that the influence of successive purchases of brand 1 on the probability of

buying brand 1 again is independent of their position in the purchase string.

To test H_1 and H_2, we estimate a constrained version of M_2 where interactions of same degree of "adjacentness" are forced to be numerically equivalent. Denoting this model by M_{2*}, we see from Table II.7 and II.8 that H_1 is strongly supported by the coffee data. Since model M_{2*} is a constrained version of M_2, the testing procedure described above applies. The increase in the chi-square statistic incurred by constraining interactions of same degree of adjacentness to be equal is marginal, given the difference in the number of parameters of the two models (10 for M_2 versus 4 for M_{2*}). E.g., the chi-square statistic went up from 112.41 to 119.39 and 22.10 to 24.66 for the ALL and NON-LOYAL segments respectively. From the fourth row of Table IV.8, we see that for all segments, the increase observed in the chi-square statistic is not statistically significant at all levels of significance considered.

Thus, on the basis of the coffee data, the "adjacentness" hypotheses H_1 and H_2 received strong empirical support. Hypothesis H_1 allows us to substantially reduce the number of parameters to be estimated without harming the fit of the model.

However, the model does not yet provide an adequate representation of the data, as evidenced by the magnitude of the chi-square goodness of fit statistics. To improve the fit, we shall have to include in the model, interactions of order higher than two. Consider the ten three-way interactions. Once again, we can partition them on the basis of their degree of adjacentness as follows:

Degree of adjacent-ness	interactions	Examples of parameter estimates					
		EXCEPT 100 P.			HEAVY		
1	(ijk), (jkl), (klm)	-.003	.009	.013	.004	.006	.004
2	(ijl), (jkm), (ikl)	.017	.026	.015	.025	.029	.016
	(jlm)	-.005			-.013		
3	(ijm), (ikm), (ilm)	.036	.006	.032	.046	.049	.050

The first row of the table consists of the three 1-adjacent three-way interactions, i.e., interactions which represent the effect of the purchase outcomes at three successive purchase occasions. The estimates displayed in the last two columns are reproduced from Table II.1. For example, the three-way interaction effect denoted by (jkm) has been estimated as .026 and .029 for the EXCEPT 100 PERCENTERS and HEAVY segments respectively. The pattern observed

earlier for the two-way interactions tend to repeat itself for the three-way interactions: Interactions of same degree of adjacentness tend to be of numerical equal importance. On the other hand, the hypothesis expressed by H_2 holds but in a reverse fashion, i.e., the lower the degree of adjacentness, the lower the interaction effect.[11] As a formal test of H_1, as applied to all interactions this time, the following models were estimated:

NAME	DESCRIPTION
M_{2*5}	This model is similar to M_{2*} but includes the five-way interaction parameter as well.
$M_{(23)*5}$	This model is similar to M_{2*5} but includes three additional parameters for the three-way interactions. (The ten three-way interactions have been partitioned in three sets as explained above.)
$M_{(234)*5}$	Same as above but with an additional parameter for the five four-way interactions that are assumed to be equal.
$M_{(234)*}$	Same as above but without the five-way interaction.

Table II.5 displays the parameters estimates for the three models M_2, M_{2*} and M_{2*5}. Recall that a * indicates that hypothesis H_1 has been acted upon. Thus, model M_2 allows each two-way interaction to have a different numerical value whereas model M_{2*} requires that all two-way interactions with the same degree of adjacentness

(11) Given the magnitude of the estimates, such a reversal may just be the spurious result of random variations.

be equal. From Table II.8, fifth row, we see that the inclusion of the five-way parameter (M_{2*} versus M_{2*5}) brings about a significant reduction of the chi-square statistic for three segments: ALL, HEAVY and LOYAL. The interaction effects that do make a difference, in terms of goodness of fit, are the three-way interactions together with either one of the four or five-way interactions. A glance at Tables II.6 and II.8 suggests that either one of $M_{(23)*5}$ or $M_{(234)*}$ (both 8-parameter models) provide an adequate representation of the coffee data.

2.7 Comparison of brand-switching behavior across segments.

Although the objective of this chapter was not to discuss the segmentation strategy followed by MASSY et al. [1970], the preceding analysis provides a further test of the segments discriminant validity, that is, the potential of the segments to identify any behavior patterns that set the buyers apart from the market in general.

On the basis of the results afforded by their Brand Loyal model, Massy et al. [1970] found no significant differences in brand-switching behavior between the HEAVY and LIGHT segments. Our analysis does not support this claim. The parameter values of the segments for the three

models considered in Table II.6 and reproduced (in part) below exhibit different patterns:

	Parameter values	Model $M_{(234)}$*
Interactions	HEAVY	LIGHT
(i,j), (j,k), (k,l), (l,m)	.243	.277
(i,k), (j,l), (k,m)	.192	.211
(i,l), (j,m)	.088	.159
(i,m)	.075	.157

As the degree of adjacentness of the interactions increases, the importance of the interaction effects declines much more rapidly for the HEAVY than for the LIGHT segments. Thus, past purchases exert less influence on recent purchases for the HEAVY than for the LIGHT segments, a somewhat surprising result given that the learning effect observed in consumer brand choice is generally assumed to be inversely related to interpurchase time: the greater the time lapse between successive purchases, the greater the possible erosion of the learning effects.

The same difference in pattern exists for the LOYAL and the NON-LOYAL segments, but this time in the expected direction. Past purchases exert much more

influence on recent purchases for the LOYAL than for the NON-LOYAL buyers, an intuitive conclusion given that successive purchases of the former are probably highly correlated.

To further investigate the extent of the differences between the different segments, the 26 parameter values of the saturated model (Table IV.1) were correlated across segments. The results are summarized in Table II.10.

On the basis of the correlation matrix based on the whole set of interactions, the various segments look very much alike. The lowest correlation observed is .800 (between LOYAL and NON-LOYAL) while the ALL and EXCEPT 100 PERCENTERS segments exhibit near perfect correlation (.996).

It is interesting to note that the LOYAL and NON-LOYAL segments behave more like the HEAVY segment than the LIGHT segment. The observed patterns discussed above repeat themselves when the correlations are taken with respect to the bivariate interactions only. It is only at the level of the trivariate interactions that the various segments do differ from one another. Many of the correlations displayed in the matrix are in the (.10, .40) range, in absolute value. The correlation between the HEAVY and LIGHT segments drops from .96

for the bivariate interactions to -.11 for the trivariate interactions. The corresponding jump for the LOYAL NON-LOYAL segments is from .84 to -.42. The HEAVY segment that was found to behave very much like the NON-LOYAL segment is now hardly correlated with it: .11.

These results corroborate our previous finding that all interaction effects must be allowed for, in order to achieve an adequate representation of brand switching data.

Had we restricted our attention to bivariate interactions only, we would have reached the same conclusion as Massy et al. [1970], namely that the various segments do not exhibit sharply different brand-switching behavior.

2.8 Summary of the results.

The results of the preceding statistical analysis of brand-switching data are summarized below.

i) Consumer choice behaviour is a complex phenomenon. The degree of complexity embodied in a contingency table can be measured by the extent to which higher order interactions add a significant contribution toward explaining the observed variance in the dependent variable. From the empirical results

presented here, it appears that all interaction effects must be allowed for, in order to achieve an adequate representation of brand-switching behavior.

ii) The above analysis gives support to the learning hypothesis of brand-switching behavior in that past purchases of a brand affect the probability of that brand being bought again in the future.

iii) There is also evidence of a "forgetting" effect in that a purchase of a brand other than the individual's favorite brand decreases the individual's probability of buying his favorite brand on the next purchase occasion. Support for this "forgetting" effect comes from the non-rejection of hypothesis #2 which states that the influence of any two purchases of a brand varies inversely with the number of purchase occasions that elapsed between those two purchases.

Conclusion

The purpose of this chapter was to perform a statistical analysis of a particular set of consumer brand-switching data. The model was formulated as a log-linear model of contingency table by decomposing the observed proportions into main and interaction effects. Model fitting and hypothesis testing procedures were presented and the model applied to a product category (coffee) that has been previously studied by several researchers. The results indicate that (stochastic) models of consumer brand switching behavior should provide ways to accommodate the kind of adaptive behavior observed in the data. The purpose of the next chapter is to present a new stochastic model of brand choice that builds on the "regularities" uncovered in this chapter.

TABLE II.1

ESTIMATES OF INTERACTION EFFECTS AMONG THE PURCHASE OCCASIONS FOR SIX SEGMENTS - COFFEE DATA.

Segments	ALL		EXCEPT 100 Perc.		HEAVY		LIGHT		LOYAL		NON-LOYAL	
Interaction	Raw	Stand.	R	S	R	S	R	S	R	S	R	S
IJ	.227	3.857	.182	3.030	.253	2.861	.277	3.329	.234	5.940	.178	2.097
IK	.181	3.070	.147	2.440	.215	2.430	.221	2.656	.218	5.539	.106	1.248
IL	.078	1.321	.043	.717	.081	.917	.147	1.766	.061	1.564	.042	.499
IM	.080	1.355	.039	.653	.071	.798	.152	1.827	.115	2.936	-.000	-.004
JK	.200	3.390	.154	2.563	.219	2.478	.253	3.043	.196	4.985	.166	1.953
JL	.154	2.604	.120	1.988	.184	2.076	.198	2.380	.187	4.764	.080	.936
JM	.100	1.688	.068	1.128	.099	1.122	.169	2.026	.081	2.062	.062	.733
KL	.224	3.796	.178	2.951	.238	2.690	.282	3.385	.232	5.897	.189	2.215
KM	.170	2.875	.137	2.279	.184	2.081	.234	2.692	.165	4.194	.126	1.476
LM	.244	4.146	.200	3.317	.258	2.917	.302	3.635	.263	6.686	.200	2.355
IJK	.015	.258	-.003	-.054	.004	.047	-.041	-.499	-.020	-.498	.028	.323
IJL	.035	.589	.017	.280	.025	.284	-.024	-.293	.075	1.895	-.021	-.241
IJM	.050	.856	.036	.604	.046	.525	-.007	-.080	.035	.889	.036	.423
JKL	.028	.468	.009	.155	.006	.020	-.021	-.258	.007	.184	.032	.379
JKM	.042	.711	.026	.424	.029	.327	-.009	-.107	.091	2.301	-.016	-.191
KLM	.034	.570	.013	.219	.004	.046	-.003	-.036	.019	.483	.040	.472
IKL	.033	.551	.015	.251	.016	.179	-.019	-.227	.064	1.634	-.005	-.051
IKM	.010	.177	.006	.100	.049	.551	-.082	-.989	.009	.225	-.014	-.160
ILM	.050	.846	.032	.526	.050	.570	-.008	-.101	.033	.836	.059	.689
JLM	.011	.193	-.005	-.077	-.013	-.151	-.030	-.361	.002	.051	-.010	-.115
IJKL	-.006	-.096	-.036	-.595	.008	.086	.041	.489	-.040	-1.019	.011	.124
IJKM	-.008	-.143	-.033	-.548	-.023	-.260	.057	.689	.013	.333	-.046	-.539
IKLM	.027	.452	.001	.023	.007	.078	.095	1.147	-.021	-.541	.058	.686
JKLM	.007	.124	-.025	-.419	.022	.248	.052	.623	-.007	-.177	.006	.073
IJLM	.012	.202	-.006	-.108	.024	.266	.056	.673	.039	.982	-.048	-.564
IJKLM	.027	.463	.009	.145	.031	.350	-.022	-.260	.052	1.321	.002	.028
Variance (λ)	.00348		.00362		.00782		.00692		.00155		.00725	
Sample size	5310		3890		2660		2650		3600		1710	

TABLE II.2

PARAMETER ESTIMATES FOR MODEL M_{234} IN WHICH THE FIVE-WAY INTERACTION IS ASSUMED ZERO - COFFEE DATA.

Segments / Interactions	ALL	EXCEPT 100 Percenters	HEAVY	LIGHT	LOYAL	NON-LOYAL
IJ	.227	.182	.252	.277	.232	.179
IK	.181	.147	.214	.221	.216	.106
IL	.079	.043	.082	.146	.068	.042
IM	.080	.039	.070	.152	.113	.000
JK	.201	.155	.221	.253	.199	.166
JL	.153	.120	.185	.199	.186	.080
JM	.100	.068	.099	.168	.085	.062
KL	.225	.178	.279	.281	.233	.189
KM	.170	.137	.185	.224	.169	.176
LM	.244	.199	.257	.303	.259	.201
IJK	.019	-.002	.009	-.045	.012	.028
IJL	.036	.017	.027	-.026	.074	.020
IJM	.054	.037	.050	-.010	.042	.036
JKL	.027	.009	.005	-.022	.007	.032
JKM	.041	.025	.027	-.009	.085	-.016
KLM	.037	.014	.009	-.006	.024	.041
IKL	.032	.015	.015	-.019	.059	-.005
IKM	.012	.006	.051	-.084	.013	-.014
ILM	.052	.032	.053	-.011	.095	.059
JLM	.014	-.004	-.010	-.032	.009	-.010
IJKL	-.002	-.035	.012	.039	.029	.011
IJKM	-.006	-.032	-.019	.056	.019	-.046
IKLM	.030	.000	.011	.093	.011	.059
JKLM	.010	-.024	.025	.050	.001	.006
IJLM	.014	-.006	.027	.055	.046	-.048
χ^2	2.145	.210	1.224	.684	3.636	.008
χ^{2gf}_{1r}	2.146	.210	1.225	.683	3.663	.008
N	5310	3890	2660	2650	3600	1710

TABLE II.3

PARAMETER ESTIMATES FOR THE UNSATURATED MODEL M_{23} IN WHICH ALL FOUR-WAY AND THE FIVE-WAY INTERACTIONS ARE 0.

Segments Interactions	ALL	EXCEPT 100 Percenters	HEAVY	LIGHT	LOYAL	NON-LOYAL
IJ	.228	.181	.254	.290	.233	.176
IK	.182	.151	.216	.222	.215	.110
IL	.076	.046	.078	.143	.062	.043
IM	.080	.041	.071	.158	.113	.003
JK	.202	.163	.222	.252	.198	.167
JL	.153	.121	.180	.195	.186	.081
JM	.100	.069	.097	.166	.083	.070
KL	.220	.180	.234	.275	.235	.185
KM	.165	.139	.182	.217	.166	.117
LM	.248	.200	.260	.310	.263	.195
IJK	.017	-.018	.006	-.026	-.015	.020
IJL	.040	.009	.038	-.012	.082	-.028
IJM	.056	.029	.052	.010	.060	-.023
IKL	.039	.008	.022	.009	.050	.009
IKM	.016	-.012	.048	-.057	.013	-.014
ILM	.066	.029	.063	.013	.048	.055
JKL	.016	-.002	.007	-.004	.000	.016
JKM	.045	.012	.030	.009	.097	-.024
JLM	.016	-.002	.007	-.007	-.000	.016
KLM	.050	.006	.018	.023	.023	.049
χ^2_{gf}	6.733	8.675	3.615	37.421	8.046	8.406
χ^2_{lr}	6.707	8.648	3.593	38.302	8.048	8.472
N	5310	3890	2660	2650	3600	1710

PARAMETER ESTIMATES OF THE UNSATURATED MODEL IN WHICH ALL 3,4 and 5-WAY INTERACTIONS ARE ASSUMED ZERO (M_2).

SEGMENTS	ALL				EXCEPT 100 PERCENTERS				HEAVY			
	J	K	L	M	J	K	L	M	J	K	L	M
I	.227	.181	.078	.080	.182	.147	.043	.039	.253	.215	.081	.071
J		.200	.154	.100		.154	.120	.068		.219	.184	.099
K			.224	.170			.178	.137			.238	.184
L				.244				.200				.258

SEGMENTS	LIGHT				LOYAL				NON-LOYAL			
	J	K	L	M	J	K	L	M	J	K	L	M
I	.277	.221	.147	.152	.234	.218	.061	.115	.178	.106	.042	.000
J		.253	.198	.169		.196	.187	.081		.166	.080	.062
K			.282	.224			.232	.165			.189	.126
L				.302				.263				.200

The parameter values displayed above are those of the saturated model. They are reproduced from Table IV.1.

TABLE II.4

PARAMETER ESTIMATES FOR THREE UNSATURATED MODELS: M_2, M_{2*} and M_{2*5}.

Segments	ALL				EXCEPT 100 PERC.				HEAVY			
Interactions												
J K L M		Model M_2				Model M_2				Model M_2		
I	.254	.202	.096	.104	.184	.149	.052	.047	.277	.233	.094	.099
J		.220	.170	.117		.162	.123	.074		.237	.193	.108
K			.240	.185			.180	.139			.247	.197
L				.272				.205				.279
χ^2_{gf}		108.92				16.77				40.78		
χ^2_{lr}		112.41				16.90				41.22		

Segments	LIGHT				LOYAL				NONLOYAL			
Interactions												
J K L M		Model M_2				Model M_2				Model M_2		
I	.286	.214	.142	.155	.270	.236	.099	.136	.177	.114	.042	.008
J		.248	.195	.170		.228	.205	.117		.167	.083	.070
K			.276	.212			.252	.200			.186	.118
L				.314				.287				.196
χ^2_{gf}		45.02				73.15				21.90		
χ^2_{lr}		45.91				74.27				22.09		

SEGMENTS	ALL		EXCEPT P.		HEAVY		LIGHT		LOYAL		NON-LOYAL	
Models / Interact.	M_{2*5}	M_{2*}	M_{2*5}	M_{2*}	M_{2*5}	M_{2*}	M_{2*5}	M_{2*}	M_{2*5}	M_{2*}	M_{2*5}	M_{2*}
ij,jk,kl,lm	.237	.245	.182	.182	.252	.259	.280	.280	.248	.258	.180	.181
ik,jl,kl	.180	.184	.136	.136	.200	.205	.205	.205	.202	.211	.105	.104
il,jm	.100	.105	.061	.061	.095	.099	.156	.156	.099	.116	.053	.053
im	.102	.110	.052	.052	.100	.106	.160	.160	.130	.140	.012	.013
ijklm	.096	-	.008	-	.089	-	-.003	-	.105	-	.024	-
χ^2_{gf}	77.5	116.	21.2	21.4	28.3	44.2	49.8	49.8	51.4	77.6	23.5	24.6
χ^2_{lr}	75.4	119.	21.3	21.5	27.2	45.0	50.7	50.8	49.6	78.5	23.6	24.7

Model M_2 Bivariate interactions only.
Model M_{2*} Bivariate interactions only with constraint (4.37) imposed.
Model M_{2*5} Same as above plus the five-way interaction term.

TABLE II.5

PARAMETER ESTIMATES FOR THREE UNSATURATED MODELS $M_{(234)*5}$, $M_{(234)*}$ and $M_{(23)*5}$.

SEGMENTS	ALL			EXCEPT 100 PERC.			HEAVY			LIGHT			LOYAL			NON-LOYAL		
Models	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Interactions																		
ij,jk,kl,lm	.224	.225	.223	.179	.179	.181	.243	.243	.242	.277	.277	.281	.231	.231	.231	.179	.180	.180
ik,jl,km	.164	.164	.163	.131	.131	.134	.192	.192	.190	.211	.211	.207	.183	.184	.182	.102	.102	.103
il,jm	.089	.090	.088	.055	.055	.059	.088	.088	.086	.159	.158	.157	.075	.079	.075	.052	.052	.053
im	.085	.085	.085	.044	.044	.045	.075	.074	.076	.157	.157	.163	.113	.111	.112	.010	.010	.010
ijk,jkl,klm	.023	.026	.027	.003	.003	-.004	.003	.006	.008	-.024	-.026	-.004	-.005	-.001	-.004	.036	.037	.035
ijl,jkm,ikl and jlm	.029	.029	.032	.012	.013	.005	.017	.017	.021	-.023	-.024	-.004	.056	.051	.057	-.014	-.013	-.015
ijm,ikl,ilm	.039	.041	.042	.024	.024	.017	.047	.049	.050	-.029	-.031	-.010	.036	.039	.037	.021	.022	.020
ijkl,ijkm, iklm,jklm and ijlm	.008	.011	-	-.018	-.018	-	.010	.013	-	.062	.060	-	.002	.008	-	-.003	-.003	-
ijklm	.026	-	.031	.008	-	-.003	.025	-	.032	-.020	-	.011	.041	-	.042	.006	-	.006
χ^2_{gf}	11.6	13.7	12.8	11.4	11.6	17.1	4.42	5.27	5.23	11.7	12.2	48.5	12.2	14.6	12.2	15.9	16.0	16.0
χ^2_{lr}	11.6	13.7	12.7	11.4	11.6	17.0	4.44	5.35	5.19	11.6	12.2	49.3	12.4	15.1	12.4	15.8	15.8	15.8
$\chi^2_{Morrison}$	8.0			7.8			2.61			10.1			8.5			13.5		

Model 1 = $M_{(234)*5}$: None of the interaction is assumed zero, but the "adjacentness" constraint is imposed
Model 2 = $M_{(234)*}$: Same as above but without the five-way interaction.
Model 3 = $M_{(23)*5}$: Same as model 1 but without the four-way interaction parameter.

TABLE II.6

MODEL	Parameters set to 0	Degrees of freedom	ALL		EXCEPT 100 P.		HEAVY		LIGHT		LOYAL		NON-LOYAL	
			χ^2_{lr}	χ^2_{gf}	χ^2_{lr}	χ^2_{gf}	χ^2_{lr}	χ^2_{gf}	χ^2_{lr}	χ^2_{gf}	χ^2_{lr}	χ^2_{gf}	χ^2_{lr}	χ^2_{gf}
234	5-way	6	2.15	2.14	.21	.21	1.23	1.22	.68	.68	3.67	3.64	.01	.01
23	4,5-way	11	6.71	6.73	8.64	8.67	3.59	3.61	38.30	37.42	8.05	8.05	8.47	8.41
2	3,4,5-way	21	112.41	108.92	16.90	16.77	41.22	40.78	45.91	45.02	74.27	73.15	22.10	21.91
2*	4,5-way	27	119.39	116.18	21.53	21.20	44.96	44.24	50.83	49.82	78.52	77.62	24.66	24.56
2*5	3,4-way	26	75.37	77.53	21.27	21.40	27.24	28.33	50.74	49.77	49.56	51.44	23.64	23.55
234*5	none	22	11.58	11.64	11.41	11.43	4.44	4.42	11.61	11.67	12.38	12.25	15.78	15.91
234*	5-way	23	13.69	13.71	11.58	11.59	5.35	5.27	12.16	12.24	15.10	14.62	15.83	15.97
23*5	4-way	23	12.69	12.81	17.00	17.05	5.19	5.23	49.26	48.49	12.38	12.25	15.87	16.00
	SAMPLE SIZE		5310		3890		2660		2650		3600		1710	

SUMMARY OF CHI-SQUARE STATISTICS FOR VARIOUS UNSATURATED MODELS.

TABLE II.7

SUMMARY OF THE LIKELIHOOD RATIO CHI-SQUARE STATISTICS FOR VARIOUS UNSATURATED MODELS.

$$\chi^2 (H|H') = \chi^2(H) - \chi^2 (H')$$

Model H	Model H'	Degrees of freed.	ALL	EXCEPT 100 P.	HEAVY	LIGHT	LOYAL	NON-1.	CRITICAL LEVEL $\chi^2_{.01}$	$\chi^2_{.05}$	$\chi^2_{.10}$
234	2345	1	2.15	.21	1.23	.68	3.67	.01	6.63	3.84	2.71
23	234	5	4.56	8.43	2.36	37.62	4.38	8.46	15.09	11.07	9.24
2	23	10	105.70	8.26	37.63	7.61	66.22	13.63	23.21	18.31	15.99
2*	2	6	6.98	4.63	3.74	4.92	4.25	2.56	16.81	12.59	10.64
2*5	2*	1	44.02	.26	17.72	.09	28.96	1.02	6.63	3.84	2.71
(234)*	234	17	11.54	11.37	4.12	11.48	11.43	15.82	32.90	27.75	24.79
(23)*5	(234)*5	1	1.11	5.59	.75	37.65	.00	.09	6.63	3.84	2.71
(234)*	(234)*5	1	2.11	.17	.92	.54	2.72	.05	6.63	3.84	2.71
(234)*5	2345	17	11.58	11.41	4.44	11.61	12.38	15.78	32.90	27.75	24.79
2*5	(23)*5	3	62.68	4.27	22.05	1.48	37.18	7.77	11.34	7.81	6.25

This table exhibits the chi-square statistics $\chi^2(H|H')$ used to discriminate between various unsaturated models (see 4.36).

TABLE II.8

CORRELATION MATRIX OF THE VARIOUS SEGMENTS BASED ON THE MODELS RESPECTIVE PARAMETER ESTIMATES FOR THE SATURATED MODEL AND TWO UNSATURATED MODELS.

SATURATED MODEL.

	Except 100 percenters	Heavy	Light	Loyal	Non-Loyal
All	.996	.985	.909	.959	.929
Except 100 p.		.978	.874	.960	.922
Heavy			.900	.949	.904
Light				.853	.834
Loyal					.799

Model M_2 (bivariate interactions only).

	Except 100 percenters	Heavy	Light	Loyal	Non-Loyal
All	.997	.989	.988	.953	.965
Except 100 p.		.993	.978	.950	.959
Heavy			.956	.957	.938
Light				.922	.977
Loyal					.845

MODEL M_3 (trivariate interactions only).

	Except 100 percenters	Heavy	Light	Loyal	Non-Loyal
All	.966	.519	.784	.611	.423
Except 100 p.		.477	.790	.603	.346
Heavy			-.111	.356	.107
Light				.428	.435
Loyal					-.425

TABLE II.10

CHAPTER III

A NEW LEARNING MODEL OF BRAND CHOICE

The purpose of this chapter is to discuss a new stochastic model of brand choice that allows for the kind of adaptive behavior observed in consumer brand-switching behavior and confirmed by statistical analyses such as the one reported in the previous chapter.

This adaptive behavior goes by the name of "learning" in the marketing literature. Learning models of consumer brand choice behavior have occupied an important position in the literature of brand choice ever since Kuehn [1958] adapted the work of Bush and Mosteller [1955] and applied the model to data on switching patterns for frozen orange juice.

The fundamental concept underlying all learning models of brand choice is that of purchase event feedback. That is, the act of purchasing and using a particular brand is assumed to affect the probability that this brand will be selected again the next time the product class is to be purchased.

The model to be presented deals explicitly with the "learning" feature of consumer brand choice.

It attempts to predict an individual's probability of selecting a particular brand on his next purchase occasion given his past purchase history.

The next section briefly describes two stochastic models of brand choice which will be used for comparison purposes to assess the performance of the new purchase - to - purchase model of brand choice.

3.1 Linear Learning and Markov Models of Brand Choice.

Several stochastic models provide some ways to accommodate the kind of adaptive behavior observed in brand choice data. Two of these will be considered for comparison purposes. They are respectively:

- Morrison's Compound Markov Model (1965a)
- Kuehn's Linear Learning Model (1962).

The rationale for choosing the above two models is based on practical considerations. They have been

extensively tested in the marketing literature and stand out for their mathematical tractability and "good" empirical performance in the face of actual data. Massy et al. [1970] have discussed them at length and reproduced some panel data which will form the basis of the empirical comparison presented in a later section.

3.1.1 The Linear Learning Model

Kuehn [1962] proposed a model of consumer brand choice behavior in which p_t, the probability of purchase at purchase occasion t, was a linear function of the probability at occasion t-1 and the outcome of the purchase (brand selected) at t-1. In symbols:

$$(3.1) \quad \begin{cases} p_t = \alpha + \beta + \gamma p_{t-1} & \text{if brand 1 is purchase at } t-1, \\ \qquad\quad \alpha + \gamma p_{t-1} & \text{if any other brand is purchased at } t-1. \end{cases}$$

These equations can be represented with the familiar diagram of figure III.1 which illustrates many of the properties of the Linear Learning Model. The two relations expressed in (3.1) are called the purchase operator and rejection operator, respectively. All families are assumed to have the same values for the parameters α, β and γ, though of course the process leads to different values of p_t for different families, due to

Diagramatical Representation of the

Linear Learning Model

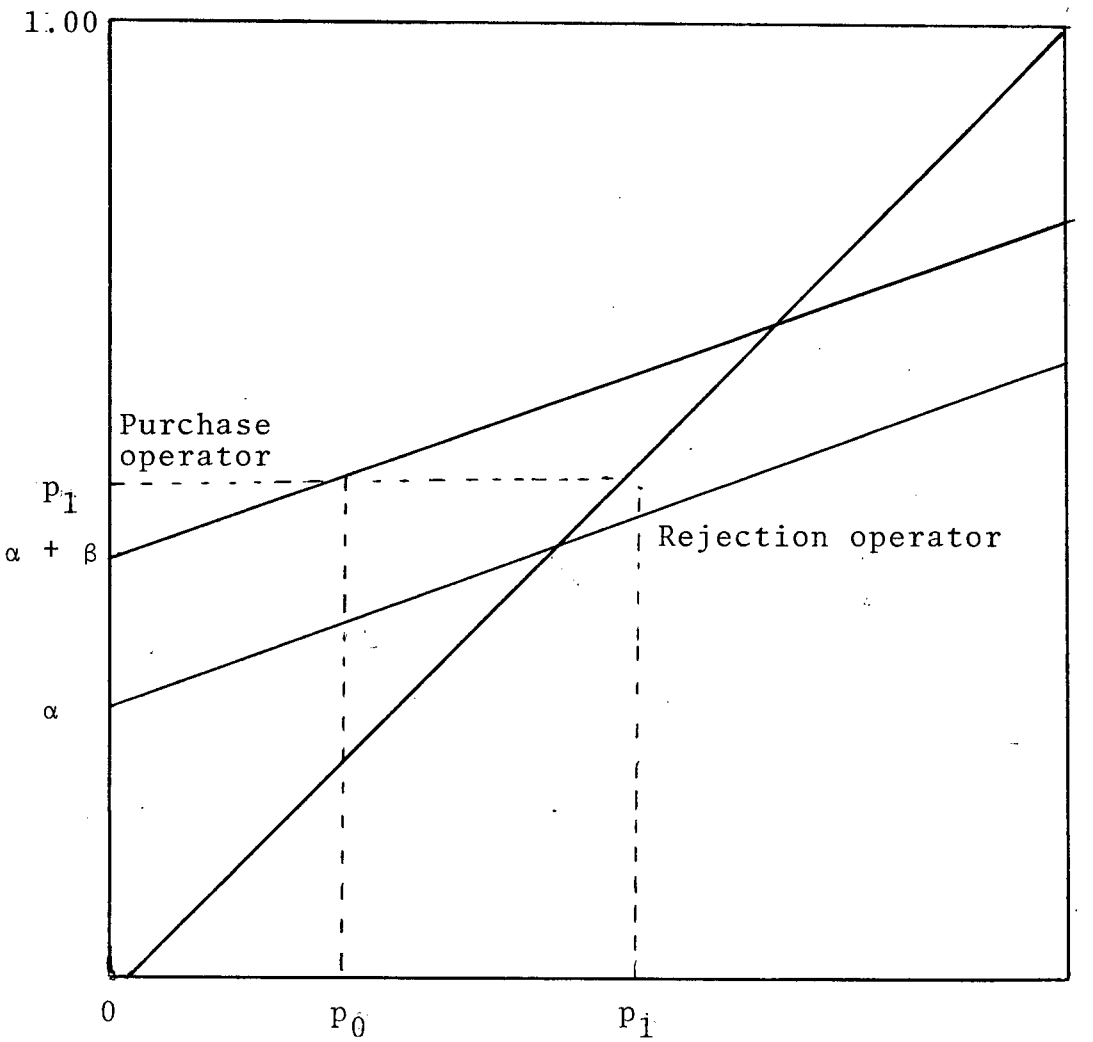


FIGURE III.1

variations in realized purchased histories. "Learning" takes place because a purchase of brand 1 leads to a larger value of p_{t+1} than a purchase of brand 0, for any trial t.

In addition to its intrinsic interest, the Linear Learning model can be viewed as a generalization of both the zero-order Bernouilli and the first-order Markov models. If $\gamma=0$, equation (3.1) becomes independent of p_{t-1}. This is equivalent to a Markov model with purchase decisions as states having transitions matrix

	1	0
1	$\alpha+\beta$	$1-\alpha-\beta$
0	α	$1-\alpha$

Similarly, if $\alpha = \beta = 0$ and $\gamma = 1$, then $p_{t+1} = p_t$ regardless of the outcome at time t, and so we have a zero-order Bernouilli model. The adaption of this model by Massy [1965] to include heterogeneity did not change the basic formulation, but involved only the method of estimating the model's parameters. Parameters α, β and γ can be estimated without assuming that the entire population has the same initial probability of purchase.

3.1.2 A Compound Markov Model: Morrison [1965a]'s Brand Loyal Model

In the Compound Markov model, the last purchase, and only the last purchase, influences the current purchase decision. That is, the Compound Markov model allows for first-order behavior. The Brand Loyal model developed by Morrison is a special case of the general compound Markov model. For Morrison's model, the brand loyal population is defined as follows:

1. Each individual is a first-order 0-1 process with transition matrix

	1	0
1	p	1-p
0	kp	1-kp

2. Different individuals may have different p. Thus, p is a random variable distributed according to some density function f(p).

3. k is a constant, the same for each individual.

The Brand loyal model says that an individual with a high probability of remaining with brand 1 will also have a higher probability of leaving brand 0 to buy brand 1 than an individual with a low probability of remaining with brand 1. Thus, people with high p are also more apt to switch to brand 1 than are people with lower p. When k=1, the model becomes a compound Bernouilli model of brand choice.

The expected probabilities of purchase, $P(1|x)$ are given by the following formula (see Massy et al. [1970] p. 69):

$$(3.2) \qquad P(1|x) = \frac{\int_0^1 p\ell(x|p).f(p)dp}{\int_0^1 \ell(x|p).f(p)dp}$$

where $\ell(x|p)$ is the conditional likelihood of the past purchase history x given p,

and $f(p)$ is the prior distribution of p.

This completes the description of Morrison's Brand Loyal model. In the next sections, we develop a new learning model of brand choice and compare results across models.

3.2 A New Purchase-to-Purchase Learning Model of Brand Choice: The Polya-Learning Model

The model, hereafter referred to as the Polya-Learning model is based on the so-called Polya urn model with contagion (see Feller [1957]). The idea to use urn models to describe after-effects (such as purchase event feedback) seems to be due to Polya. His simple urn scheme is described below to motivate the modifications that were brought to it in order to make it more germane to the brand choice context.

3.2.1 The Polya Urn Model

From an urn containing Np white and N(1-p) black balls, a series of drawings is made, but instead of replacing the ball drawn, $1 + N\beta$ balls of the colour last drawn are laid into the urn before a new drawing is made. Like p, the constant β is a multiple of 1/N, but while p is a value between zero and one, the constant β can take every positive multiple of 1/N, zero included. However, in the limit, i.e., as N goes to ∞, β and p can take any value in the real line and unit interval, respectively. The unknown results of the individual drawings are random variables x_n which a priori have the same probability distribution as in a Bernouilli model. However, unless $\beta = 0$, these variables are no longer independent, for, k white balls having been obtained in the first n drawings, the urn will, after these n drawings, contain $N(p+k\beta)$ white balls and $N[(1-p)+(n-k)\beta]$ black balls. Thus, the probability of drawing white in the (n+1)st drawing will under these conditions be equal to:

$$\text{(3.3)} \qquad \text{Pr.}\left(\begin{matrix}\text{drawing white}\\ \text{at trial } n+1\end{matrix} \;\middle|\; \begin{matrix}\text{k white balls have}\\ \text{already been drawn}\end{matrix}\right) = \frac{p + k\beta}{1 + n\beta}$$

It is clear from (3.3) that in the Polya-urn model, results of the drawings made have a "contagious" influence on the results of the subsequent drawings. The strength

of this contagion is measured by the parameter β. A value of zero for this parameter indicates that the process is Bernouilli, i.e., the drawings are independent from one another.

The application of this urn model to brand choice theory is immediate. If drawing white and black balls becomes purchasing brand 1 and brand 0, we have a "learning" model of brand choice.

3.2.2 Limitation of the Simple Polya Urn Model as Applied to Brand Choice

A special feature of the Polya urn model is that the probability of drawing white at the (n+1)st drawing only depends on the number of white balls drawn at all previous trials, whatever the order in which they were drawn. In the brand choice context, this analytical simplicity implies the absence of any "recency" effect, i.e., all past purchases of brand 1 contribute to increase the purchase probability on future purchase occasions, irrespective of their occurrence time.

The practical implications are that the following groups of past purchase histories yield the same conditional probability of purchase:

Past Purchase history n	Observed conditional[1] probability of purchase $P(1\|x)$	Model predicted conditional probability of purchase $\hat{P}(1\|x)$
0 1 1 1	.748	$\frac{P + 3\beta}{1 + 4\beta}$
1 0 1 1	.753	
1 1 0 1	.700	
1 1 1 0	.606	
0 0 1 1	.684	$\frac{P + 2\beta}{1 + 4\beta}$
0 1 0 1	.575	
0 0 1 1	.684	
1 0 0 1	.605	
0 1 1 0	.590	
1 1 0 0	.489	
0 0 0 1	.553	$\frac{P + \beta}{1 + 4\beta}$
0 0 1 0	.477	
0 1 0 0	.384	
1 0 0 0	.368	

(1) These observed brand choice probabilities were obtained from panel data on coffee consumption. See Massy et al. [1970, p. 126].

The statistical analysis of the previous chapter made it clear that both the number of past purchase of the brand and their position in the purchase string must be reckoned with in order to provide an adequate representation of brand choice data. The Polya urn model provides a mean to allow for learning behavior in terms of purchase frequency but fails to recognize the impact of purchase recency. In the next section, we shall extend the simple Polya urn model in several ways to accommodate the influence of purchase recency on subsequent purchase probabilities.

3.2.3 Mathematical development of the Polya-Learning Model

Assumptions

The assumptions underlying the model will be divided into three basic classes:

a) Model specification assumptions.

b) Response assumptions.

c) Response probability change assumptions.

Each of these are discussed in turn.

a) Model Specification Assumptions

S1: Each respondent's response is generated by the same stochastic process.

S2: At any given response occasion n, there are two mutually exclusive and collectively exhaustive responses: response 1 (purchase of the brand under study) and response 2.

S3: Each respondent is associated with an urn containing Np black and N(1-p) red balls. The urn composition in terms of black and red balls may vary across individuals. That is, we view p as a random variable distributed according to some density f(p).

Assumptions S1 and S2 deserve no comments. Assumption S3 provides a mechanism to allow for consumer heterogeneity through the consumer's initial purchase probability. This heterogeneity may result from differences in perceptions, preferences and attitudes toward the brand. These differences are not taken into account explicitly but are modeled via the parameter p that is allowed to be a random variable.

b) Response Assumptions

R1: At any given purchase occasion, each individual's probability of purchasing brand 1 is equal to the proportion of black balls in his urn. Initially, this probability is equal to p as stated in S3.

R2: The actual response made by an individual at occasion n depends on his responses at all previous response occasions. The form of this dependence is specified in the response change probability assumptions.

R3: Individuals respond independently of one another.

The last assumption which is required for testing purposes, assumes away the role of interpersonal influence in brand choice decisions. The nature of the data to be used[2] ensures that this assumption is reasonable satisfied in our particular case.

---

2) Consumer panel data.

c) Response Probability Change Assumption

C1: After each drawing (purchase occasion), the ball drawn is replaced and, moreover, β_n black balls are added to the urn if a black ball was drawn. If, instead, a red ball was drawn, λ_n red balls are added to the urn.

Assumption C1 constitutes a marked departure from the simple Polya urn model, and provides us with a way to model the influence of purchase recency. The strength of the "learning" and "forgetting" effects depend on the absolute value of β_n and λ_n respectively. If $\beta_n \geq \lambda_n$ for all n, we assume that consumers "learn" faster than they "forget". In other words, it will take more purchases of brand 0 (drawing of a red ball) to bring an individual's probability of purchasing brand 1 down from p_1 to p_o than it took purchases of brand 1 to bring that same probability from p_o up to p_1. Also, brand loyal consumers will tend to exhibit higher values for β_n and lower values for λ_n than less loyal consumers.

If $\beta_{n+1} > \beta_n$, recent purchases of brand 1 will exert more influence on subsequent purchase probabilities than will less recent ones.

Mathematical development

To turn our picturesque description into mathematics, we shall introduce some notation.

Let

$$X_n = \begin{cases} 1 & \text{if brand 1 was purchased at occasion } n \\ 0 & \text{otherwise} \end{cases}$$

$I(X_n)$ be the indicator function for the random variable X_n, i.e.

$$I_1\ (X_n) = \begin{matrix} 1 & \text{if } X_n = 1 \\ 0 & \text{otherwise} \end{matrix}$$

and similarly,

$$I_o\ (X_n) = \begin{matrix} 1 & \text{if } X_n = 0 \\ 0 & \text{otherwise.} \end{matrix}$$

According to the above assumptions, the probability that an individual with initial purchase probability p purchases brand 1 on the (n +1)st trial given his past purchase history for the first n purchases is given by

(3.4) $\Pr[X_{n+1} = 1 | X_n = x_n, \ldots, X_1 = x_1, p] =$

$$\frac{p + \sum_{k}^{n} \{\beta_k \; I_1 \; (X_k)\}}{1 + \sum_{k}^{n} \{\beta_k \; I_1 \; (X_k) + \lambda_k I_0(X_k)\}}$$

This probability reduces to that of the Polya simple urn model when β_n and λ_n equal some positive constant, for all n. To reduce the number of parameters to be estimated from the data, we introduce a simplifying assumption for the non-stationarity of the learning parameters β_n and λ_n, namely:

(3.5) $$\beta_n = n\beta$$
$$\lambda_n = n\lambda$$

Alternative formulations are of course possible[3]. Support for the above formulation comes from its simplicity and the results from the previous chapter that demonstrated the importance of purchase re cency on subsequent purchase probability.

Substituting (3.5) in (3.4) yields:

$$\Pr(1|x,p) \stackrel{\text{def}}{=} \Pr(X_{n+1}=1|X_n=x_n,\ldots, X_1=x_1,p)$$

(3.6) $$= [p + \beta\sum_{k=1}^{n} k \; I_1(X_k)]/[1 + \sum_{k=1}^{n} k\{\beta I_1(X_k) + \lambda I_0(X_k)\}]$$

---

(3) An alternative formulation is the following: $\beta_n = a + n\beta$, $\lambda_n = b + n\lambda$

It has the advantage over (3.5) to reduce to the Polya case for $\beta = \lambda = 0$ and $a = b$. However, it introduces another two parameters (a and b) and was thus discarded in favor of the more parsimonous formulation in (3.5).

Updating the prior

If we know an individual's initial purchase probability p, his probability of purchasing brand 1 on any trial is given by (3.6), provided we also know his past purchase history, say x. However, we never know a consumer's true value for p, so that it is for us a random variable with prior distribution f(p). Once we observe a past history, x, our prior distribution gets updated to a posterior distribution $f(p|x)$ (See e.g., Massy et al. [1970]). Hence, an individual with past history x picked at random will have a probability of purchasing brand 1 equal to:

$$\Pr(1|x) = \int_0^1 \Pr(1|x,p)f(p|x)dp. \tag{3.7}$$

From Bayes' theorem

$$f(p|x) = \ell(x|p).f(p)/\ell(x) = \ell(x|p).f(p)/\int_0^1 \ell(x|p).f(p)dp$$

where $\ell(x|p)$ is the likelihood of the purchase sequence x given p. We first need to derive the likelihood $\ell(x|p)$ for all x. They follow directly from (3.6). For example, consider the purchase sequence denoted by 0101. For each purchase occasion, the individual's probability of purchasing brand 1 is given by (3.6). Thus, we obtain:

$$\ell(0101|p) = (1-p)\ \frac{p}{(1+\lambda)} \cdot \frac{(1-p+\lambda)}{(1+\lambda+2\beta)} \cdot \frac{(p+2\beta)}{(1+4\lambda+2\beta)}$$

since

$$\Pr(0|p) = 1-p$$

$$\Pr(1|X_1 = 0,p) = p/(1-\lambda)$$

$$\Pr(0|X_2 = 1,X_1 = 0,p) = (1-p+\lambda)/(1+\lambda+2\beta)$$

$$\Pr(1|X_3 = 0,X_2 = 1,\ X_1 = 0,p) = (p+2\beta)/(1+4\lambda+2\beta).$$

The remaining likelihoods $\ell(x|p)$ are similarly derived. Displayed in table III.1, they are all of the form:

$$\ell(x|p) = \frac{1}{d_x} \sum_{k=0}^{4} a_{kx}\ p^k \tag{3.8}$$

where the a_{kx}, $k=0,\ldots,4$ and d_x are functions of β, λ and the particular purchase sequence x being dealt with. Substituting (3.8) in (3.7) yields:

$$\Pr(1|x) = \left\{ \int_0^1 \Pr(1|x,p)\ \{ \frac{1}{d_x} \sum_{k=0}^{4} a_{kx}p^k \} f(p)dp. \right\} / \ell(x) \tag{3.9}$$

From (3.6), we can express $\Pr(1|x,p)$ as:

$$\Pr(1|x,p) = (p + b_x)/c_x \tag{3.10}$$

where b_x and c_x are functions of β, λ and the particular purchase sequence being dealt with. Substituting (3.10) in (3.9) yields after some rearrangement and integrating p out:

$$(3.11) \qquad Pr(1|x) = \left\{ \sum_{k=0}^{5} \mu_{5-k}[a_{kx}+b_x a_{(k-1)x}] \right\} \Big/ [d_x c_x \sum_{k=0}^{4} a_{kx}\mu_{4-k}].$$

where

$$a_{kx} = 0 \text{ for all } k<0$$

$$\mu_k = \int_0^1 p^k f(p)dp = k^{th} \text{ moment of } p.$$

The values of $\{a_{kx}\}_{k=0}^{4}$, b_x, c_x and d_x for all purchase sequences x are displayed in table III.1. As in the Linear Learning model, the first five moments of the random variable p are treated as parameters to be estimated from the data, together with the "learning" parameters β and λ. The parameters must satisfy the following constraints:

$$(3.12) \quad 0 \leqq \beta, \lambda$$

$$(3.13) \quad 0 \leqq \mu_5 \leqq \mu_4 \leqq \mu_3 \leqq \mu_2 \leqq \mu_1 \leqq 1$$

$$(3.14) \quad 0 \leqq \mu_2 - \mu_1^2$$

$$(3.15) \quad 0 \leqq \mu_4 - \mu_2^2$$

Constraint (3.12) will ensure that the predicted probabilities lie in the unit interval. The last three inequalities must hold since the μ_k's are the moments[4] of the random variable p.

---

4) Note that $\mu_2 - \mu_1^2$ is just the variance of p. These inequalities can be shown to hold my making use of the Jensen or the Cauchy inequalities.

INTERMEDIARY RESULTS TO COMPUTE THE LIKELIHOOD $\ell(x|p)$ AND $P(1|x)$.

Purchase seq. x	a_{0x}	a_{1x}	a_{2x}	a_{3x}	a_{4x}
1111	1	10β	$27\beta^2$	$18\beta^3$	0
0000	1	$-(4+10\lambda)$	$[(2+10\lambda+18\lambda^2)+(2+\lambda)(2+9\lambda)]$	$-[(2+9\lambda)(\lambda+1)+(2+\lambda)(1+9\lambda+18\lambda^2)]$	$(\lambda+1)(1+9\lambda+18\lambda^2)$
0111	-1	$(1-7\beta)$	$\beta(7-10\beta)$	$10\beta^2$	0
1011	-1	$(1-5\beta)$	$\beta(5-4\beta)$	$4\beta^2$	0
1101	-1	$(1-4\beta)$	$\beta(4-3\beta)$	$3\beta^2$	0
1110	-1	$(1-4\beta)$	$\beta(4-3\beta)$	$3\beta^2$	0
0011	1	$-(2+\lambda-3\beta)$	$[1+\lambda-3\beta(2+\lambda)]$	$3\beta(1+\lambda)$	0
0101	1	$-(2+\lambda-2\beta)$	$[1+\lambda-2\beta(2+\lambda)]$	$2\beta(1+\lambda)$	0
1001	1	$-(2+2\lambda-\beta)$	$[-\beta+(1+2\lambda)(1-\beta)]$	$\beta(1+2\lambda)$	0
0110	1	$-(2-2\beta+\lambda)$	$[1+\lambda-2\beta(2+\lambda)]$	$2\beta(1+\lambda)$	0
1010	1	$-(2-\beta+2\lambda)$	$[1+2\lambda-\beta(2+2\lambda)]$	$\beta(1+2\lambda)$	0
1100	1	$-(2+3\lambda-\beta)$	$[-\beta+(1+3\lambda)(1-\beta)]$	$\beta(1+3\lambda)$	0
0001	-1	$(3+4\lambda)$	$-(3+8\lambda+3\lambda^2)$	$(1+4\lambda+3\lambda^2)$	0
0010	-1	$(3+4\lambda)$	$-(3+8\lambda+3\lambda^2)$	$(1+4\lambda+3\lambda^2)$	0
0100	-1	$(3+5\lambda)$	$-(3+10\lambda+4\lambda^2)$	$(1+5\lambda+4\lambda^2)$	0
1000	-1	$(3+7\lambda)$	$-(3+14\lambda+10\lambda^2)$	$(1+7\lambda+10\lambda^2)$	0

	b_x	c_x	d_x
1111	10β	$(1+10\beta)$	$(1+\beta)(1+3\beta)(1+6\beta)$
0000	0	$(1+10\lambda)$	$(1+\lambda)(1+3\lambda)(1+6\lambda)$
0111	9β	$(1+\lambda+9\beta)$	$(1+\lambda)(1+\lambda+2\beta)(1+\lambda+5\beta)$
1011	8β	$(1+2\lambda+8\beta)$	$(1+\beta)(1+\beta+2\lambda)(1+4\beta+2\lambda)$
1101	7β	$(1+3\lambda+7\beta)$	$(1+\beta)(1+3\beta)(1+3\beta+3\lambda)$
1110	6β	$(1+4\lambda+6\beta)$	$(1+\beta)(1+3\beta)(1+6\beta)$
0011	7β	$(1+3\lambda+7\beta)$	$(1+\lambda)(1+3\lambda)(1+3\lambda+3\beta)$
0101	6β	$(1+4\lambda+6\beta)$	$(1+\lambda)(1+\lambda+2\beta)(1+4\lambda+2\beta)$
1001	5β	$(1+5\lambda+5\beta)$	$(1+\beta)(1+\beta+2\lambda)(1+\beta+5\lambda)$
0110	5β	$(1+5\lambda+5\beta)$	$(1+\lambda)(1+\lambda+2\beta)(1+\lambda+5\beta)$
1010	4β	$(1+6\lambda+4\beta)$	$(1+\beta)(1+\beta+2\lambda)(1+4\beta+2\lambda)$
1100	3β	$(1+7\lambda+3\beta)$	$(1+\beta)(1+3\beta)(1+3\beta+3\lambda)$
0001	4β	$(1+6\lambda+4\beta)$	$(1+\lambda)(1+3\lambda)(1+6\lambda)$
0010	3β	$(1+7\lambda+3\beta)$	$(1+\lambda)(1+3\lambda)(1+3\lambda+3\beta)$
0100	2β	$(1+8\lambda+2\beta)$	$(1+\lambda)(1+\lambda+2\beta)(1+4\lambda+2\beta)$
1000	β	$(1+9\lambda+\beta)$	$(1+\beta)(1+\beta+2\lambda)(1+\beta+5\lambda)$

TABLE III.1

This completes the mathematical description of the Polya-Learning model.

3.2.4 Model fitting procedures and data

A minimum chi-square procedure was utilized to estimate the seven parameters of the model (β, λ and the first five moments of the random variable p). Given the non-linearity of the chi-square function to be minimized and the various constraints imposed on the parameters, an iterative method of constrained optimization had to be resorted to[5].

The coffee data that has already been described and analyzed in the previous chapter will provide the empirical basis for testing and comparing the performance of the newly developed learning model with alternative models of brand choice.

3.2.5 Empirical results

The parameter estimates for each of the six segments[6] of coffee buyers are displayed in table III-2

---

5) The algorithm that has been used in this study has been developed by M.J. Box [1965].

6) These segments were described in chapter II page 40.

together with the corresponding chi-square goodness of fit statistics.

TABLE III.2

Parameter estimates \ SEGMENTS	ALL	EXCEPT 100%	HEAVY	LIGHT	LOYAL	NON-LOYAL
β	.142	.120	.257	.051	.163	#
λ	.082	.053	.047	.104	.050	
μ_1	.586	.487	.350	.776	.587	
μ_2	.357	.237	.122	.612	.373	
μ_3	.238	.121	.065	.491	.266	
μ_4	.176	.064	.052	.403	.210	
μ_5	.142	.034	.046	.338	.278	
χ^2	8.053	7.340	3.962	8.310	12.300	
p - level	.43	.50	.86	.41	.15	
degrees of freedom	8	8	8	8	8	
Sample size	5,310	3,890	2,660	2.650	3,600	1,710
$\chi^2_{8,.05}$	= 15.507					

No estimates are available for this segment due to convergence problems.

The model fits the data quite well, as evidenced by the low values of the chi-square statistics or the high p - values[7], except in the case of "Loyal" buyers.

The magnitude of the two Learning parameters β λ indicates the strength of the Learning of each segment. The greater the values of β and λ, the greater the learning effect observed in the data. That is, the lower the values of β and λ, the less effect the last purchases have, or to put it differently: the lower the values of β and λ, the more "Bernouilli" the population of consumers becomes.

Heavy versus Light consumers.

Heavy and Light consumers exhibit different behavior patterns, as suggested by their parameter values:

---

7) The p - level associated with a chi-square statistic is defined as:

$$p - \text{level} = \int_{\chi^2}^{\infty} f(x)\, dx$$

where f(x) is the chi-square distribution with the appropriate number of degrees of freedom. A low p - level indicates that the model is not a viable representation of the process. For models which have a different number of parameters, p - levels rather than chi-square values should be compared to correct for the different degrees of freedom.

	"Learning" parameter β	"Forgetting" parameter λ
Heavy	.257	.047
Light	.051	.104

The Light segment is the only group for which $\lambda > \beta$. For these consumers, the relative decrease in the probability of purchasing their favorite brand following a purchase of a competing brand is greater than the relative increase following a purchase of their favorite brand. The reverse holds for the Heavy segment, where the learning effect proves to be much stronger than the forgetting effect. An example will clarify those points.

Suppose we have two consumers, one "Light" and one "Heavy", each with an initial probability of buying their favorite brand equal to .5. After the first purchase, their probability of purchasing their favorite brand at the next trial increases or decreases depending on the purchase outcome. The following table summarizes the possible outcomes.

Segments	Initial purchase probability	Purchase probability after the first purchase	
		Favorite brand was purchased	Favorite brand was not purchased
Light	.5	.524 (4.8)*	.453 (10.6)
Heavy	.5	.602 (20.4)	.478 (4.4)

* Figures in parenthesis correspond to relative increase or decrease in the purchase probability following a purchase.

Following a purchase of the favorite brand, the purchase probability jumps form .5 to .6 for the "Heavy" consumer as compared with a mere .524 for the "Light" consumer, or a 20.4% and 4.8% increase respectively. If the favorite brand was not selected, the corresponding probabilities decrease to .478 and .453 for the Heavy and Light consumer respectively. In other words, Heavy consumers exhibit stronger brand loyalty than Light ones. Purchase of competing brands will not affect their probability of purchasing their favorite brand very much, that is, they experience mild forgetting effects and strong learning effects. The opposite is true for Light consumers.

The definition of brand loyalty in terms of learning (β) and forgetting (λ) effects introduces a fine distinction between disloyal behavior and mere brand switching. This distinction carries much importance for the brand manager who would like to know the extent to which brand-switching is evidence of disloyal behavior or mere variety seeking. For the coffee data, we could tell our manager that a Light consumer who switches brands is less "loyal" than a Heavy consumer who does so. The latter is more likely to return to his favorite brand as evidenced by the low value of the forgetting parameter. This conclusions stands at variance with Massy et al.

[1970]'s who found the two segments to be virtually identical, but is consistent with the findings of the previous chapter.

Comparison of the Brand Loyal, Linear Learning and Polya Learning models.

How does the Polya-Learning model perform as compared with other competing brand choice models? Table III.3 displays the values of three different models' chi-square statistics and their corresponding p-levels. The models are:

i) Kuehn's Linear Learning model.

ii) Morrison's Brand Loyal model.

iii) The Polya - Learning Model.

The two Learning models outperform the Brand Loyal (Markov) model. In most cases, the fit is far better as can be seen by comparing the p - values. There seems to be little doubt that the Learning models provide a better representation of the coffee data than the Markov models, a conclusion already reached by Massy et al. [1970].

The two learning models (Linear Learning and Polya Learning) performed equally well. The p-values of the PL model were higher than those of the LL model for three segments (ALL, HEAVY and LOYAL) and lower for the remaining three. Thus, none of the two models is dominated by the other.

Comparison of Linear Learning (LL), Brand Loyal (BL) and Polya - Learning (PL) Goodness of Fit: Coffee data

SEGMENTS	LL		BL		PL	
	x^2	p	x^2	p	x^2	p
ALL	8.9	.36	16.7	.04	8.05	.43
EXCEPT 100%	7.1	.52	18.1	.02	7.34	.50
HEAVY	6.0	.65	15.2	.06	3.96	.86
LIGHT	5.8	.67	9.1	.37	8.31	.41
LOYAL	13.2	.11	14.4	.08	12.3	.15
NON-LOYAL	3.8	.87	19.0	.02		
DEGREE OF FREEDOM	8		8		8	

Table III.3

While the need to develop alternative models to accommodate the adaptive behavior observed in consumer brand choice is not obvious, the Polyà - Learning model does exhibit some interesting properties:

1) It provides one possible way to accommodate adaptive behavior into stochastic models of brand choice. Although both the Linear Learning and the Polyà - Learning models derive from the same basic assumption (namely, that past purchases influence future ones), the actual feedback mechanisms are different. As indicated by its name, the Linear Learning model assumes that the adaptive process is linear in the probabilities. While linearity is not an unreasonable assumption in many situations, there are times where others might be appropriate. The Polyà - Learning model is an example of such a "non-linear" learning.

2) The limiting form of the Polyà distribution is the so-called negative binominal distribution (see Feller [1957], p. 143), which is the backbone of both Bass' [1974] brand choice and Ehrenberg's [1972] purchase incidence models. While both authors have justified their use of the negative binomial distribution on grounds other than adaptive behavior, the good fit they observed could be explained in terms of learning behavior.

CONCLUSION.

The purpose of this chapter was to present and discuss a new model of brand choice based on learning assumptions. The model was a purchase-to-purchase pure brand choice model which allows an individual's purchase probability to vary with his purchase history. Non-stationarity of purchase probability was assumed to be caused by "learning" effects.

The definition of brand loyalty in terms of learning and forgetting effect introduced a fine distinction between disloyal behavior and mere brand-switching. For the coffee data, it was shown that a LIGHT consumer who switched brands was less loyal than a HEAVY consumer who did so. The latter was more likely to return to his favorite brand as evidenced by the low value of the forgetting parameter.

The next three chapters are devoted to the study of consumer brand-switching behavior in multi-brand markets.

CHAPTER IV

A MODEL OF CONSUMER BRAND-SWITCHING: MATHEMATICAL PRELIMINARIES

The previous two chapters dealt with consumer brand choice as opposed to consumer brand-switching. The stochastic models developed in chapter III are limited to analysing consumer brand choice since by construction they collapse the market into two mutually exclusive categories: brand 1, or the brand under study, and brand 0, a catch all category for the remaining brands. For these models, the variable of interest is the probability that an individual chooses the brand under study at a particular purchase occasion. The probability of that individual purchasing any other specific brand is not known. What is known is just the individual's probability of not purchasing the brand under study. As a result, those models are inadequate for the study of consumer brand-switching behavior.

In contrast, this chapter develops a general class of brand-switching models that acknowledge both the stochastic and deterministic features of the brand-switching phenomenon. We will introduce a class of models which implies aggregate consumer brand-switching and repeat

purchase probabilities, but also directly incorporates the impact of a priori relevant variables in its structure. This approach, while preserving some of the features associated with brand choice models, will allow researchers (with more faith in human rationality) to include in the model variables of managerial and behavioral significance.

4.1 Problem Definition

Given a finite set of n brands, which includes all brands from which a given customer group makes its purchases, suppose that for each pair (i,j) of brands we are given a "datum" A_{ij} representing the similarity, substitutability, association or in general proximity between them. This datum may be a function of the variables controlled by the sellers of the brands, e.g., the sellers' advertising and promotional expenditures, the price of the brands, the reputation of the company etc... It may also be a function of the perceived similarities between the brands as derived from standard composition (multi-attribute attitude approach) or decomposition procedures (multi-dimensional scaling). Indeed, A_{ij} could be a function of both perceived and actual differences between the brands.

We postulate the existence of some functional relation between the observed aggregate brand-switching data, say P_{ij} (for the proportion of consumers purchasing brands i and j on two successive purchase occasions) and the similarity measure A_{ij} so that a general class of brand-switching models can be written in the form:

$$(4.1) \qquad P_{ij} = f_{ij}\,(A_{ij}), \qquad i,j = 1, \ldots, n$$

for some function f_{ij}. In this section, we do not address ourselves to the problem of specifying meaningful general forms for the similarity measure A_{ij}. This is dealt with in the next chapter. Our problem instead is that of deducing from some weak assumptions a functional form for f_{ij} given A_{ij}. Thus, we will not derive any specific results about brand-switching behavior, but rather, some mathematical preliminaries.

4.2 Mathematical Development

Let

$B = \{1, 2, \ldots, n\}$ be a set of n brands

P_{ij} = Predicted proportion of consumers buying brand i and j on two successive purchase occasions. This proportion can be interpreted as the probability of choosing at random from the population of consumers, a consumer who purchased brand i and j on two adjacent purchase occasions.[1]

t_{ij} = Observed proportion of consumers buying brand i and j on two successive purchase occasions.

$m_{it} \doteq \sum_j t_{ij}$ = Observed proportion of consumers who bought brand i on purchase occasion 1.

$m_{j2} \doteq \sum_i t_{ij}$ = Observed proportion of consumers who bought brand j on purchase occasion 2.

$i, j \in B$

---

(1) As no confusion can arise, it is convenient to delete the time subscript form the P_{ij}'s. A more rigorous notation would be $P_{(i,1)(j,2)}$ to reflect the time dependence (purchase occasion) of the proportions.

We now present two alternative methods that can be used to specify the function f_{ij}. The first one is based on the maximum likelihood principle.The second one makes use of the concept of entropy developed by information theorists (Shannon [1949] and recently applied in the social sciences (see e.g Theil [1967] and Wilson [1970]).

4.2.1 The Maximum Likelihood Solution.

Suppose we have available some panel data involving a total sample of N individuals. If n_{ij} denotes the observed number of consumers purchasing brands i and j on two successive purchase occasions, then we must have:[2]

$$\sum_{i,j} n_{ij} = N. \tag{4.2}$$

We may wish to make inferences from the sample results as to the distribution of brand-switching for the whole population from which our sample was drawn. Since the random variables n_{ij} (i,j=1,...,n) follow a multinomial distribution, the likelihood function is proportional to :[3]

---

(2) For simplicity's sake, the number of individuals in the panel is assumed to remain constant over time.

(3) For the multinomial distribution to hold exactly, individuals must make purchase decisions independently of one another. The n_{ij}, however, are dependent because of (4.2).

$$(4.3) \qquad L = \prod_{i,j} P_{ij}^{n_{ij}} .$$

If the similarity measure A_{ij} is a function of K parameters we can write:

$$(4.4) \qquad P_{ij} = f_{ij} [A_{ij} (\Theta_1, \ldots, \Theta_K)] \qquad i,j = 1, \ldots, n$$

where the Θ_k are parameters to be estimated from the data. We assume[4]:

$$(4.5) \qquad f_{ij} [A_{ij} (\Theta_1, \ldots, \Theta_K)] = a_i b_j A_{ij} (\Theta_1, \ldots, \Theta_K)$$

$$\doteq a_i b_j A_{ij} (\tilde{\Theta})$$

where a_i and b_j are additional parameters to be estimated along with the Θ_k and $\tilde{\Theta}$ denotes the vector of parameters $\{\Theta_1, \ldots, \Theta_K\}$. We want to maximize equation (4.3) or equivalently, its logarithm, with the P_{ij}'s expressed in terms of (4.5) subject to the restrictions:

$$(4.6) \qquad \sum_{i,j} P_{ij} = 1$$

$$(4.7) \qquad P_{ij} \geq 0 .$$

---

(4) Motivation for this assumption is presented in Appendix A where equation (4.5) is shown to result from four weak assumptions.

Letting $t_{ij} = n_{ij}/N$ and λ be a lagrangean multiplier, we find values of $\{a_i\}_{i=1}^{n}$, $\{b_j\}_{j=1}^{n}$, $\{\Theta_k\}_{k=1}^{K}$ and λ which maximize

$$(4.8) \qquad H = \sum_{i,j} t_{ij} \log P_{ij} - \lambda[\sum_{i,j} P_{ij} - 1]$$

$$= \sum_{i,j} t_{ij} [\log a_i + \log b_j + \log A_{ij}(\tilde{\Theta}) -$$

$$\lambda[\sum_{i,j} a_i b_j A_{ij}(\tilde{\Theta}) - 1].$$

Setting the partial derivatives of H, with respect to each model parameter, to zero, i.e.,

$$(4.9) \qquad \frac{\partial H}{\partial a_i} = 0 \quad \forall_i$$

$$(4.10) \qquad \frac{\partial H}{\partial b_j} = 0 \quad \forall_j$$

$$(4.11) \qquad \frac{\partial H}{\partial \Theta_k} = 0 \quad \forall_k$$

we obtain respectively:

$$(4.9a) \qquad \sum_j t_{ij}/a_i - \lambda \sum_j b_j A_{ij}(\tilde{\Theta}) = 0 \qquad i = 1, \ldots, n$$

$$(4.10a) \qquad \sum_i t_{ij}/b_j - \lambda \sum_i a_i A_{ij}(\tilde{\Theta}) = 0 \qquad j = 1, \ldots, n$$

(4.11a) $$\sum_{i,j} \frac{t_{ij}}{A_{ij}(\tilde{\Theta})} \frac{\partial A_{ij}(\tilde{\Theta})}{\partial \Theta_k} = \lambda \sum_{i,j} a_i b_j \frac{\partial A_{ij}(\tilde{\Theta})}{\partial \Theta_k} \qquad k=1,\ldots, K.$$

Using (4.6), (4.9a) - (4.11a) become

(4.12) $$\lambda \sum_j P_{ij} = \sum_j t_{ij} \qquad \forall_i.$$

(4.13) $$\lambda \sum_i P_{ij} = \sum_i t_{ij} \qquad \forall_j.$$

(4.14) $$\sum_{i,j} \frac{P_{ij}}{A_{ij}(\tilde{\Theta})} \frac{\partial A_{ij}(\tilde{\Theta})}{\partial \Theta_k} = \sum_{i,j} \frac{t_{ij}}{A_{ij}(\tilde{\Theta})} \frac{\partial A_{ij}(\tilde{\Theta})}{\partial \Theta_k} \quad \text{for each } k.$$

Since $\sum_{i,j} P_{ij} = \sum_{i,j} t_{ij} = 1$, we have

(4.15) $$\lambda = 1.$$

Further manipulation of (4.9a)-(4.14) yields:[5]

(4.16) $$P_{ij} = \alpha_i \beta_j m_{i1} m_{j2} A_{ij} (\tilde{\Theta})$$

where

(4.17) $$\alpha_i = [\sum_j \beta_j m_{j2} A_{ij} (\tilde{\Theta})]^{-1}$$

(4.18) $$_j = [\sum_i \alpha_i m_{i1} A_{ij} (\tilde{\Theta})]^{-1}$$

---

(5) The P_{ij}'s will all be non-negative as long as the similarity measure $A_{ij}(\tilde{\Theta})$ is itself non-negative.

and as before

(4.19) $$m_{i1} = \sum_j t_{ij}$$

(4.20) $$m_{j2} = \sum_i t_{ij}.$$

The normalizing constants α_i and β_j are such that the maximum likelihood constraints expressed in (4.12) and (4.13) are mechanically satisfied as can be readily checked by summing (4.16) with respect to j and i respectively. The maximum likelihood method not only specifies the functional form of the a_i's and b_j's, but also indicates how to estimate the Θ_k's: just solve equation (4.14) for Θ_k, k = 1, ..., K.

We will now present an alternative derivation of equations (4.16)-(4.18) which makes uses of the concept of entropy.

4.2.2 Maximum Entropy Solution

The concept of entropy was first applied in the study of thermodynamics (Jaynes [1957]) and has found more recent application in the field of information theory (Khinchin [1957]) where researchers are concerned with the measuring of the amount of information conveyed by a given message. In addition, it has recently aroused additional

interest in the social sciences, either as a descriptive measure of uncertainty such as in Theil [1972] and Carman [1970], or as a more subjective concept used by the analyst as a model-building tool to maximize the use of information available to him (see e.g., Wilson [1970] and Herniter [1973]).

4.2.2.1 Entropy and Information

To illustrate the notion of entropy, consider an event E with probability of occurrence p. At some point in time, we receive a reliable message stating that E in fact occurred. The question is: how should one measure the amount of information conveyed by this message? Suppose that p is close to one (e.g., p = .95). Then, one may argue, the message conveys little information, because it was virtually certain that E would take place. But suppose that p = .01, so that is is almost certain that E will not occur. If E nevertheless does occur, the message stating this will be unexpected and hence contains a great deal of information. These intuitive ideas suggest that to measure the information received from a message in terms of the probability p that prevailed prior to the arrival of the message, we should select a decreasing function of p. The function proposed by Shannon [1949] is:

$$(4.21) \qquad h(p) = \log \frac{1}{p} = -\log p$$

which decreases from ∞ (infinite surprise and hence infinite information when the probability prior to the message is zero) to zero (zero information when the probability is one). In this instance the logarithmic definition for information in (4.21) can be shown to be the only possible definition where certain simple axioms are accepted (see e.g. Khinchin [1957]).

4.2.2.2 The Entropy of a Distribution

The information received from the message which states that event E has occurred is not the same as the information concerning the complementary event that E has failed to occur. If p is the probability of E, the information provided by the latter message is:

$$h(1-p) = -\log(1-p).$$

Therefore, as far as event E is concerned, the information to be received is either $h(p)$ or $h(1-p)$ and we do not know which as long as the message of occurrence or non-occurrence has not been received. However, we can compute the expected information content of this message prior to its arrival, i.e.,

$$(4.22) \qquad H = p\,h(p) + (1-p)\,h(1-p)$$
$$= -p\,\log(p) - (1-p)\,\log(1-p).$$

The function H is also known as the entropy of any distribution that assigns probabilities p and (1-p) to two different events. It follows directly from (4.22) that the entropy function is symmetric in p and 1-p. It is non-negative, takes the zero value at p=0 and p=1 and reaches a maximum at p=1/2. Finally, the entropy function in (4.22) can be extended to the case of n events $E_1, \ldots, E_n$ with probabilities $p_1, \ldots, p_n$:

$$(4.23) \qquad H = \sum_{i=1}^{n} p_i\,h(p_i) = -\sum_{i=1}^{n} p_i\,\log(p_i)$$

4.2.2.3 Entropy as a Model-Building Tool

This section presents the ideas of Jaynes [1957] and will enable us to offer a second and more useful interpretation of the concept of entropy. Let X be a random variable which can take on values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, \ldots, p_n$. The probabilities are not known. All we know is the expectation of some function f(X):

$$(4.24) \qquad E[f(X)] = \sum_i p_i\,f(x_i)$$

and

(4.25) $\sum_i p_i = 1, \quad p_i \geq 0 \quad \forall_i.$

Given this information only, what is our best estimate of the probability distribution p_i? Jaynes writes:

> "Just as in applied statistics, the crux of a problem is often the devising of some method of sampling that avoids bias, our problem is that of finding a probability assignment which avoids bias while agreeing with whatever information is given. The great advance provided by information theory lies in the discovery that there is a unique, unambiguous criterion for the 'amount' of uncertainty represented by a discrete probability distribution, which agrees with our intuitive notions that a broad distribution represents more uncertainty that does a sharply peaked one, and satisfies all other conditions which make it reasonable".

This criterion is the one expressed in (4.23). Jaynes then writes:

> "It is now evident how to solve our problem; in making inferences on the basis of partial information, we must use that probability distribution which has maximum entropy subject to whatever is known. This is the only unbiased assumption we can make; to use any other would amount to arbitrary assumption of information which by hypothesis we do not have".

Thus, to solve the problem posed above, we simply have to maximize entropy in (4.23) subject to equation (4.24) and (4.25), which represent what we know. This gives:

$$(4.26) \qquad p_i = \frac{\exp[-\mu f(x_i)]}{\sum_j \exp[-\mu f(x_j)]}$$

where μ is a lagrangean multiplier associated with (4.24).

In the marketing field, Herniter [1973] provides a good illustration of the use of entropy as a model-building tool to maximize the use of information available to the researcher. Building on the ideas presented above, he developed a probabilistic model of consumer purchase behavior. The model is completely determined by specifying only the market shares. All other brand selection statistics, such as repeat purchase probabilities (P_{ii}) and brand-switching probabilities (P_{ij}, $i \neq j$) are derived from the model. The assumptions of Herniter's model are best expressed in his own words:

> "It is assumed that each consumer has a set of preferences for the brands in the market, and there is a distribution of preferences over the population. The probability of a consumer purchasing a particular brand is numerically equal to her preference for the brand. Rather than specifying the joint distribution of preferences and fitting the parameters to empirical data, the concept of entropy is employed and the distribution is selected that maximizes the entropy of the system subject only to the empirical market share values."

By doing so, Herniter developed a genuine heterogeneous consumer behavior model based on micro-theoretic

assumptions. For all its elegance, his approach did not allow him to build into the model exogeneous variables of behavioral or managerial interest. The method to be presented below overcomes this deficiency but offers no postulates for the underlying individual behavior.

4.2.2.4 Maximum Entropy Solution

Let (X,Y) be a pair of random variables which can take on values (x_1,y_1), (x_2,y_1),....., (x_n,y_n) with probabilities p_{11}, p_{21},....., p_{nn}. In the brand-switching context, the random variable X= i will represent a purchase of brand i at purchase occasion t-1, and similarly, Y=j will represent a purchase of brand j at purchase occasion t, i, j=1,...,n. The probabilities are not known. However, suppose we know the expected value of some function of the random variables X and Y, say A(X,Y):

$$(4.27) \qquad E[A(X,Y)] = \sum_{i,j} P_{ij} A[X=i,\ Y=j].$$

To ease notation, let us write A_{ij} for A(X=i, Y=j). Suppose there is available some panel data, involving a sample size of N individuals. If n_{ij} denotes the observed number of consumers purchasing brands i and j on two adjacent purchase occasions, then we must have:

(4.28) $$\sum_{i,j} n_{ij} = N.$$

Let $$t_{ij} = n_{ij}/N.$$

As before, we would like to make inferences from the sample results as to the distribution of brand-switching for the whole population from which our sample was drawn. Following Jaynes [1957], we shall use that probability distribution P_{ij} which has maximum entropy under the known constraints. That is, we want to maximize the entropy of the joint probability distribution P_{ij}:

(4.28) $$H = -\sum_{i,j} P_{ij} \log (P_{ij})$$

subject to

(4.27) $$E[A(X,Y)] = \sum_{i,j} P_{ij} A_{ij}$$

(4.29) $$\sum_{i} P_{ij} = \sum_{i} t_{ij} \doteq m_{j2} \qquad j=1,\ldots, n$$

(4.30) $$\sum_{j} P_{ij} = \sum_{j} t_{ij} \doteq m_{i1} \qquad i=1,\ldots, n$$

where the symbol $\doteq$ means equality by definition. Constraints (4.29) and (4.30) require the predicted marginals (which can be interpreted as market shares) to be equal to the observed marginals.

The mathematical problem represented by (4.28) to (4.30) is that of finding the maximum of a function subject to a set of equality constraints. This may be solved through the method of lagrange multipliers as follows:

Define the lagrangean

$$(4.31) \qquad L = \sum_{i,j} p_{ij}\log(P_{ij}) - \sum_i \lambda_i(\sum_j P_{ij} - m_{i1})$$

$$- \sum_j \mu_j(\sum_i P_{ij} - m_{j2}) - \beta[\sum_{i,j} P_{ij} A_{ij} - E(A_{ij})].$$

where $\{\lambda_i\}$, $\{\mu_j\}$ and β are lagrangean multipliers, and $E(A_{ij})$ is known. The P_{ij} that maximize L are solution of:

$$(4.32) \qquad \frac{\partial L}{\partial P_{ij}} = \frac{\partial L}{\partial \lambda_i} = \frac{\partial L}{\partial \mu_j} \quad \frac{\partial L}{\partial \beta} = 0, \quad i,j=1,\ldots, n.$$

From (4.32):

$$\frac{\partial L}{\partial P_{ij}} = -\log(P_{ij}) - 1 - \lambda_i - \mu_j - \beta A_{ij}$$

or

$$(4.33) \qquad P_{ij} = \exp[-1 - \lambda_i - \mu_j - \beta A_{ij}].$$

Substituting in (4.30) for P_{ij} yields

$$\sum_j \exp[-1 - \lambda_i - \mu_j - \beta A_{ij}] = m_{i1}$$

or $$\exp(-\lambda_i) = m_{i1} \{\sum_j \exp[-1 - \mu_j - \beta A_{ij}]\}^{-1} \qquad i=1,\ldots,n.$$

Substituting in (4.29) for P_{ij} similarly yields:

$$\exp(-\mu_j) = m_{j2}\{\sum_i \exp[-1 - \lambda_i - \beta A_{ij}]\}^{-1} \qquad j=1,\ldots,n.$$

Upon defining

$$\bar{a}_i = \exp(-\lambda_i)/m_{i1}$$

$$\bar{b}_j = \exp(-\lambda_j)/m_{j2}$$

one obtains:

(4.34) $$P_{ij} = \bar{a}_i \bar{b}_j m_{i1} m_{j2} \exp(-\beta A_{ij})$$

where

(4.35) $$\bar{a}_i = [\sum_j m_{j2} \bar{b}_j \exp(\beta A_{ij})]^{-1}$$

(4.36) $$\bar{b}_j = [\sum_i m_{i1} \bar{a}_i \exp(-\beta A_{ij})]^{-1}.$$

Comparing equations (4.16)-(4.18) with equations (4.34)-(4.36) shows that the two approaches yield similar "looking" equations.[6] This is not surprising. It has always been recognized that

---

(6) There are differences, however, due to the logarithmic definition of entropy. This explains the presence of the exponential function in (4.34). The two equations, however, are completely different in terms of their mathematical structure.

there is a close connection between entropy maximizing methods and maximum likelihood methods (see e.g., Hyman [1969] and Wilson [1970]), since entropy is the negative of the expected value of the log-likelihood function. The former method has the advantage of flexibility.

We have shown above that the entropy maximizing procedure is a way of obtaining a probability distribution taking account of all the information available. If the available information is modified or added to (e.g., if some other constraints of the type expressed in (4.27) are shown to hold) then the estimate will be changed to reflect the knowledge of the new information. Finally, the concept of entropy will also provide us with a theoretical justification for the problem of specifying an operational form for A_{ij}, which is the subject of the next chapter.

SUMMARY

This chapter has laid the theoretical base for the development and empirical investigation of the brand-switching models offered here. A model building strategy has been outlined that explicitly incorporates into the model variables of behavioral and managerial significance to brand choice decision making.

Three alternative derivations were presented. Of particular importance for the remainder of this research is the entropy maximizing procedure that provides a set of rules for model construction which will guarantee compatibility with known information and an internal consistency which is not otherwise easily achievable. In the next chapters, the concept of entropy will be used as a descriptive statistic to help us interpret the empirical results. The maximum likelihood derivation will prove useful for estimation purposes.

The next chapter provides some operational formulations for the similarity measure $A_{ij}(\tilde{\Theta})$ that will allow a detailed empirical comparison of the model's performance in terms of goodness of fit and diagnostic potential with that of alternative brand-switching models.

CHAPTER V

THE DETERMINANTS OF BRAND SWITCHING BEHAVIOR

Introduction

In the last chapter, a model testing strategy has been outlined that will allow the researcher to express consumer brand-switching probabilities in terms of variables of managerial significance to brand choice decision making. The purpose of this chapter is to develop a theory of consumer brand-switching behavior based on sound psychological premises and empirically test it with the procedures developed in chapter IV.

To this end, the central concepts underlying the proposed theory will be presented in the next section. The mathematical model developed in the preceding chapter will then be operationalized, and the empirical data described. The chapter will end with a discussion of the test results and some concluding comments about the general pertinence of the theory.

5.1 Stochastic and Deterministic Theories

Best's [1976] study not withstanding, researchers have traditionally treated brand choice behavior for an individual consumer as being completely stochastic or

entirely deterministic. Quite recently, however, some studies[1] have strongly suggested that brand choice is mainly a stochastic process and that the outcome of any particular choice decision cannot be predicted precisely. Their strong results have caused stochastic models to upstage the traditional brand choice correlates studies and have sanctioned the new shift toward predicting rather than explaining the actions of consumers. Indeed, the fundamental premise of social psychology which postulates that all behavior is rational and therefore can be explained has been put aside if not put down.

Clearly, consumer behavior is both a stochastic and a cognitive (deterministic) process. Stochastic, since brand selection on a given trial cannot be predicted precisely; and cognitive because the steady state choice probabilities observed over a sequence of choices reveal a choice pattern consistent with the consumers perceptions, preferences and beliefs toward a particular set of brands.

The next section will develop and discuss a cognitive process model which incorporates the stochastic reality of the consumer choice process.

---

(1) See e.g. Herniter [1973] and Bass [1974].

5.2 A Joint Space Theory of Brand Choice

A useful approach to examining a brand's competitive position is to consider a set of brands as points located in a space where the axes are defined in terms of the "relevant" characteristics of the brands. The concept of a brand as a position on a set of attributes has been suggested by economists (Lancaster [1966]), social psychologists (Rosenberg [1960], Fishbein [1963]) and mathematical psychologists (Shepard [1962], Kruskal [1964]). Numerous exploratory and a few significant applied marketing studies have been reported in the marketing literature. They often appear under the title "Market Structure Analysis" (see e.g. Stefflre [1968]), "Perceptual Mapping" or "Joint-Space Theory" (see e.g. Best [1976]).

In the brand choice context, a joint space configuration consists of a set of brands and consumers positioned in the same "perceptual" space. The dimensions of this perceptual space reflect the consumers' perception of those attributes which they use in making discriminate judgments among the brands.

Consumers may be characterized as having a unique stimulus or ideal point in the perceptual space. The interpretation holds that consumers prefer some particular combination of values on the perceived dimensions to all other combinations. When both the consumers' ideal points

and the brands' characteristics are mapped onto the same "perceptual" space, the latter is often referred to as a "joint-space".

The central concept underlying the joint space theory of brand choice is that of cognitive consistency. Consumers strive to maintain a cognitive equilibrium between their perceptions and preferences of the brands, on the one hand, and their actual brand choice, on the other hand. While they may occasionally alternate between brands and exhibit varying degrees of brand loyalty, they do so in a "long-term rational" fashion rather than in a purely random one. That is, they tend to organize their choice behavior so as to achieve a cognitive equilibrium between their perceptions, preferences and brand choice.

Joint space configurations[2] provide a geometric picture of consumers' perceptions and preferences for a particular set of brands. Inter-brands and ideal point - brand distances are meant to reflect consumers cognitive structure in terms of perceptions and preferences. A theory of joint space brand choice postulates that:

---

(2) A joint-space configuration for eight brands of soft brands of soft drink is shown in figure V.1.

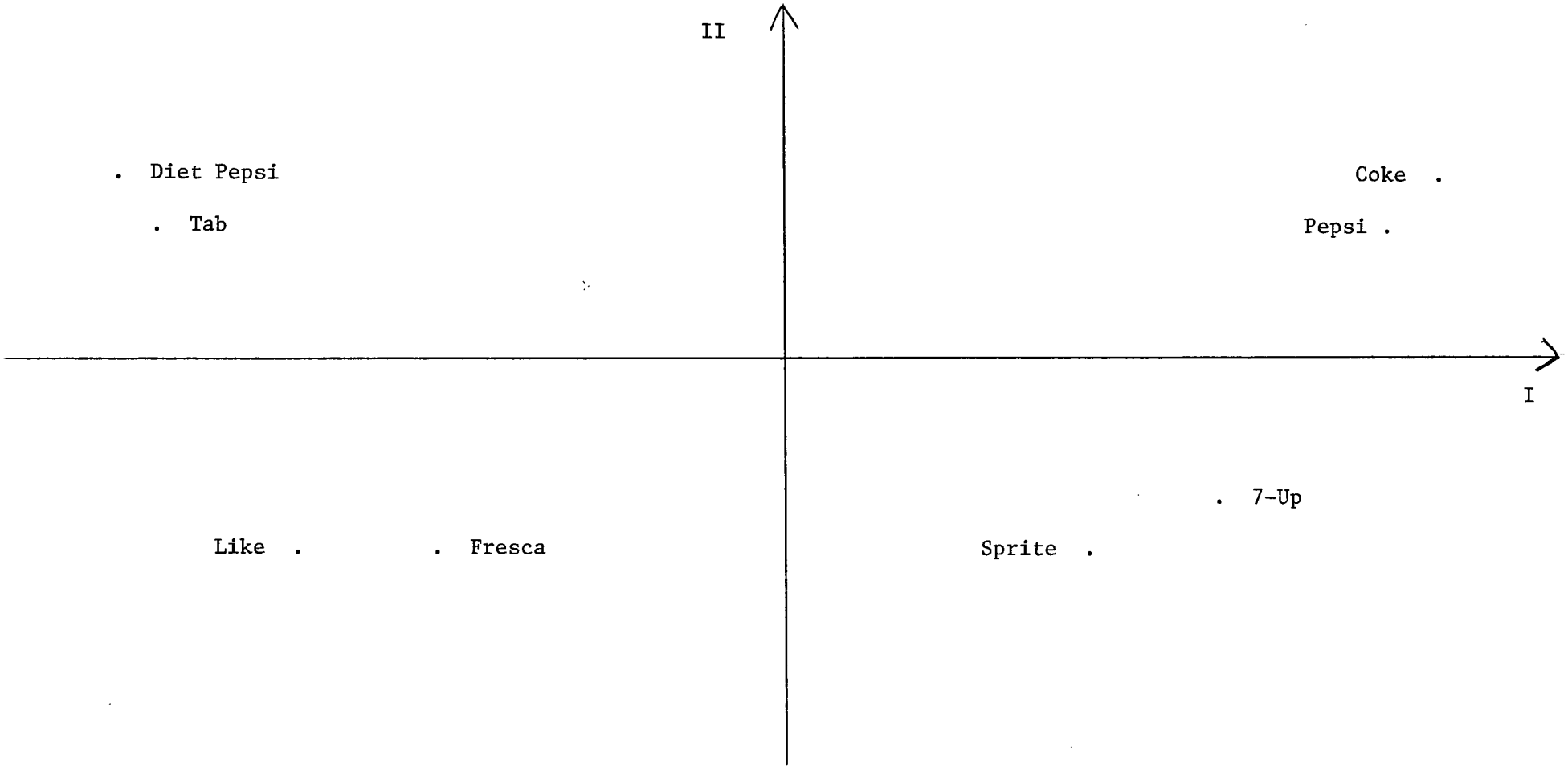


FIGURE V.1 Position of the Eight Brands in 2-Dimensional Discriminant Configuration

i) Brand choice probabilities are related functionally to the distance between a consumer's ideal point and each brand in a particular choice set. The smaller the distance between a particular brand and the consumer's ideal point, the greater the likelihood of it being chosen on a particular choice occasion. Moreover, the consumers may differentially weight the dimensions of the perceptual space in terms of their relative importance to them in an evaluative context. The distance of specific brands from his ideal point is assumed to reflect the differential weighting which he applies to the dimensions of interest.

ii) Brand-switching[3] probabilities are related functionally to the distances between the brands in the perceptual space. The premise is that consumers will tend to switch to similar rather than to dissimilar brands. The smaller the perceived psychological distance between the brands, the greater their similarity or substitutability, hence the greater the switching between them.

---

(3) Brand choice probability is defined as the probability that a consumer selects a particular brand on a given choice occasion.

Brand-switching probability is defined as the joint probability that a consumer selects any two distinct brands on two adjacent purchase occasions.

To test these two hypotheses, the model developed in the preceding chapter will be applied and elaborated upon in the next section.

5.3 Model Development

In model form, brand-switching probabilities can be written as:

(5.1) $P_{ij} = f_{ij}(D_{ij}, d_{it})$, $i,j = 1, \ldots, n$,

where

P_{ij} = joint probability that a consumer chooses brands i and j on two successive purchase occasions

D_{ij} = "perceptual" distance between brands i and j in the joint space

d_{it} = "perceptual" distance between the consumers' average ideal point and brand i in the joint space, at purchase occasion t

f_{ij} = some mathematical function

n = number of brands in the market.

Similarly, brand-choice probabilities can be expressed as:

$$m_{i1} \overset{def}{=} \sum_j P_{ij} = g(d_{i1})$$

(5.2) $$m_{j2} \overset{def}{=} \sum_i P_{ij} = g(d_{j2})$$

for some mathematical function g.

Equations (5.1) and (5.2) merely restate in symbolic notation the two hypotheses described earlier, namely:

i) brand-switching probabilities (P_{ij}, $i \neq j$) depend on inter-brand distances (D_{ij}) in the joint space.

ii) brand choice probabilities ($\sum_j P_{ij}$, $\sum_i P_{ij}$) and repeat purchase probabilities (P_{ii}) vary inversely with the distance between the brands and the consumers' average ideal point (d_{it}).

An extensive rationale was provided in chapter IV to justify the use of the following mathematical relation between the dependent variables [brand-switching (P_{ij}, $i \neq j$), repeat purchase (P_{ii}) and brand choice (m_{it}) probabilities] and the independent variables (joint space distance D_{ij} and d_{it}):

(5.3) $$P_{ij} = a_i b_j \, g(d_{i1}) \, g(d_{j2}) \, h(D_{ij}, d_{it}) \qquad i,j = 1,\ldots,n$$

where

(5.4) $$a_i = [\sum_{j=1}^{n} b_j \, g(d_{j2}) \, h(D_{ij}, d_{it})]^{-1} \qquad i = 1,\ldots,n$$

(5.5) $$b_j = [\sum_{i=1}^{n} a_i \, g(d_{i1}) \, h(D_{ij}, d_{it})]^{-1} \qquad j = 1,\ldots,n$$

g, h = some mathematical functions to be specified later.

The normalizing constants a_i and b_j are such that if equation (5.3) is summed with respect to either i or j one obtains:

$$\sum_j P_{ij} \doteq m_{i1} = g(d_{i1}) \tag{5.6}$$

and

$$\sum_i P_{ij} \doteq m_{j2} = g(d_{j2}). \tag{5.7}$$

The above equations show that brand choice probabilities are functions of the distances between the brands and the ideal point as previously hypothesized. Equation (5.3) conveniently expresses brand-switching and repeat purchase probabilities as the product of brand choice probabilities and a function that depends on joint space distances. The last step in the model-building procedure consists of providing operational definitions for both the functions g and h. This task is accomplished in the next two subsections. When this is done, we shall have an operational model with which to test our joint space theory of brand choice and brand-switching behavior.

5.3.1 Specification of the function g

While marketing theory suggests that brand choice probability should vary inversely with the distance between the brands and the consumers' average ideal point, it stops short of specifying the appropriate mathematical relation between these two quantities[4]. As there is a priori no reason to suspect that one mathematical function is better than the other, several functional forms will be examined and fitted to the empirical data. They include:

(5.8) Hyperbolic model:

$$g(d_{it}) = \frac{d_{it}^{\lambda}}{\sum_j d_{jt}^{\lambda}} \qquad \begin{array}{l} i = 1,\ldots, n \\ t = 1,2 \end{array}$$

(5.9) Exponential model:

$$g(d_{it}) = \frac{\exp[\lambda\, d_{it}]}{\sum_j \exp\,[\lambda d_{jt}]} \qquad \begin{array}{l} i = 1,\ldots, n \\ t = 1,2 \end{array}$$

(5.10) Polynomial model:

$$g(d_{it}) = \frac{\sum_{k=1}^{k} \lambda_k\, d_{it}^{\,k}}{\sum_j \sum_k \lambda_k\, d_{jt}^{\,k}} \qquad \begin{array}{l} i = 1,\ldots, n \\ t = 1,2 \end{array}$$

---

(4) Best [1976] provided some empirical evidence suggesting that different functions may be appropriate for different individuals.

The parameter(s) λ measures the extent to which brand choice probabilities vary with perceptual distances between the brands and the consumers' average ideal point. Since brand choice probability is assumed to be inversely related to perceptual distance, the parameter λ should take on non positive values. When λ vanishes, the hyperbolic and exponential models yield the same limiting form: $m_{it} = 1/n$, i.e., all brands are equiprobable. As the absolute value of λ becomes large, brand choice probability decreases more sharply as perceptual distance increases. In the limit, the brand which is closest to the ideal point is chosen with probability one and all the other brands are assigned zero probability mass. In other words, the greater the absolute value of λ, the greater the brand loyalty to the most preferred brand.

The polynomial model was included to accommodate possible non-monotonic relationships between brand choice probability and perceptual distances.

5.3.2 Specification of the function h

To specify the function h that controls the extent to which brand-switching probabilities vary with joint space distances, we will use the concept of entropy

as a model-building tool[5]. Our problem is that of finding a probability assignment which avoids bias while agreeing with whatever information is given. In our particular case, we hypothesized that both brand choice and repeat purchase probabilities were functions of the distance between the brands and the consumers' ideal point. We also held that brand-switching probabilities vary inversely with inter-brand joint space distances. For simplicity's sake, let us assume that repeat purchase and brand-switching probabilities are proportional to (normalized) joint space distances, that is:

(5.11) $P_{ii} = k_1 \, d_{i2}^{\gamma}, \; i = 1,\ldots,n$

and

(5.12) $P_{ij} = k_2 \, D_{ij}^{\mu} \qquad i \neq j, \quad i,j = 1, \ldots,n$

where k_1 and k_2 are proportionality constant,

γ and μ are distance parameters that control the extent to which brand-switching probabilities vary with joint space distances.

As before, we require that:

(5.13) $m_{i1} \doteq \sum_j P_{ij} = g(d_{i1})$

---

(5) The use of entropy as a model-building tool was reviewed in chapter IV.

(5.14) $$m_{j2} \doteq \sum_j P_{ij} = g(d_{j2})$$

Let us assume that equations (5.11) to (5.14) represent all the information we have about consumer brand-switching behavior. In making inferences about the P_{ij}'s on the basis of partial information, we must, as for Jaynes [1957], use that probability distribution (P_{ij}) which has maximum entropy subject to whatever information is known. Thus, to solve the problem posed above, we simply have to:

(5.15) $$\text{Maximize } E = - \sum_{i,j} P_{ij} \log P_{ij}$$

subject to the constraints expressed by (5.11) to (5.14). This maximization problem can be solved through the method of Lagrange multipliers as follows:

Define the Lagrangean as:

(5.16) $$L = - \sum_{i,j} P_{ij} \log P_{ij} - \sum_i \lambda_i \left[\sum_j P_{ij} - g(d_{i1})\right]$$

$$- \sum_j \mu_j \left[\sum_i P_{ij} - g(d_{i2})\right] - \sum_i \beta_i \left(P_{ii}\, d_{i2}^{-\gamma} - k_1\right)$$

$$- \sum_{\substack{i,j \\ i \neq j}} \gamma_{ij} \left(P_{ij}\, D_{ij}^{-\mu} - k_2\right)$$

where

δ_{ij} is the Kroenecker delta, i.e.

$$\delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

$\{\lambda_i\}$, $\{\mu_j\}$, $\{\beta_i\}$ and $\{\alpha_{ij}\}$ are the Lagrangean multipliers associated with the four sets of constraints (5.11 to 5.14).

Upon setting the partial derivatives of L equal to zero and solving the resulting equations, one obtains:

$$P_{ij} = a_i b_j \, g\,(d_{i1}) \, g\,(d_{i2}) \exp\left[- \beta_i \, \delta_{ij} \, d_{i2}^{-\gamma} - \alpha_{ij} \, D_{ij}^{-\mu}\right] \tag{5.17}$$

where the terms a_i and b_j are defined as in (5-4) - (5-5) with the obvious modifications. When either one of the hyperbolic (5.8) exponential (5.9) or polynomial (5.10) models is substituted in equation (5.17) in lieu of the brand choice functions $g\,(d_{it})$, one obtains a fully operational model which can be used to submit the joint space brand-switching theory to an empirical test. For testing purposes, simplified versions of the model expressed in (5.17) will be submitted to empirical data. The need for simpler models arises from the desire to limit the number of parameters to be estimated from the data.

As it stands, the model (5.17) is overparameterized: $(n^2 + 3)$ parameters $(\lambda, \gamma, \mu, \{\beta_i\}_{i=1}^{n}$ and $\{\alpha_{ij}\}_{\substack{i,j=1 \\ i \neq j}}^{n})$ for just $(n^2 - 1)$ degress of freedom. To reduce the number of parameters to be estimated without drastically altering the model's general character, some parameters had to be removed or set equal to one another. From this simplification process, a number of models were retained for testing purposes. For convenience, they are summarized in table V.4 and briefly discussed in the next section.

5.3.3 Models for testing purposes

The model are classified in three groups, in order of increasing complexity.

a) Group 1

This is the simplest class of models to be entertained in this chapter. Four different version will be considered:

SUMMARY OF MODELS

Basic Equation: $P_{ij} = a_i b_j m_{i1} m_{j2}\ h(D_{ij}, d_{it})$ $\quad i,j = 1,\ldots,n$

where $a_i = [\sum_j b_j m_{j2}\ h(.)]^{-1}$ $\quad i = 1,\ldots,n$

$b_j = [\sum_i a_i m_{i1}\ h(.)]^{-1}$ $\quad j = 1,\ldots,n$

MODEL	BRAND CHOICE PROBABILITIES m_{i1}, m_{j2}	$h(D_{ij}, d_{it})$
I.1	Equality between predicted and observed brand choice probabilities is forced by setting m_{it} $(i = 1,\ldots,n,\ t = 1,2)$ equal to the observed brand choice probabilities.	$\exp(\beta\delta_{ij})$
I.2		$\exp(\beta_i\delta_{ij})$
I.3		$\exp[\beta(\delta_{ij} - D_{ij})]$
I.4		$\exp[\beta_i(\delta_{ij} - D_{ij})]$
II	$m_{it} = \dfrac{d_{it}^{\lambda}}{\sum_j d_{jt}^{\lambda}}$ $\quad i = 1,\ldots,n$ $\quad t = 1,2$	$\exp[\beta_i(\delta_{ij} - D_{ij})]$
III.1	$m_{it} = \dfrac{d_{it}^{\lambda}}{\sum_j d_{jt}^{\lambda}}$ $\quad i = 1,\ldots,n$ $\quad t = 1,2$	$\exp[\beta_i(\delta_{ij} d_{i2}^{\mu} - D_{ij})]$
III.2		$\exp[\beta_i(\delta_{ij} d_{i2}^{\mu} - \log(\delta_{ij} + D_{ij}))]$
III.3		$\exp[\mu\delta_{ij} \dfrac{d_{i2}^{\lambda}}{\sum_j d_{j2}^{\lambda}} - \beta_i D_{ij}]$

TABLE V.4

Model version	Functional form for $h(D_{ij}, d_{it})$	$g(d_{it})$
I.1	$\exp[\beta\ \delta_{ij}]$	Equality between predicted and observed brand choice probabilities is forced by setting $g(d_{it})$ in (5.17) equal to the observed brand choice probabilities
I.2	$\exp[\beta_i\ \delta_{ij}]$	
I.3	$\exp[\beta(\delta_{ij} - D_{ij})]$	
I.4	$\exp[\beta_i\ (\delta_{ij} - D_{ij})]$	

where δ_{ij} is the Kroenecker delta i.e.

$$\delta_{ij} = \begin{cases} 1 \text{ when } i = j \\ 0 \text{ when } i \neq j \end{cases}$$

D_{ij} = "perceptual" distance between brands i and j in the joint space. The set of all inter-brands distances is an input to the model, and must be derived externally.[6]

The β's are parameters to be estimated from the empirical data. The distinguishing features of those four simple models are two-fold:

i) No attempt is made to "predict" or "fit" the brand choice probabilities. That is, the observed brand choice probabilities are substituted into equation (5.17) in lieu of $[g(d_{it}),\ t = 1,2]$ so as to force equality between predicted and observed brand choice probabilities.

ii) No attempt is made to "fit" the repeat purchase probabilities. That is no attempt is made to express them as a function of the distance between the brands and the consumer's average ideal point in the joint space.

---

(6) Techniques to derive joint spaces are mentioned in the next section.

It should be noted at this stage that part of the information contained in the joint space is deliberately ignored in the above formulations, in order to focus exclusively on consumer brand-switching. The perceptual distances between the brands and the consumers' average ideal point will prove to be useful predictors for both brand choice and repeat purchase probabilities in later applications. However, it is assumed that brand-switching probabilities depend on the brands' respective position in the point space and not on their position with respect to the consumers' average ideal point[7].

Models belonging to group I can be written as follows:

(5.18) $$P_{ij} = a_i b_j \, t_{i.} \, t_{j.} \, h(D_{ij})$$

where

$t_{i.} = \sum_j t_{ij}$

$t_{.j} = \sum_i t_{ij}$

t_{ij} = observed brand-switching probabilities

P_{ij} = corresponding predicted brand-switching probabilities

$h(D_{ij})$ = either one of the four functions expressed above

---

(7) This is an hypothesis which needs to be empirically verified

If one sums both side of (5.18) with respect to i or j, one obtains:

$$\sum_j P_{ij} = t_{i.}$$

$$\sum_i P_{ij} = t_{.j}$$

In group I, predicted and observed brand choice probabilities are equal by construction. Thus, no attempt is made to express brand choice probabilities in terms of joint space distances, as will be the case for the next group of models.

b) Group II

Models in group II differ from those in group I in that the brand choice probabilities can now be expressed in terms of the joint space distances between the various brands and the consumers' average ideal point. The three mathematical functions already discussed in (5.8) to (5.10) will be fitted to empirical data. While all of the four functional forms that were developed for the function $h(D_{ij})$ could be used in conjunction with any of the three brand choice models described above (exponential, hyperbolic and polynomial), it was decided to confine the subsequent analysis to the more general form afforded by model I.4. That is, models of group II can be written as:

(5.19) $P_{ij} = a_i b_j \, g(d_{i1}) \, g(d_{j2}) \quad \exp\,[\beta_i(\delta_{ij} - D_{ij})]$

where the expressions a_i, b_j, δ_{ij} and D_{ij} are defined above and the $g(d_{it})$ $(i = 1, \ldots, n, t = 1, 2)$ can take one any one of the three functional forms specified in (5.8) through (5.10).

c) Group III

This is the most general group. In this group of models the brand-switching probabilities are made a function of inter-brand perceptual distances, as in group I. Brand choice probabilities are expressed in terms of the distances between the various brands and the customers' average ideal point, as in group II.

The distinguishing feature of models in group III consists of allowing the repeat purchase probabilities (the P_{ii}'s) to depend on the perceptual distances between the brands and the ideal point by analogy with brand choice probabilities.

Four different versions were submitted to an empirical test. They are respectively:[8]

---

(8) Actually, several other versions were empirically tested but due to their poor empirical results, they are not discussed in this study.

$$(5.20) \qquad P_{ij} = a_i b_j \frac{d_{i1}^{\lambda} d_{j2}^{\lambda}}{(\sum_k d_{k1}^{\lambda})(\sum_k d_{k2}^{\lambda})} h(D_{ij}, d_{it})$$

where

$$(5.21) \qquad h(.) = \exp [\beta_i (\delta_{ij} d_{i2}^{\mu} - D_{ij})] \quad \text{for model III.1}$$

$$(5.22) \qquad h(.) = \exp [\beta_i (\delta_{ij} d_{i2}^{\mu} - \log(\delta_{ij} + D_{ij}))] \quad \text{III.2}$$

$$(5.23) \qquad h(.) = \exp [\mu \delta_{ij} \frac{d_{i2}^{\lambda}}{\sum_j d_{j2}^{\lambda}} - \beta_i D_{ij}] \quad \text{for model III.3}$$

The fourth model, model III.4, combines the functional form of model III.1 with a set of inter-brand distances based on the Minkovski metric with $p = 1$ (City-Block) rather than the usual Euclidean distance (Minkovski metric with $p = 2$)[9].

No new terms were introduced in the above formulations but an extra parameter μ that controls the extent to which the repeat purchase probabilities vary with perceptual distances in the joint space. When $\mu = 0$, model III (version 1 and 2) reduces to model II. As the

---

(9) In some instances (see e.g. Bass et al. [1972], the City block metric proved to be superior to the usual Euclidean metric in terms of empirical performance. It is included here for completeness rather than for any theoretical reasons.

absolute value of μ gets large, the repeat purchase probabilities decrease sharply as perceptual distances from the ideal point increase.

Note that equations (5.21) through (5.23) can be rewritten as:

(5.24) Model III.1 $$h(.) = \begin{cases} e^{\beta_i d_{i2}^{\mu}} & i = j \\ e^{-\beta_i D_{ij}} & i \neq j \end{cases}$$

(5.25) Model III.2 $$h(.) = \begin{cases} e^{\beta_i d_{i2}^{\mu}} & i = j \\ D_{ij}^{-\beta_i} & i \neq j \end{cases}$$

(5.26) Model III.3 $$h(.) = \begin{cases} e^{\mu\, d_{i2}^{\lambda}/\sum_j d_{j2}^{\lambda}} & i = j \\ e^{-\beta_i D_{ij}} & i \neq j \end{cases}$$

This notation makes it clear that while the repeat purchase probabilities depend on the perceptual distances between the brands and the consumers' ideal point, the brand switching probabilities are but functions of inter-brand perceptual distances. In models III.1 and III.2, the parameters β_i (i=1,..., n) act upon both the repeat purchase and the brand switching probabilities. In model III.3, the β's are uniquely associated with brand switching probabilities and the repeat purchase probabilities are assumed to vary with relative rather than absolute weighted perceptual distances.

This completes the description of the models that will be submitted to the empirical data to test the two hypotheses mentioned earlier[10]. For convenience, the models are summarized in table V.4.

5.4 Estimation procedure

As in chapter II, the model parameters were estimated by the minimum chi-square procedure. That is, the parameters were chosen so as to minimize the usual goodness-of-fit chi-square statistic defined as:

$$x^2 = N \cdot \sum_{i,j} \frac{(t_{ij} - P_{ij})^2}{P_{ij}}$$

where

t_{ij} = observed brand switching probabilities

P_{ij} = model predicted brand switching probabilities

N = sample size upon which the observed probabilities were derived.

---

(10) The hypotheses were:

i) Brand choice and repeat purchase probability are functionally related to the perceptual distance between the brands and the consumers' ideal point.

ii) Brand-switching probabilities are functionally related to inter-brand perceptual distances in the joint space.

5.5 The data

The data used in this analysis were collected as part of a laboratory experiment conducted in the summer of 1969 by Bass, Pessemier and Lehmann [1972]. The study focused on preferences for eight brands of soft drink which were selected to represent two major segments of the soft-drink product market: Lemon-Lime versus Cola, and Diet versus Non-Diet. The brands included in table V.1 account form the bulk of the soft drink purchases in the area where the research was conducted. Two hundred and sixty four subjects chose one of the available soft drinks each Monday through Friday morning for three weeks, thus producing a total of 12 "purchase" occasions. The subjects also completed attitudinal and background questions at four points during the experiment. These questions included, among other items, ratings of the brands on eight attributes and paired similarity judgments for the brands.

5.6 Joint space construction

To test our hypotheses, we need a joint space configuration from which the perceptual distances between each brand and the consumers' average ideal point together with all inter-brand perceptual distances can be obtained.

BRANDS OF SOFT-DRINK SELECTED FOR THE STUDY.

	NON-DIET	DIET
COLA	COKE PEPSI	TAB DIET-PEPSI
LEMON-LIME	7-UP SPRITE	LIKE FRESCA

TABLE V.1

Several techniques[11] have been developed to infer joint space configurations from similarity and preferences judgments, depending on the nature of the input data (metric versus non-metric).

For convenience, the joint-space configuration that will be used in this study is the one derived by Lehmann and Pessemier [1973]. The approach they followed was to develop a joint space configuration based on determinant attributes. They did this by submitting the attribute levels for each brand as judged by all the subjects in the study to a Cooley and Lohnes [1971] discriminant analysis program. The attributes used in the Lehmann and Pessemier study were:

- carbonation
- calories
- sweetness
- thirst quenching
- popularity with others
- packaging
- after-taste
- flavor preference

---

(11) For a review of multidimensional scaling techniques, see GREEN [1975].

The resulting configuration provides a reduced space coordinate system in which the perceptual distances between the brands may be represented. Lehmann & Pessemier [1973] found two major dimensions which accounted for about half of the total variance and over 97% of the explained variance[12]. As can be seen in table V.2, the groups are clearly different. In order to better visualize the discriminant configuration, the position of the eight brands along the two dominant dimensions are plotted in figure V.1. Lehmann & Pessemier interpreted dimension 1 as overall popularity and calories content of the brands. This interpretation was supported by the discriminant structure (i.e., the "calories" and "popularity with others" attributes were highly correlated with the first dimension .898 and .741 respectively). The second dimension is clearly flavor type: lemon-lime versus cola.

Having derived a dimensional representation of the brands in the market, the last step calls for positioning an average ideal point into the derived perceptual space. This task was accomplished with the aid of the Carroll-Chang PREFMAP algorithm [1970]. From the joint

---

(12) These figures represent the percentage of variance explained accounted for by the first two discriminant functions.

POSITIONS ON THE DISCRIMINANT FUNCTIONS[1]

BRANDS	DIMENSIONS			
	1	2	3	4
COKE	2.74	.43	.17	.15
7-UP	1.70	-.21	.15	-.29
TAB	-2.33	.39	-.14	.11
LIKE	-1.59	-.43	-.20	.11
PEPSI	2.42	.30	-.17	.08
SPRITE	.78	-.57	-.31	-.07
D-PEPSI	-2.49	.60	.05	.21
FRESCA	-1.21	-.50	.46	.12
Respective % of variance explained by particular discriminant function	90.36	6.92	1.69	.54

1) From Lehmann and Pessemier [1973].

TABLE V.2

space so derived, all perceptual distances (inter-brand and between each brand and the average ideal point) were obtained.

5.7 Analysis and Results

There were six time periods in which all of the brands were available and in which conditions were stable. From these six choice occasions it was possible to compute five joint probability matrices on the basis of the actual choice behavior of the 264 subjects who participated in each of the six choice occasions. The average of these joint matrices is presented in table V.3a.

A summary of parameter estimates and goodness of fit statistics for various models is presented in table V.5, while tables V.3b and V.3c display the predicted brand-switching probabilities for selected models.

While the actual and predicted joint probabilities match remarkably well, the chi-square goodness-of-fit statistic is significant beyond the .01 level for all models. However, caution should be exercized when interpreting the chi-square test results. Since the chi-square goodness-of-fit statistic is proportional to the sample size, small deviations from the observed joint probabilities get amplified as the sample size increases.

(a)

OBSERVED AVERAGE JOINT PROBABILITY MATRIX FOR SOFT DRINK BRANDS.

	Coke	7-Up	Tab	Like	Pepsi	Sprite	D-Pepsi	Fresca	Market share at t
Coke	.188	.033	.003	.010	.041	.017	.004	.011	.307
7-Up	.032	.077	.001	.011	.024	.017	.002	.008	.172
Tab	.002	.003	.004	.009	.002	.001	.002	.002	.025
Like	.004	.007	.004	.007	.011	.002	.006	.005	.041
Pepsi	.047	.035	.002	.008	.137	.020	.007	.010	.266
Sprite	.008	.013	.002	.005	.011	.023	.002	.006	.070
D-Pepsi	.004	.002	.008	.004	.005	.004	.011	.005	.043
Fresca	.017	.007	.004	.008	.011	.008	.005	.015	.075
Market share at t+1	.304	.177	.028	.062	.242	.092	.038	.062	1.000

(b)

THEORETICAL AVERAGE JOINT PROBABILITY MATRIX FOR SOFT DRINK BRANDS, MODEL I.4

$h(\cdot)=\exp\{\beta_i(\delta_{ij} - D_{ij})\}$

	Coke	7-Up	Tab	Like	Pepsi	Sprite	D-Pepsi	Fresca	Market Share at t
Coke	.170	.034	.002	.008	.065	.016	.003	.009	.307
7-Up	.028	.076	.002	.008	.028	.018	.003	.009	.172
Tab	.002	.002	.004	.005	.002	.002	.004	.004	.025
Like	.009	.006	.003	.006	.007	.004	.005	.006	.046
Pepsi	.069	.036	.003	.008	.119	.017	.004	.009	.266
Sprite	.008	.011	.002	.007	.008	.025	.003	.008	.070
D-Pepsi	.004	.003	.005	.008	.003	.003	.009	.007	.043
Fresca	.013	.009	.006	.012	.010	.007	.008	.010	.075
Market share at t+1	.304	.177	.028	.062	.242	.092	.038	.062	1.000

(c)

THEORETICAL AVERAGE JOINT PROBABILITY MATRIX FOR SOFT DRINK BRANDS, MODEL III.3

	Coke	7-Up	Tab	Like	Pepsi	Sprite	D-Pepsi	Fresca	Market share at t
Coke	.174	.030	.008	.010	.046	.020	.008	.011	.307
7-Up	.022	.056	.004	.006	.026	.017	.004	.007	.172
Tab	.002	.003	.013	.007	.003	.002	.010	.006	.025
Like	.007	.007	.005	.007	.008	.006	.005	.005	.046
Pepsi	.041	.030	.007	.009	.140	.018	.007	.009	.266
Sprite	.011	.017	.004	.006	.013	.035	.003	.007	.070
D-Pepsi	.003	.004	.009	.007	.004	.004	.012	.006	.043
Fresca	.007	.007	.005	.006	.008	.006	.005	.007	.075
Market share at t+1	.304	.177	.028	.062	.242	.092	.038	.062	1.000

TABLE V.3

SUMMARY OF PARAMETER ESTIMATES FOR VARIOUS MODELS - SOFT DRINK DATA.

Parameter Estimates	MODEL I.2	MODEL I.4	MODEL II	MODEL III.1	MODEL III.2	MODEL III.3	MODEL III.4
B_1 COKE	1.700	.564	.544	.318	.466	.248	.282
B_2 7-UP	1.287	.609	.601	.285	.321	.430	.255
B_3 TAB	.814	.201	.334	.273	.347	.404	.251
B_4 LIKE	.291	-.068	-.042	.040	.057	.042	.035
B_5 PEPSI	1.410	.559	.550	.315	.448	.298	.283
B_6 SPRITE	1.359	.534	.539	.254	.283	.516	.219
B_7 D-PEPSI	1.460	.214	.221	.216	.337	.313	.206
B_8 FRESCA	1.070	-.020	-.012	.075	.120	.040	.064
λ	-	-	-1.83	-1.84	-1.85	-1.82	-1.84
μ	-	-		.33	.35	4.17	.35
χ^2	155.07	79.40	159.40	121.94	138.42	121.13	125.26
Normal Approxim. ≠			7.50	5.39	6.38	5.31	5.58
Degrees of Freedom	41	41	54	53	53	53	53

Sample Size: 1310

$\chi^2_{.99,40} = 63.69$

	MODEL I.1	MODEL I.3
β	1.378	.376
χ^2	176.54	155.58

TABLE V.5

≠ For large values of df, the expression $\sqrt{2\chi^2} - \sqrt{2(df) - 1}$ may be used as a normal deviate with unit standard error.

An element by element comparison of the predicted with the actual joint probabilities indicates substantially accurate predictions for the P_{ij}, as illustrated below for the repeat purchase probabilities predicted by model I.2:

Repeat purchase probabilities (P_{ii})	Observed	Predicted (MODEL I.2)
Coke	.188	.190
7-up	.770	.780
TAB	.004	.002
Like	.007	.006
Pepsi	.137	.138
Sprite	.023	.023
D - Pepsi	.011	.080
Fresca	.015	.015

When the 64 observed and predicted joint probabilities were correlated, all models produced high correlation coefficients (as large as .98 for model I.4). Therefore it is fair to say that the evidence is in agreement with the theory and that consumers' cognitive structure (perceived similarities among brands) is a good predictor of actual brand switching. In other words, the findings of this study are supportive of behavioral inferences commonly derived from interpretations of joint space configurations. A more detailed analysis of the results afforded by the various models follows.

Models I

The two one-parameter models (I.1 and I.3) provided a reasonably good fit to the data (chi-square values of 176.54 and 155.58 respectively). Model I.3 tends to underestimate the repeat purchase probabilities (P_{ii}) although it does quite well in reproducing the brand switching probabilities. In contrast, Model I.1 which does not capitalize on the information afforded by the joint space is best at reproducing the repeat purchase probabilities.

Allowing the parameter β to vary across brands produced mixed results:

i) Model I.1's performance is not significantly enhanced by the introduction of the extra parameters. The decline in chi-square from 176.54 to 155.07 is not statistically significant at the usual levels. Note that model I.3's predicted power is equivalent (in terms of chi-square) to that of model I.3 based on the perceived distances between the brands, although the latter is best at predicting brand-switching probabilities whereas the former's strenght rests on its ability to reproduce repeat purchase

probabilities, being in that respect similar to BASS' stochastic model [1974].

ii) The dramatic improvement of model I.4 over model I.3 is worth noting. By allowing the parameter β to vary across brands, the chi-square value shrank from 155.58 to 79.40, a decrease sharp enough to compensate for the loss of seven degrees of freedom.

The estimates of the β's confirm earlier contentions about their interpretation for models with or without perceptual distances as an input. One can readily check from the figures exhibited below that the β's of model I.2 are associated with the conditional repeat purchase probabilities[13].

---

(13) Conditional repeat purchase probabilities are defined as:

$$P_{i/i} = P_{ii}/\sum_j P_{ij} \text{ where } P_{ij} \text{ is the joint probability.}$$

They represent the probability that a consumer will purchase some brand, say i, given his last purchase was a purchase of brand i.

	Conditional Repeat purchase probability	Corresponding value for β_i Model I.2 $h(.) = \exp(\beta_i \delta_{ij})$	Model I.4 $h(.) = \exp[\beta_i(\delta_{ij} - D_{ij})]$
COKE	.612	1.700	.564
PEPSI	.515	1.410	.559
7-UP	.448	1.287	.609
SPRITE	.329	1.359	.534
D-PEPSI	.256	1.460	.214
FRESCA	.200	1.070	-.020
TAB	.160	.814	.201
LIKE	.152	.291	-.068

When the perceptual distances are used, as in model I.4, the β's are still somewhat associated with the repeat purchase probabilities. Non-diet brands (Coke, Pepsi, 7-Up and Sprite) which enjoy relatively high conditional repeat purchase probabilities exhibit higher values for β than do diet brands with low conditional repeat purchase probabilities. But within each group (diet versus non-diet), the β's are no longer associated with repeat purchase probabilities. Rather, they indicate the extent to which the observed brand-switching probabilities are consistent with the derived joint space. That this is indeed the case is best illustrated by the following table:

SWITCHING

FROM / TO

								Switching inferred from
FRESCA	COKE LIKE	PEPSI SPRITE	LIKE TAB	SPRITE D-PEPSI	7-UP 7-UP	D-PEPSI PEPSI	TAB COKE	Actual choice Joint space
SPRITE	7-UP	PEPSI	COKE	FRESCA	LIKE	TAB	D-PEPSI	Actual choice
	7-UP	FRESCA	PEPSI	COKE	LIKE	TAB	D-PEPSI	Joint Space

If the derived perceptual map pictured in figure V.1 is a proper specification of consumers' cognitive brand structure, and assuming that consumers would rather switch to similar than dissimilar brands, one would expect regular consumers of Fresca to switch to brands such as Like and Sprite as indicated in the above table. It turns out, however, that Coke and Pepsi are the two brands which Fresca consumers switched to most of the time, in spite of the fact they were perceived as being more unlike Fresca than any other brand (see figure V.1). In other words, consumers' switching behavior to and from Fresca (or Like) was not consistent with the joint-space theory of stochastic choice. Accordingly, the derived perceptual distances for those two brands were given negligible (negative) weights: -.020 and -.068 respectively.

In contrast, switching from and to Sprite agreed with what would be expected on the basis of consumers' perceptions. Brands that were perceived as being similar to Sprite, such as 7-Up, Pepsi and Coke were also those that were selected more often as second choice by regular Sprite consumers. As a result, the β parameter associated with Sprite was large as were those of brands for which consumers' brand-switching behavior behaved according to expectations such as Coke, Pepsi and 7-Up.

Model II

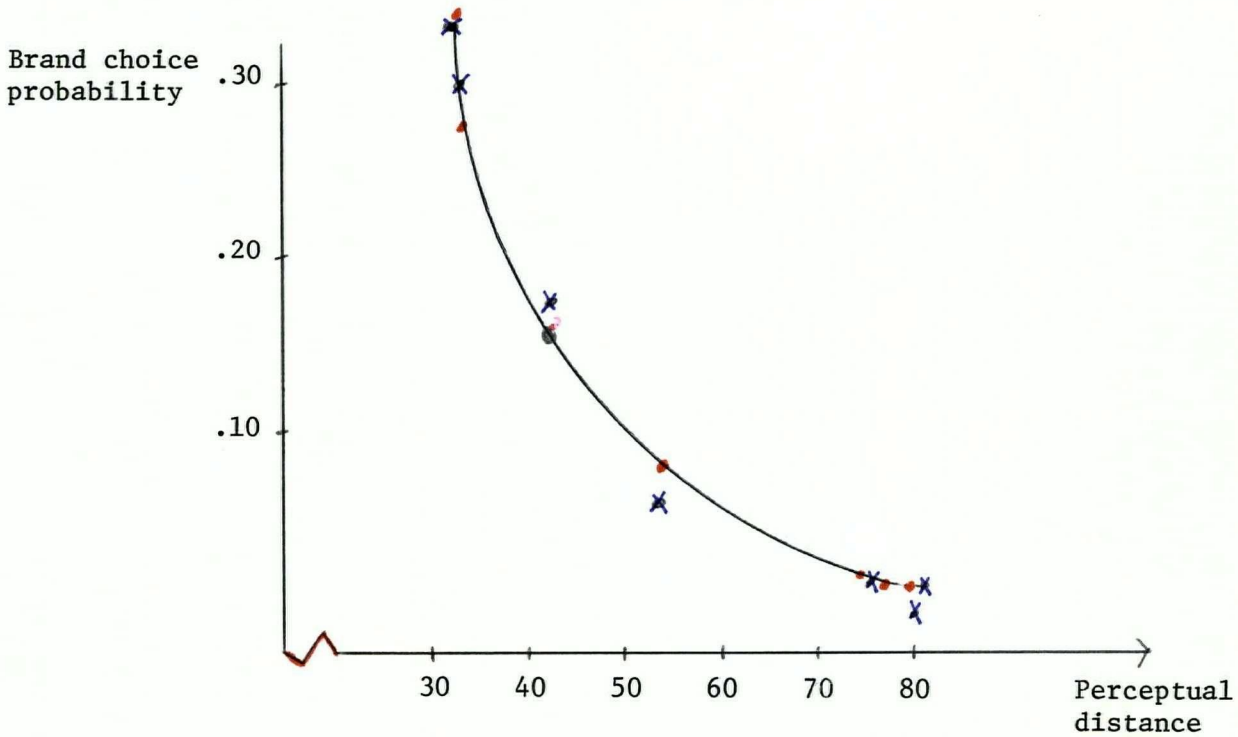
Of the three functional forms (hyperbolic, exponential and polynomial) that were experimented with, the hyperbolic function provided the best fit to the observed brand choice probabilities. Figure V.2 portrays the relationship between perceptual distances and brand choice probabilities for one of the two choice occasions. As the perceptual distance between a brand and the consumers' average ideal point increases, the probability of that brand being selected on any purchase occasion dwindles dramatically.

The excellent match between observed and predicted brand choice probabilities lends some additional support to the joint space theory of brand choice.

The β parameters that control the extent to which brand-switching probabilities vary with interbrand perceptual distances are very similar to those obtained under model I.4. All comments about their interpretation extend to model II as well.

FIGURE V.2

Relationship Between Brand Choice Probabilities and Perceptual Distances - Soft Drink Experiment. Hyperbolic Model



Legend: • Predicted

X Observed

Observed brand choice probabilities	Predicted brand choice probabilities	Perceptual distance	BRANDS
.307	.309	34.04	COKE
.172	.141	46.46	7-UP
.025	.046	82.21	TAB
.046	.050	79.01	LIKE
.266	.262	35.58	PEPSI
.070	.096	56.51	SPRITE
.043	.047	82.49	D-PEPSI
.075	.050	78.81	FRESCA

Functional form: $m_i = \frac{d_i^{\lambda}}{\sum_j d_j^{\lambda}}$, with $\lambda = -1.83$

Relationship Between Observed Repeat Purchase Probabilities and Perceptual Distances. Soft Drink Experiment.

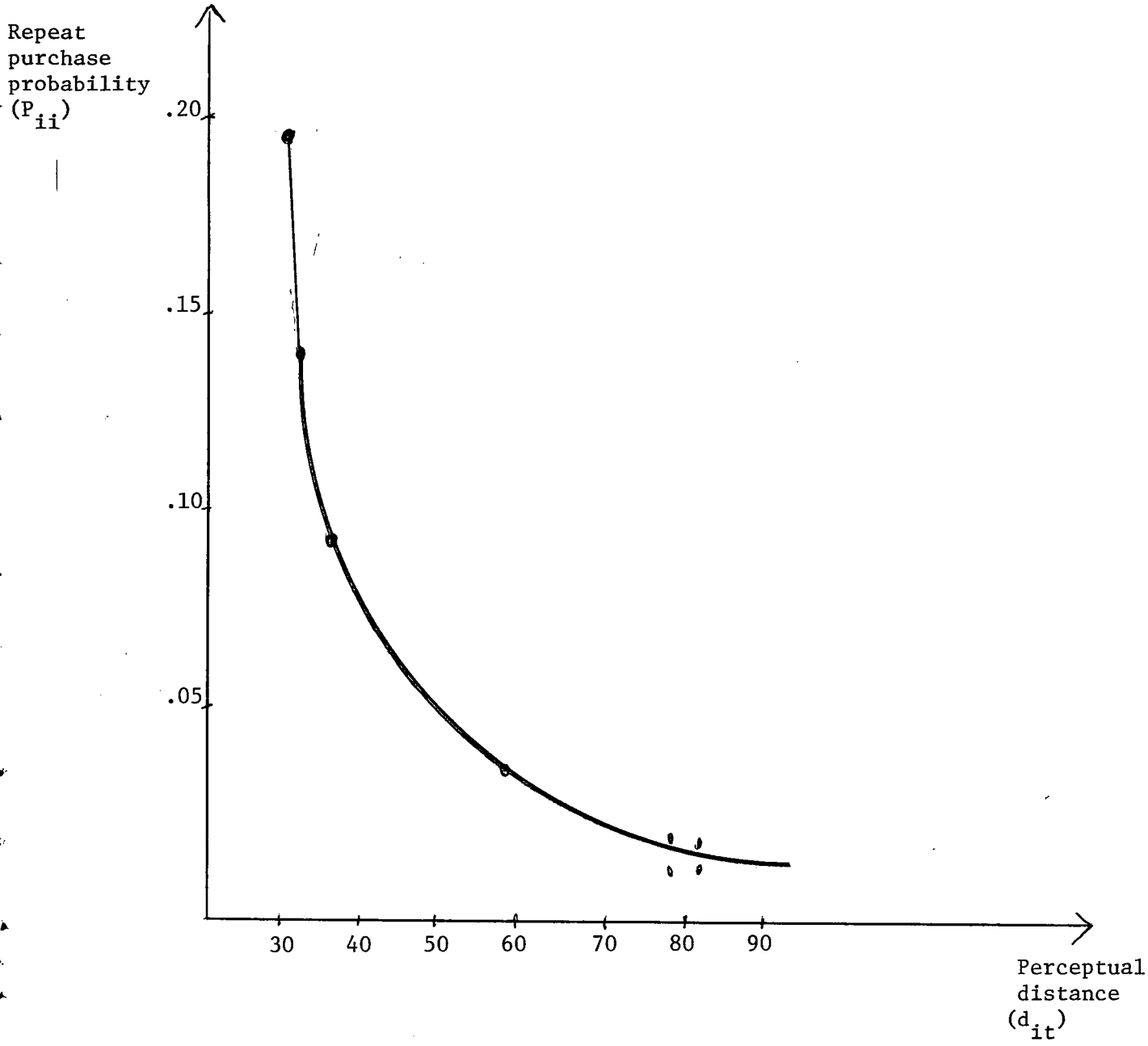


FIGURE V.3

Model III

While all four versions of model III produced similar results in terms of goodness of fit, model III.3 proved to be the only one for which no disturbing sign reversal occurred for the parameter μ. By construction and from consumer choice theory, one would expect the parameter μ in models III.1 and III.2 to be negative. As the distance between the brands and the consumers' ideal point increases, the repeat purchase probabilities are assumed to decrease, as illustrated in figure V.3.

The sign reversal for the parameter μ is a consequence of the dual role played by the β parameters as alluded to earlier. When these parameters are uniquely associated with the inter-brand distances as in model III.3, no such sign reversal occurs. The parameter μ, which in this model was expected to be positive, did turn out to be positive.

Model III.3 is therefore recommended as a most general and most appropriate model for the study of consumer brand switching behavior as applied to frequently purchased consumer goods.

Conclusions

This chapter investigated a stochastic theory of brand choice and brand switching derived from concepts of perception, preference, cognitive structure and consistency. A muti-dimensionally scaled configuration was used as a specification of consumers' cognitive structure. Perceptual distances derived from this configuration were then related to brand-choice and brand switching probabilities through a model that takes into account the constraints imposed on the various probabilities (limited range and sum constraints).

The following conclusions were reached:

1) There exists a relationship between brand choice probability and joint-space distance. The hyperbolic function proved to be the best of the three mathematical functions that were a-priori postulated.

2) There also exists a relationship between repeat purchase and brand swtiching probabilities on the one hand, and joint-space distance on the other. Several mathematical functions were experimented with and one of them was singled out as the most appropriate function for the data at hand.

The approach followed in this study represents a marked departure from earlier attemps at brand switching modeling such as the two stochastic models developed by Herniter [1973] and Bass [1974]. In these models, brand switching behavior is assumed to be a pure random process. No attempt was made to relate brand switching probabilities to variables of managerial significance such as perception, preference or other marketing variables of interest.

It is hoped that the methodology developed in this chapter will trigger future research toward greater internal and external validation.[14]

The next chapter is devoted to the reverse problem that consists of inferring a joint space configuration from a matrix of brand-switching probabilities.

---

(14) Such as the use of more realistic experimental settings, different scaling mechanisms, and more disaggregate data.

CHAPTER VI

MULTI-DIMENSIONAL SCALING OF BRAND SWITCHING DATA

6.1 Introduction

This chapter discusses a procedure for deriving a joint-space configuration from brand-switching probabilities rather than the more widely used judged similarity or preference measures (Green and Carmone [1969]). Since a joint space configuration implies a particular brand-switching structure, as demonstrated in the previous chapter, it seems likely than a brand-switching structure will in turn imply a joint-space configuration.

In chapter V, we addressed ourselves to the problem of predicting actual brand switching probabilities from an average joint space configuration derived from the individuals' perceived similarities between and preferences toward the various brands. Thus, we entertained the problem symbolically expressed by:

(6.1) $P_{ij} = f(D_{ij})$

where P_{ij} = brand-switching probability between brand i and j

D_{ij} = perceptual distance between brands i and j in the joint space

In this chapter, we address ourselves to the reverse problem expressed by:

$$D_{ij} = f^{-1}(P_{ij}) \tag{6.2}$$

where f^{-1} denotes the inverse of the function f. Here, the concern is with deriving a joint space configuration that is as consistent as possible with the observed brand switching probabilities. The purpose of this chapter is to present a methodology to perform such multidimensional scaling. To this end, the remaining discussion has been organized as follows:

The next section demonstrates the inability of traditional multidimensional scaling techniques to cope with the problem at hand. An alternative methodology is then developed and illustrated with the soft drink experimental data described in the preceding chapter.

The discussion ends on some concluding remarks about the limitations and future prospects for the recommended methodology.

6.2 Methodological Implications

Since Shepard [1962]'s pioneering work, literally scores of computer programs have been developed for metric or non-metric scaling of similarities or preference data[1]. But in spite of this intense algorithmic activity, none of the models produced so far can adjust to the restricted nature of brand-switching data.

The limitations of brand-switching probabilities as an appropriate input to traditional multidimentional scaling (MDS) techniques are best appreciated when the properties of brand-switching probabilities are compared with those of the usual distance and similarity measures.

If we let d_{ij} stand for the distance between two objects i and j from a set of n such objects, say $B = \{1, \ldots, n\}$, the quantity d_{ij} satisfies the following axioms:

(i) $d_{ii} = 0 \qquad \forall\, i \in B$

(ii) $d_{ij} = d_{ji} \qquad \forall\, i \text{ and } j \in B$ (symmetry)

(iii) $d_{ij} \leq d_{ik} + d_{kj} \qquad \forall\, i \neq j \neq k \in B$ (triangle inequality)

When axiom (iii) is not satisfied, the quantity d_{ij} is referred to as a measure of similarity rather than

---

(1) For an assessment of marketing applications of multidimensional scaling techniques, see GREEN [1975].

as a measure of distance. Such similarity measures are traditionally used as input to metric multi-dimensional scaling algorithms.

In contrast, brand-switching probabilities generally satisfy none of the above axioms since:

(i) the repeat purchase probability P_{ii} need not and in general will not be equal to zero,

(ii) the brand switching matrix need not be symmetric, that is $P_{ij} \neq P_{ji}$ for $(i, j) \in B \times B$,

(iii) the triangle inequality need not be satisfied by the P_{ij}'s and is void of any meaning in the brand-switching context.

Moreover,

(iv) brand-switching probabilities are range constrained ($P_{ij} \geq 0 \ \forall_{i,j} \in B$),

(v) brand-switching probabilities are sum-constrained ($\sum_{i,j} P_{ij} = 1$).

Nevertheless, attempts have been made to submit brand-switching probabilities to ordinary MDS techniques[2].

---

(2) See for instance Lehmann [1972].

The probabilities predicted by metric MDS models, however, may not necessarily lie within the unit interval or add up to unity since there is no built-in mechanism to ensure that they do.

Alternatively, the brand-switching probabilities could be transformed into a set of rank order input data. The set of ranked pairs would then be submitted to a non-metric scaling algorithm such as Kyst (Kruskal et al [1973]) or Torsca 8 (Torgerson [1967]). While this approach would obviate the two limitations mentioned above, it disregards the metric information provided by the brand-switching probabilities and thus cannot be fully efficient.

Another difficulty arises from the possible asymmetry of the brand-switching matrix. While similarity measures can be constructed from brand-switching probabilities, the question remains of which half of the matrix should be deleted or whether some kind of "data massaging" would be more appropriate.

Last, none of the existing algorithms is able to extract from the repeat purchase probabilities the information required to position the consumers' average ideal point into the joint-space. Such algorithms have to forgo the information contained in the repeat purchase probabilities and limit themselves to brand positioning only.

It is clear at this point that existing metric or non metric multidimensional scaling techniques are not equipped to infer joint space configurations from brand-switching probabilities. Failure to recognize the restricted nature of such probabilities as an input to traditional MDS techniques may consequently produce spurious results[3].

What is needed is a methodology that makes explicit provision for the particular nature of brand-switching probabilities. Such a methodology is outlined in the following section.

6.3 Multidimensional Scaling of Brand-Switching Probabilities

The objective of multidimensional scaling of brand-switching probabilities can be stated non rigourously as: Given a matrix of brand-switching probabilities (similar to that shown in chapter V, table 3), find a configuration whose distances - in a specified dimensionality - best reproduce the original brand-switching probabilities.

---

(3) Lehman's [1972] attempts to infer a perceptual map from brand-switching probabilities provides such an example of spurious results. Using Bass et al [1972] soft drink experimental data and a non-metric scaling algorithm (TORSCA), he found little agreement between the perceptual maps derived from brand-switching probabilities and similarity measures.

The conceptual basis of using brand-switching matrices as a basis for perceptual mapping is not difficult to grasp intuitively. If one assumes that brand-switching probabilities are a function of perceptual distances, it is possible to work backward and infer a perceptual map from a brand-switching matrix.

The chief difficulty rests upon the specification of an appropriate model that will cause the predicted probabilities to satisfy all requirements for logical consistency. The model developed in the previous chapter provides a structural framework that brings about such logical consistency.

6.3.1 Problem Definition

The problem is to find that joint space configuration which best reproduces the original brand-switching probabilities. That is, one must select for each brand i and for each dimension k, the set of coordinates x_{ik} ($i = 1,\ldots, n$; $k = 1,\ldots, r$) that provides the best fit to the observed probabilities. The coordinates of the consumers' average ideal point must also be derived simultaneously with the brands' coordinates for each dimension.

To ease the subsequent discussion, let us introduce some notation. Let

$B = \{1, \ldots, n\}$ = a set of n brands

t_{ij} = Observed proportion of consumers that "switched" from brand i to brand j on two successive purchase occasions, $i,j \in B$

P_{ij} = Corresponding model predicted proportion, $i,j \in B$

$$X = [\bar{x}_1, \ldots \bar{x}_n] = \begin{bmatrix} x_{11}, & x_{21}, & \ldots, & x_{n1} \\ \vdots & & & \vdots \\ x_{1r}, & \ldots\ldots\ldots, & & x_{nr} \end{bmatrix}$$

x_{ik} = Joint space coordinate of brand i on dimension k, $i \in B$; $k = 1,\ldots,r$

r = number of dimensions of the joint space

$$Y = [y_1, y_2] = \begin{bmatrix} y_{11}, & y_{21} \\ \vdots & \vdots \\ y_{1r}, & y_{2r} \end{bmatrix}$$

y_{tk} = Joint space coordinate of the consumers' average ideal point on dimension k for purchase occasion t. $k = 1 \ldots, r$

$t = 1, 2$

D_{ij} = Joint space distance between brands i and j, $i,j \in B$

$$= \left[\sum_k |x_{ik} - x_{jk}|^p\right]^{1/p}$$

p = Minkovski metric

d_{it} = Joint space weighted distance between brand i and the consumers' average ideal point at purchase occasion t, $i \in B$, $t = 1, 2$.

$$= \left[\sum_k \omega_k |x_{ik} - y_{tk}|^p\right]^{1/p}$$

$m_{it} \doteq \sum_{i \text{ or } j} t_{ij}$ = Observed proportion of consumers who purchased brand i at purchase occasion t, $t = 1, 2$.

As in chapter V, we express the brand-switching probabilities (P_{ij}) in terms of joint space distances (D_{ij} and d_{it}) as follows:

$$P_{ij} = a_i\, b_j\, g\,(d_{i1}, \tilde{\lambda})\; g\,(d_{j2}, \tilde{\lambda})\; h\,(D_{ij}, d_{it}, \tilde{\Theta}) \qquad i,j \in B$$

Where

$$a_i = \left[\sum_j b_j\, g(D_{j2}, \tilde{\lambda})\; h(d_{ij}, d_{it}, \tilde{\Theta})\right]^{-1} \qquad i \in B$$

$$b_j = \left[\sum_j a_i\, g(d_{i1}, \tilde{\lambda})\; h\,(D_{ij}, d_{it}, \tilde{\Theta})\right]^{-1} \qquad j \in B$$

and the functions g and h are as specified in the previous chapter. The quantities $\tilde{\lambda}$ and $\tilde{\Theta}$ denote parameter vectors to be estimated along with the brands' coordinates by minimizing some function of the discrepancy between observed and predicted brand-switching probabilities. Such a measure is the usual chi-square goodness-of-fit statistic defined as:

$$(6.3) \qquad \chi^2 = N \sum_{i,j} (t_{ij} - P_{ij})2 \; / \; P_{ij}$$

where N is the sample size from which the observed brand-switching probabilities were obtained. The smaller the chi-square value, the better the fit between observed and predicted probabilities.

Besides its potential usefulness as an indication of "correct" dimensionality, the chi-square statistic, unlike the usual fit measures of non-metric MDS like Kruskal' "stress", proves very convenient for statistical inference. Since the sampling distribution of the chi-square statistic is known, tests can be carried out to determine whether the derived joint space is consistent with the observed brand-switching probabilities.

6.3.2 Estimation Procedure

For a specified dimensionality r and a given Minkovski p - metric, the problem is to:

$$\text{MINIMIZE } \chi^2 = N \sum_{i,j} (t_{ij} - P_{ij})^2 / P_{ij}$$

with respect to X, Y, $\{\omega_k\}_{k=1}^{r}$, $\tilde{\lambda}$ and $\tilde{\Theta}$.

This minimization process is repeated in the next higher dimensionality until the decrease in the chi-square value brought about by the additional dimension is not statistically significant. The estimation of the model parameters was accomplished with the aid of a computer algorithm that is described below. This is followed by a discussion of the stopping rule that was adopted to select the "correct" dimensionality.

6.3.2.1 Computer Algorithm

The algorithm combines a gradient search with an iterative procedure similar to that used in Torgerson's [1967] non-metric MDS algorithm (TORSCA 8). Given the large number of parameters to be estimated, a gradient

search alone was considered unmanageable[4]. Fortunately, there exists a simple heuristic iterative procedure to improve a starting configuration until it converges to a stable value for a given dimensionality.

For expository purposes, assume that we have a set of starting values[5] for Y, $\tilde{\omega}$, $\tilde{\lambda}$ and $\tilde{\Theta}$, say Y^o, $\tilde{\omega}^o$, λ^o and $\tilde{\Theta}^o$. The first step consists of finding an initial value for the matrix of brand coordinates X, say X^o.

Finding an initial configuration

Assume there exists an integer $r < n$ and n vectors $\bar{x}_1, \ldots \bar{x}_n \in \mathbb{R}^r$ such that[6]

$$P_{ij} = \alpha D_{ij}^2 = (x_i - x_j)' (x_i - x_j) \quad \forall i, j \in B,$$

---

(4) For the most general model (Model III), there are altogether n + r (n + 3) + 2 parameters where n is the number of brands and r is the number of dimensions in the joint space. Even as few as eight brands and two dimensions produce 32 parameters. Ten brands and five dimensions yield a staggering 77 parameters.

(5) For notational convenience, we shall let $\tilde{\omega}$ stand for $\{\omega_k\}_{k=1}^r$

(6) And setting p = 2 in the distance formula D_{ij}.

That is, we assume that the brand-switching probabilities P_{ij} are proportional to the square of inter-brand distances in the joint space. This, of course, will not hold exactly but will allow us to derive an initial set of values for X.

The P_{ij} are known. The coordinate matrix $X = [x_1, \ldots, x_n]$ is wanted up to a similarity transform (rotation of the configuration about the origin, translation of the origin, reflection) since such a transform preserves proportionality between the P_{ij} and D_{ij}^2. Without loss of generality, we let α be equal to one and let the center of gravity of the $\bar{x}_j$ be the origin.

It can be shown (Torgerson [1958]) that the scalar product between two vectors $\bar{x}_j$ and $\bar{x}_k$, denoted by $e_{jk} = \bar{x}_j' \bar{x}_k$, depends only on the inter-point distances as follows:

$$(6.4) \qquad e_{jk} = -\frac{1}{2}\left[P_{jk} - \frac{1}{n}\sum_{e=1}^{n} P_{ek} - \frac{1}{n}\sum_{e=1}^{n} P_{je} + \frac{1}{n^2}\sum_{e,s} P_{es}\right].$$

Let $E = [e_{jk}]$ be the matrix of scalar products between the unknown $\bar{x}_j$. Since the P_{ij} are known, the matrix E can be computed from (6.4). The wanted initial configuration can be obtained by factoring the known matrix $E = X'X$. Since E is equal to the product $X'X$, E is a positive semi-definite matrix of rank r. As a result, E can be written as

$$(6.5) \qquad E = Z' \begin{bmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_r \end{bmatrix} Z,$$

where the μ's are the eigenvalues and Z is the matrix of eigenvectors associated with the matrix E. The desired initial configuration is just

$$(6.6) \qquad X^o = \begin{bmatrix} \sqrt{\mu_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\mu_r} \end{bmatrix} Z.$$

Searching the parameter space.

From the starting values X^o, Y^o, $\tilde{\omega}$, $\tilde{\lambda}^o$ and $\tilde{\Theta}^o$, a chi-square value may be computed from (6.3). Suppose, however, that this chi-square value is high; that is, the fit between observed and predicted brand-switching probabilities is poor. The next step consists of finding a new configuration X^1 and Y^1 that will produce a better match between observed and predicted probabilities, keeping the other parameters ($\tilde{\omega}$, $\tilde{\lambda}$ and $\tilde{\Theta}$) constant . From the new configuration X^1 and Y^1, a gradient search will then be performed to obtain improved values for $\tilde{\omega}$, $\tilde{\lambda}$ and $\tilde{\Theta}$. We first deal with the estimation of X^1 and Y^1.

Improving the configuration

The objective is to move the points (brands and ideal) around so that the new configuration yields a better match between observed and predicted probabilities. In particular, consider a specific brand i and its relationship to each of the brands j in turn. We would like to move brand i in the perceptual space so as to decrease the average discrepancy between the observed (t_{ij}) and predicted (P_{ij}) probabilities.

If P_{ij} is larger than t_{ij}, we could move brand i away from j by an amount which is proportional to the size of the discrepancy. Conversely, if t_{ij} is larger than P_{ij}, then brand j is to be moved towards i by an amount proportional to the discrepancy. Let α represent the coefficient of proportionality.

To find a new coordinate x^1_{ik} for brand i on dimension k, as related to brand j, we can use the formula:

$$x^1_{ik(j)} = x^o_{ik(j)} + \alpha \left(1 - \frac{P_{ij}}{t_{ij}}\right) (x^o_{jk} - x^o_{ik}). \tag{6.7}$$

The above formula would move brand i in the appropriate direction with respect to brand j, but we must consider all (n - 1) brands insofar as their effect on brand i is concerned. When this is done, equation (6.7) generalizes into

$$(6.8) \qquad x_{ik}^{1} = x_{ik}^{o} + \frac{\alpha}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(1 - \frac{P_{ij}}{t_{ij}}\right) (x_{jk}^{o} - x_{ik}^{o}).$$

That is, we move brand i along dimension k in such a way as to take into account the discrepancies involving all other brands. This is, of course, done for all brands in all dimensions and repeated until successive configurations converge.

A similar procedure is resorted to for the ideal point in order to select that set of coordinates that yields the best fit between observed and predicted brand choice probabilities. In particular, consider a specific brand, say brand i. We would like to move the ideal point in the joint space so as to decrease the discrepancy between the observed (m_{it}) and the predicted [$g(d_{it}, \tilde{\lambda})$] brand choice probabilities for brand i.

If m_{it} is larger than $g(.)$[7] we could move the ideal point towards brand i by an amount that is proportional to the size of the discrepancy. Conversely, if $g(.)$ is larger than m_{it}, the ideal point is to be moved away from brand i by an amount proportional to the discrepancy.

---

(7) To ease notation, we will write $g(.)$ instead of $g(d_{it}, \tilde{\lambda})$.

Since a change in the ideal point location affects the entire set of distances between the brands and the ideal point, we would like to move the ideal point so as to decrease the average discrepancy between the observed and predicted brand choice probabilities. To do this, we merely use the expression:

$$(6.9) \qquad y_{kt}^{1} = y_{kt}^{o} + \frac{\hat{\beta}}{n} \sum_{i=1}^{n} \left(1 - \frac{m_{it}}{g(.)}\right)(y_{kt}^{o} - x_{ik}^{o}).$$

where β represents the coefficient of proportionality[8].

This is, of course, done for each dimension and repeated until the coordinates of the ideal point converge to some stable value.

Searching the parameter space for $\tilde{\omega}$, $\tilde{\lambda}$ and $\tilde{\Theta}$

Once the best configuration afforded by the temporary parameter vectors $\tilde{\omega}^{o}$, $\tilde{\lambda}^{o}$ and $\tilde{\Theta}^{o}$ is obtained, the next step consists of searching the parameter space for improved parameters vectors $\tilde{\omega}^{1}$, $\tilde{\lambda}^{1}$ and $\tilde{\Theta}^{1}$ that will further reduce the chi-square goodness-of-fit statistic. This is best done by a gradient procedure[9].

---

(8) Both coefficients of proportionality (α and β) must be specified by the user.

(9) The gradient code used in this study was developed by BOX [1965].

The procedure then starts afresh with the last configuration X^1, Y^1 and the improved parameter vectors $\tilde{\omega}^1$, $\tilde{\lambda}^1$ and $\tilde{\theta}^1$ until the absolute differences between the values of each parameter for any two successive iterations satisfy some prespecified tolerance level.

Selecting the Number of Dimensions in the Joint Space.

The goodness of fit chi-square statistic provides valuable information to decide whether or not to add an extra dimension in the joint space. To select the number of dimensions, one would keep up adding dimensions (starting from a unidimensional space) until the chi-square statistic falls below some prespecified tolerance level.[10] If we let χ^2_r denote the chi-square statistic and df(r) denote the number of degrees of freedom associated with the model assuming a space of r dimensions, the stopping rule can be formulated as follows:

---

(10) It is tempting to resort to an F-test to decide whether the loss of degrees of freedom incurred by adding an extra dimension is matched by a significant reduction in the chi-square statistic. This temptation should,however, be resisted since

Starting from r=1, choose the first r for which

(6.10) $\chi^2_r < \chi^2_{\alpha,df(r)}$,

where α is the prespecified tolerance level. This stopping rule achieves a reasonable compromise between parsimony and the desire to achieve a good fit between the derived joint space and the observed brand-switching probabilities.

6.4 An Empirical Illustration.

To illustrate the use of brand-switching data to construct joint space configurations, the soft drink data described in the preceding chapter have been submitted to the multidimensional procedure just developed. Three different versions of the model were entertained. They are:

(6.11) $P_{ij} = a_i b_j \, g(d_{i1}, \tilde{\lambda}) \, g(d_{j2}, \tilde{\lambda}) \, h(D_{ij}, d_{it}, \tilde{\theta})$

---

the chi-squares obtained at successive iterations are not independent (they are based on the same set of data). Lack of independence between the chi-squares combined with the fact that the F-test is very sensitive to departure from independence precludes its use as a criterion to decide on the space dimensionality.

Where

MODEL I:
$$\begin{cases} g(d_{it}, \tilde{\lambda}) = m_{it} & t = 1, 2 \\ h(D_{ij}, d_{it}, \tilde{\Theta}) = \exp\,[\Theta_i\,(\delta_{ij} - D_{ij})] \end{cases}$$

MODEL II:
$$\begin{cases} g(d_{it}, \tilde{\lambda}) = \dfrac{d_{it}^{\lambda}}{\sum_j d_{jt}^{\lambda}} & t = 1, 2 \\ h(D_{ij}, d_{it}, \tilde{\Theta}) = \exp\,[\Theta_i\,(\delta_{ij} - D_{ij})] \end{cases}$$

MODEL III:
$$\begin{cases} g(d_{it}, \tilde{\lambda}) = \dfrac{d_{it}^{\lambda}}{\sum_j d_{jt}^{\lambda}} & t = 1, 2 \\ h(D_{ij}, d_{it}, \tilde{\Theta}) = \exp\,[\mu\delta_{ij}\,\dfrac{d_{i2}^{\lambda}}{\sum_j d_{j2}^{\lambda}} - \Theta_i\,D_{ij}]. \end{cases}$$

The reader is referred to the preceding chapter for a detailed description of and justification for the functional forms specified for the functions g and h. The three models (I, II, and III) correspond to models I.4, II and III.3 of chapter V. Model I is appropriate for deriving a configuration without ideal points. The output of such a model consists of a set of coordinates for each brand. The other two models produce a configuration where both brands and ideal points (one for each time period) are simultaneously positioned.

Analysis and Discussion

The results of the analysis are conveniently summarized in Tables VI.I to VI.3. Table VI.1 exhibits for each model the coordinates of each brand for each dimension, the coordinates of the average consumers' ideal point for each time period, the estimates of the remaining parameters as well as some statistics required for testing purposes.

As in the preceding chapter, Θ_i indicates the extent to which the observed brand-switching probabilities are consistent with the derived joint space. Two brands, LIKE and FRESCA were assigned negligible and/or negative values for the parameter Θ. In other words, consumers' switching behavior to and from LIKE and FRESCA was not consistent with the joint-space theory of stochastic choice. This lack of consistency, however, did not prevent a reasonably fair recovery of the consumers' perceptual map.

As can be checked from figure VI.1 (map derived from brand-switching probabilities) and figure V.1 of the previous chapter (map derived from judged similarity measures), the two maps are in close agreement. In both configurations, the major dimension can be interpreted as calories content of the brands: diet versus non diet. The

Parameter Estimates for Three Models - Joint Space Configuration Derived from Brand-Switching Data

Soft Drink Experiment

	MODEL I			MODEL II			MODEL III		
BRANDS	Θ_i	Dimension I	Dimension II	Θ_i	Dimension I	Dimension II	Θ_i	Dimension I	Dimension II
COKE	.169	9.72	-.49	.161	9.59	-.46	.145	9.46	2.24
7-UP	.228	5.09	-.59	.170	5.35	-.61	.193	4.91	.92
TAB	.072	-10.42	.36	.144	-10.22	.38	.055	-11.47	.07
LIKE	-.024	-5.57	-6.13	-.029	-5.85	-6.14	-.044	-2.20	-8.58
PEPSI	.180	6.72	4.68	.161	6.77	4.73	.151	5.85	5.72
SPRITE	.214	1.85	1.31	.237	1.76	1.29	.177	1.04	2.15
D-PEPSI	.111	-6.73	6.69	.145	-6.72	6.70	.113	-7.38	5.50
FRESCA	.008	-2.55	-5.89	.038	-2.57	-5.97	-.017	-.08	-9.47
Dimension Weights		-	-		2.84	.018		8.37	.101
λ	-			-.732			-.963		
μ	-			-			1.98		
Ideal Point Coordinates									
Time 1		-	-		8.54	7.46		7.86	8.07
Time 2		-	-		8.67	13.87		7.72	14.29
χ^2		32.23			43.22			42.68	
Degrees of freedom		25			32			31	

MULTI-DIMENSIONAL SCALING OF THE SOFT DRINK EXPERIMENT BRAND-SWITCHING MATRIX. MODEL I.

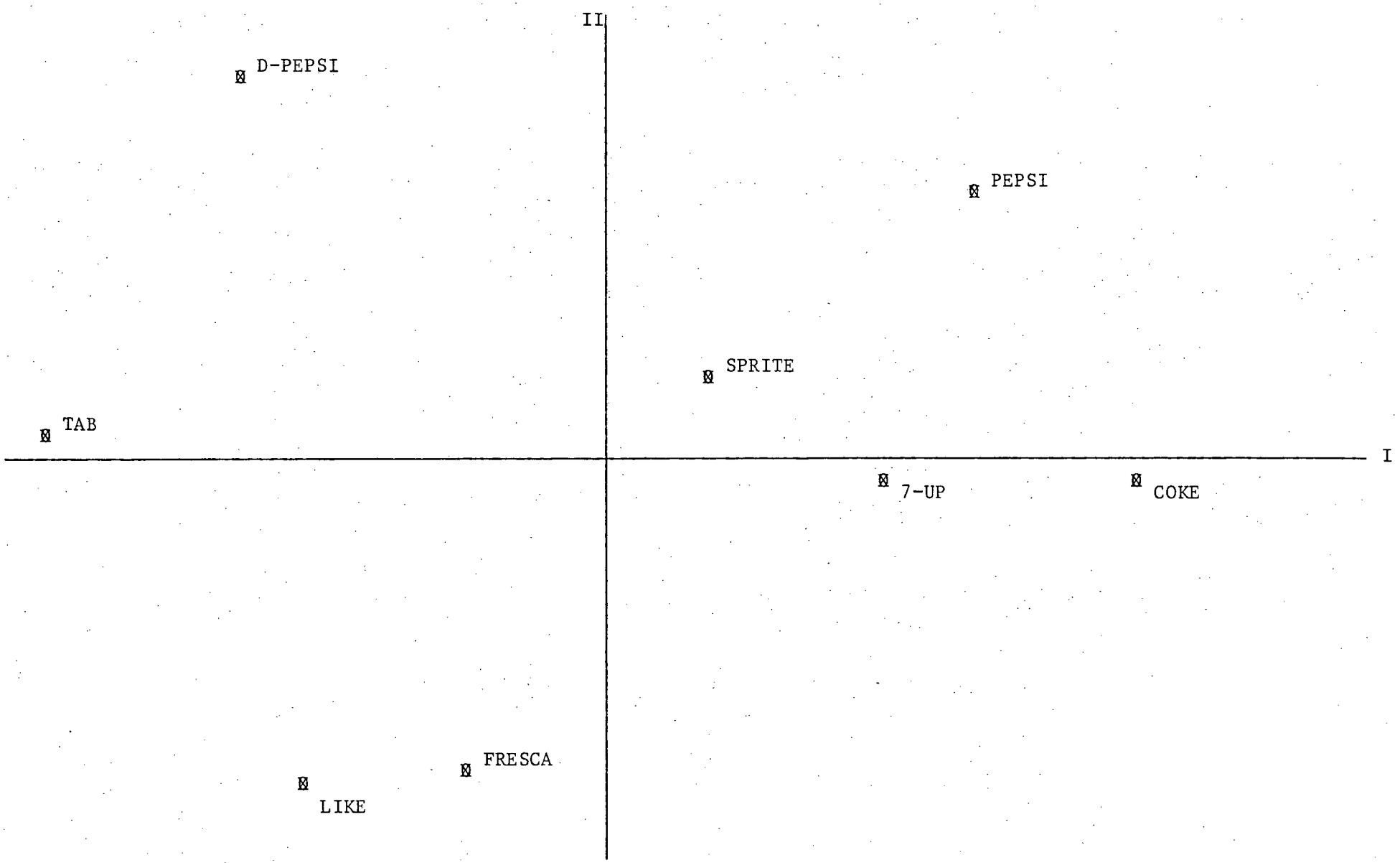


FIGURE VI.1

second dimension is clearly flavor type[11]: lemon-lime versus cola. Table VI.2 exhibits the correlation matrix of inter-brand distances across the 28 pairs for various models. The magnitude of the correlation coefficients (.87 for MODEL I) indicates a fairly good degree of agreement between the two configurations. There is almost perfect correlation between the two maps (brand-switching versus judged similarities) for the first dimension (.98 for Model I) and moderate agreement for the second dimension ($\simeq$ .70) as evidenced in Table VI.3.

There are differences, however. Based on the individuals' rating, the two diet brands Tab and Diet-Pepsi are perceived to be much more similar than they are on the basis of the brand-switching data. From the individuals' rating, we would think that consumers of Tab would switch to Diet-Pepsi, Like and Fresca respectively, should they be denied their favorite brand. However, when faced with actual choice decisions, it appears that they would rather switch to Like (36%), 7-Up (12%), or even to Coke, Pepsi and Fresca (8% each). Only 8% of them did switch to Diet-Pepsi. In the perceptual map based on brand-switching data (see figure VI.2),

---

(11) Notwithstanding COKE and SPRITE whose position is somewhat in disagreement with the flavor interpretation.

TABLE VI.2

CORRELATION MATRIX OF PERCEPTUAL DISTANCES (D_{ij}) ACROSS THE 28 PAIRS FOR VARIOUS MODELS

	MODEL II	MODEL III	Lehmann & Pessemier *
MODEL I	.999	.943	.870
MODEL II		.940	.873
MODEL III			.750

* Based on the perceptual distances obtained from Lehmann & Pessemier [1973].

TABLE VI.3

CORRELATION MATRIX OF THE COORDINATE VECTORS FOR VARIOUS MODELS (DIMENSION BY DIMENSION)

		MODEL II		MODEL III		Lehmann & Pessemier	
DIMENSION		1	2	1	2	1	2
MODEL I	1	.999	.119	.972	.352	.980	-.015
	2	.126	.999	-.0837	.947	.138	.687
MODEL II	1		.128	.968	.361	.982	-.008
	2			-.082	.968	.1405	.690
MODEL III	1				.140	.932	-.160
	2					.393	.671

7-Up is closer to Coke than Pepsi is, an unexpected result given that consumers of Coke switched to Pepsi more often than to any other brands and vice-versa. However, the predicted perceptual distance between Coke and Pepsi does not depend solely on the observed amount of switching between the two brands but also on that observed between all pairs of brands.

The differential weights (see the ω_k in the expression d_{it}) assigned to each dimension reinforce the conclusion that the first dimension (diet versus non-diet) was the most determinant factor for brand choice purposes: dimension I was assigned a weight of 8.37 as against .101 for the second dimension. This result indicates that the cognitive process underlying the brand choice decisions of the subjects is essentially unidimensional. This conclusion, however, may not apply to the soft drink market as a whole, due to the restrictive experimental setting used by Lehmann & Pessemier [1973] (laboratory experiment, student population, free consumption, etc...).

Finally, a most interesting finding from this analysis of brand-switching probabilities relates to the discrepancy observed for two brands, COKE and PEPSI between the perceptual distances as derived from judged similarity measures on the one hand, and brand-switching probabilities

on the other. As can be checked by a glance at figures V.1, VI.1 and table V.3a, the brand-switching activity between COKE and PEPSI observed from the subjects' brand choice over time was less dramatic than expected on the basis of their perceived similarity. In other words, there should have been more "switching" between COKE and PEPSI for the joint space theory of brand choice to hold exactly. While the subjects do perceive COKE and PEPSI as being two very similar soft drink brands, they do not consider them as being perfect substitutes for brand choice purposes. This finding bears far-reaching implications for brand positioning and advertising strategy.

As a case in point, consider the advertising campaign launched by the Pepsi Cola company during the spring of 1977. The chief vehicle for this advertising campaign was a brochure urging the reader to perform a "blind test". The instructions read as follows:

i) have somebody fill up two identical glasses with Pepsi-Cola and Coca-Cola respectively while you are watching away;

ii) take a sip at the two glasses and decide which is which.

In case the reader had trouble drawing the "right" conclusion, the brochure marveled at how difficult it was to tell the difference and discreetly suggested that since this was so, the reader might as well drink Pepsi-Cola instead of Coca-Cola.

The rationale underlying Pepsi-Cola Inc's advertising strategy rests on the intuitively appealing theory that consumers tend to organize their choice behavior so as to achieve a cognitive equilibrium between their perceptions and brand choice. According to this theory a consumer who perceives two brands as being virtually identical would exhibit no particular preference for one or the other and thus would tend to act as though they were perfect substitutes.

The analysis conducted in this chapter, however, shows what appears to be diminishing returns on "psychic investment". In this context, psychic investment denotes all marketing expenses incurred in the process of altering the consumer perceptual space. As the perceptual distance between two brands gets small, one observes that further decreases in perceptual distances do not yield proportionate increases in brand-switching probabilities. At this stage, further psychic investment to alter the consumers' perceptual space may yield diminishing returns

in that decreases in perceptual distances are not matched by substantial increases in brand-switching probabilities. As a result of these diminishing returns, the economic value of the Pepsi-Cola positioning campaign is open to question. The results obtained in this study suggest that it will take more than a mere psychic investment to trans- from a Coke "addict" into a Pepsi zealot since perceptions and preferences are not the only determinants of brand choice behavior.

SUMMARY

This chapter dealt with the problem of deriving a joint-space configuration form brand-switching probabilities rather than the more widely used judged similarity or preference measures. The shortcomings associated with applying traditional MDS techniques to brand-switching probabilities have been stressed and a more appropriate procedure discussed and illustrated with empirical data. This procedure satisfies all constraints for logical consistency arising for the particular nature of the input data.

Using brand-switching matrices as a basis for perceptual mapping is not without peril. For example, brand-switching data require a sample of individuals or measures of an individual over time, and the implicit assumption of homogeneity of perception over the group of individuals or time period. Yet, these problems exist in any application of perceptual mapping that is not individually based.

Last, perceptual mapping from brand-choice data should not be considered as a substitute for perceptual mapping from judged similarity measures, but rather as a useful complement. Comparison of the two maps derived from perceptions on the one hand and actual brand choice on

the other may give the brand manager valuable hints for brand positioning purposes.

CHAPTER VII

CONCLUSION

The purpose of this research was to analyse, discuss and extend the analytic methodology associated with the study of brand-switching data by developing stochastic models to be used as constructs for organizing and interpreting brand-switching data. Models for both two-brand and multi-brand markets were constructed.

The approach taken in this study was that of description leading to generalization. In chapter II, a set of consumer brand switching data were submitted to a statistical analysis in an attempt to uncover possible "regularities" in the data. These "regularities" were then exploited in chapter III to construct brand choice models based on the results of the preliminary statistical analysis. The statistical analysis gave support to the learning hypothesis of brand-switching behavior in that past purchases of a brand affect the probability of that brand being bought again in the future.

While the need to develop alternative models to accommodate the adaptive behavior observed in consumer brand choice was not obvious, the model (Polya-Learning) developed in chapter II exhibits some interesting properties:

i) It does not assume that the adaptive process is linear in the purchase probabilities.

ii) The limiting form of the Polya distribution is the so-called negative binomial distribution which is the backbone of Bass' [1974] brand choice and Ehrenberg's [1972] purchase incidence model.

iii) The model introduces a fine distinction between disloyal behavior and mere brand-switching. This distinction carries much importance for the brand manager who would like to know the extent to which brand-switching is evidence of disloyal behavior or mere variety seeking.

If only for those characteristics, the model could be considered as a viable alternative to existing two-brand stochastic choice models for the study of consumer brand choice behavior.

The dissertation's major contribution is contained in chapters IV to VI which developed a general class of brand-switching models that acknowledge both the stochastic and deterministic features of the brand-switching phenomenon. This approach stands at variance with that of previously developed multi-brand switching models (see e.g. Herniter [1973], Bass [1974] that focused exclusively on the probabilistic components of consumer brand-switching behavior.

A multi-dimensionally scaled configuration was used as a specification of consumers' cognitive structure. Perceptual distances derived from this configuration were then related to brand choice and brand-switching probabilities through a model that took into account the constraints imposed on the various probabilities. The empirical results demonstrated that perceptions, preferences and cognitive structure were indeed significant determinants of consumer habitual brand-switching behavior.

Finally, chapter VII dealt with the problem of deriving a joint-space configuration from brand-switching probabilities rather than the more widely used judged similarity or preference measures. Comparison of the two maps derived from perceptions on the one hand and actual brand choice on the other was shown to provide the brand manager with valuable hints for brand positioning purposes.

As with many other studies of consumer choice behavior, this dissertation raised more questions than it solved, leaving open numerous possible avenues for future research. Among the topics it is important to explore are:

i) Replication of the study with different sets of data.
ii) Replication of the study at the individual level.

iii) Finding ways to discriminate between brand-switching as a result of mere variety seeking versus brand-switching as evidence of disloyal behavior.

The conceptual framework of analysis developed in this study can handle aggregate as well as dis-aggregate input. The purpose of chapter V was to explain an aggregate brand-switching matrix with an aggregate perceptual map. Our understanding of brand-switching behavior would be further enriched by dealing with each individual's brand-switching matrix and perceptual map in an attempt to identify different types of consumer behavior. While the constructing of perceptual maps at the individual level poses no conceptual nor operational difficulties, data constraints severely limit the gathering of a brand-switching matrix at the individual level when the number of alternative brands is large, as in the soft drink experiment. For an n-brand market, there are n^2 possible brand-switching probabilities to "fill" in with actual data. Given that one needs a minimum of 5 observations per cell to have any confidence in the statistical tests to be used, no fewer than $5n^2$ purchase occasions are required. In the soft drink experiment where a choice of eight brands was offered to the subjects, one would need a whopping 320 purchase occasions to carry out the analysis at the individual level, an unrealistic number for all practical purposes.

Traditional brand choice experiments, such as Bass et al. [1972] and MacConnell [1968] beer experiments, will yield between 10 to 30 purchase occasions. The alternative is to collapse the market into subsets of brands or collapse the individuals with the same pattern of purchase into homogeneous groups. The latter procedure is more appealing than the former that would involve comparing individuals with different "evoked" sets of brands. Clustering techniques such as the ones suggested by Johnson [1967] or Lehmann and Pessemier [1973] could be employed to form homogeneous groups of consumers with respect to their purchase behavior. For each group, the analysis outlined in chapter V could be carried out to answer questions such as:

i) In a product category, does the perceptual configuration of brand-loyal consumers differ from disloyal consumers?

ii) Do brand-loyal consumers perceive the various brands in a product category as being substantially dissimilar, whereas disloyal consumers perceive all brands as being very similar?

iii) To which extent do the perceptual maps derived from similarity measures and brand-switching data differ across groups?

This we shall leave for future research.

BIBLIOGRAPHY

Aaker D.A., "A New Method for Evaluating Stochastic Models of Brand Choice," Journal of Marketing Research, Vol. 7, No. 3 (August 1970), pp 300-6.

_____ (1971) "The New Trier Stochastic Model of Brand Choice," Management Science, Vol.17, No.8, pp B435-50.

Anderson L.K., J.R., Taylor and R.J., Holloway (1966) "The Consumers and His Alternatives: An Experimental Approach," Journal of Marketing Research, Vol.3, pp 62-67.

Are There Consumer Types, New York, Advertising Research Foundation, (1964).

Bass F.M., E.A., Pessemier and D.R., Lehmann (1972) "An Experimental Study of Relationship Between Attitude, Brand Preference and Choice," Behavioral Science, Vol.17, No.6.

Bass F.M., "The Theory of Stochastic Preference and Brand-Switching," Journal of Marketing Research, Vol. XI (February 1974), pp 1-20.

Bass F.M., (1974) "An Empirical Analysis of the Theory of Stochastic Preference," Working Paper No. 455, Krannert Graduate School of Industrial Administration, Purdue University.

Bass F.M., A. Jeuland and G.P., Wright (1976) "Equilibrium Stochastic Choice and Market Penetration Theories: Derivation and Comparison," Management Science, Vol.22 #10, pp 1051-64.

Best R., (1976) "The Predicted Aspects of a Joint-Space Theory of Stochestic Choice", Journal of Marketing Research, Vol. XIII, No.2, pp 198-204.

Birch M.W., (1963) "Maximum Likelihood in Three-Way Contingency Tables, Journal of The Royal Statistical Society, Ser. B, 25, pp 220-33.

Blattberg R.C., and S.K., Sen (1975) "A Bayesian Technique To Discriminate Between Stochastic Models of Brand Choice," Management Science, Vol.21, No.6.

_____ (1974) "Market Segmentation Using Models of Multidimensional Purchasing Behavior," Journal of Marketing, Vol. 38, pp 17-28.

Box M.J., (1965) "A New Method of Constrained Optimization and a Comparison with Other Methods," The Computer Journal, 8, pp 42-52.

Brown G., (1952-3), "Brand Loyalty - Fact of Fiction?", Advertising Age, Vol. 23, June 9, June 30, October 6, December 1, 1952; Vol.24, January 25, 1953.

Brown K.M., (1973), "Computer Oriented Algorithms for Solving Systems of Simultaneous Non-Linear Algebraic Equations," Numerical Solution of Systems of Non-Linear Algebraic Equations, ed. by G.D., Byrne and C.A., Hall, Academic Press, pp 281-348.

Carman J.M., (1970), "Correlates of Brand-Loyalty: Some Positive Results," Journal of Marketing Research, Vol. VII, pp 67-76.

Carroll J.D., and J.J. Chang (1970), "Analysis of Individual Differences in Multidimensional Scaling via an N-Way Generalization of "Eckart-Young" Decomposition", Psychometrika, Vol.35, No.3.

Cinlar E., (1975), Introduction to Stochastic Processes, Prentice-Hall Inc., Englewood Cliffs, N.J.

Cox D.R., and Lewis P.A., (1968), The Statistical Analysis of Series of Events, London: Methuen & Co. Ltd.

Cunningham R.M., (1956), "Brand Loyalty - What, Where and How Much?" Harvard Business Review, Vol. XXXIV, No.1, pp 166-28.

Day G.S., (1969), "A Two-Dimensional Concept of Brand-Loyalty," Journal of Advertising Research, Vol.9, pp 29-35.

Duhamel W.F., (1966), "The Use of Variable Markov Processes As a Partial Basis for the Determination and Analysis of Market Segments," Unpublished PhD. Dissertation at the Graduate School of Business, Stanford University.

Ehrenberg A.S., (1959), "The Pattern of Consumer Purchases," Applied Statistics, Vol. VIII, pp 26-41.

_____ (1965), "An Appraisal of Markov Brand-Switching Models," Journal of Marketing Research, Vol.2, 347-62.

_____ (1972), Repeat Buying, Amsterdam, North Holland.

Engel J.F., Kollat D.T., and Blackwell R.D., (1968) Consumer Behavior, N.Y.: Holt, Rinehart and Winston.

Farley J.U., (1964), "Why Does Brand Loyalty Vary Over Products?" Journal of Marketing Research, Vol.1, pp 9-14.

Farley J.U., and W.L., Ring (1970), "An Empirical Test of the Howard and Sheth Model of Buyer Behavior," Journal of Marketing Research, 7, pp 427-38.

Feller W., (1957), An Introduction to Probability Theory and Its Applications, Vol.I, Second Edition, N.Y. John Wiley and Sons, Inc.

Fishbein M.A., (1963) "A Consideration of Beliefs, Attitudes and their Relationships," in I.D., Steiner and M.A., Fishbein, eds., Current Studies in Social Psychology, N.Y., Holt, Rinehart and Winston, pp 107-20.

Frank R.E., (1960), "Prediction and Brand Choice", Unpublished Dissertation (PhD.), University of Chicago.

_____, and H., Boyd (1965), "Are Private-Brand Prone Grocery Customers Really Different," Journal of Advertising Research, Vol.5, pp 27-35.

_____, P.E., Green and H.F., Sieber (1967), "Household Correlates of Purchase Price for Grocery Products," Journal of Marketing Research, 4, pp 54-8.

_____, (1967), "Is Brand-Loyalty a useful Basis for Market Segmentation," Journal of Advertising Research, 7, pp. 27-33.

_____, (1967a), "Correlates of Buying Behavior for Grocery Products, Journal of Marketing, 31, pp 48-53.

_____, and P.E., Green (1968), "Numerical Taxonomy In Marketing Analysis: A Review Article," Journal of Marketing Research, 5, pp 81-94.

_____, W.F., Massy and T.M., Lodhal (1969), "Purchasing Behavior and Personal Attributes," Journal of Advertising Research, Vol.9, pp 15-24.

Goodman L.A., (1968), "The Analysis of Cross-Classified Data: Independence, Quasi-Independence, and Interactions in Contingency Tables with or without Missing Entries," Journal of the American Statistical Association, 63, pp 1091-1131.

_____, (1974), "Interactions in Multi-dimensional Contingency Tables", Annals of Mathematical Statistics, 35, pp 632-46.

_____, (1969), "On Partitioning Chi-Square and Detecting Partial Association in Three-way Contingency Tables", Journal of the Royal Statistical Society, Ser. B., 31, pp 486-98.

_____, (1970), "The Multivariate Analysis of Qualitative Data: Interactions Among Multiple Classifications", Journal of the American Statistical Association, 65, pp 226-56.

_____, (1971a), "The Analysis of Multi-Dimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications", Technometrics, 13, pp. 33-61.

_____, (1971b), "Partitioning of Chi-square, Analysis of Marginal Contingency Tables, and Estimation of Expected frequencies in Multi-dimensional Contingency Tables", Journal of the American Statistical Association, 66, pp 339-44.

_____, (1972), "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables", American Sociological Review, 37, pp 28-46.

_____, (1972a), "A General Model for the Analysis of Surveys", American Journal of Sociology, 77, pp 1035-86.

Green P.E., and Carmone F.J., (1969), "Multidimensional Scaling: An Introduction and Comparison of Non-Metric Unfolding Techniques", Journal of Marketing Research, 4, pp 330-41.

Green P.E. (1975), "Marketing Application of MDS: Assessment and Outlook", Journal of Marketing, 39, pp 24-31.

Greenwood M., and G.U., Yule (1920), "An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings, with Particular Reference to the Occurence of Multiple Attacks of Desease or of Repeated Accidents", Journal of the Royal Statistical Society, Vol. 83, pp 225 ff.

Haines G.H., (1964), "A Theory of Market Behavior After Innovation", Management Science, Vol. X, No. 4, pp 634-58.

Hald A., (1967), Statistical Theory with Engineering Applications, J. Wiley & Sons, Inc.,N.Y.

The Hendry Corporation (1966), "Hendrodynamics: Fundamental Laws of Consumer Dynamics".

Hansen F., (1976), "Psychological Theories of Consumer Choice", Journal of Consumer Research, Vol. 3, No. 3, pp 117-42.

Herniter J., (1973), "An Entropy Model of Brand Purchase Behavior", Journal of Marketing Research, pp 361-75.

_____, (1974), "A Comparison of the Entropy Model and the Hendry Model", Journal of Marketing Research, February, pp 21-29.

_____, (1971), "A Probabilistic Market Model of Purchase Timing and Brand Selection", Management Science, Vol. 18, No. 4, Part II.

Howard J.A., and J.N. Sheth (1969), The Theory of Buyer Behavior, N.Y., Wiley.

Howard R.A., (1963), "Stochastic Process Models of Consumer Choice Behavior", Journal of Advertising Research, Vol. 3, No. 3, pp 35-42.

_____, (1965), "Dynamic Inference", Journal of the Operations Research Society of America, Vol. 13, No. 2, Sept., pp 712-33.

Hyman G.M., (1969), "The Calibration of Trip Distribution Models", Environment and Planning, Vol. 1, pp 105-112.

Jaynes E.T., (1957), "Information Theory and Statistical Mechanics", Physical Review, 106, pp 620-30.

Jeffreys H., (1961), Theory of Probability, (3rd ed.), Oxford: Clarendon.

Johnson R.M., (1970), "Multiple Discriminant Analysis: An Application to Marketing Research", Chicago, Market Facts Inc.

Johnson S.C., (1967), "Hierarchical Clustering Schemes", Psychometrica, 32, pp 241-54.

Jones J.M., (1969), "A Non-Stationary Probability Diffusion Model of Consumer Brand Choice Behavior", Unpublished Doctoral Dissertation, Stanford University.

Jones J.M., (1970), "A Dual Effects Model of Brand Choice", Journal of Marketing Research, Vol. 7, No. 4, pp 458-64.

_____, (1971), "A Stochastic Model for Adaptive Behavior in a Dynamic Situation", Management Science, Vol. 17, No. 7, pp 487-97.

_____, (1973), "A Composite Heterogeneous Model for Brand Choice Behavior", Management Science, Vol. 19, No. 5, pp 499-509.

Karlin S., and Taylor H.M., (1975), A First Course in Stochastic Processes, Academic Press, N.Y.

Khinchin A.I., (1957), Mathematical Foundations of Information Theory, N.Y., Dover Publications.

Kuehn A.A., (1958), "An Analysis of the Dynamics of Consumer Behavior and its Implications for Marketing Management," Unpublished Ph.D. dissertation, Graduate School of Industrial Administration, Carnegie Institute of Technology.

_____, (1962), "Consumer Brand Choice - A Learning Process?" Journal of Advertising Research, Vol.II, pp 10-17.

_____, and A.C., Rohloff (1965), "New Dimensions in Analysis of Brand-Switching", Presented at the Twenty-Seventh National Meeting of the Operations Research Society of America, Boston, Massachussets, May 7.

_____(1967), "Consumer Response to Promotions", in P.J., Robinson et al., Promotional Decisions Using Mathematical Models, Boston: Allyn and Bacon, Inc.

Kruskal J.B., (1964), "Multidimensional Scaling by Optimizing Goodness of fit to a Non-Metric Hypothesis", Psychometrica, 29, pp 1-27.

Lancaster K.J., (1966), "A New Approach to Consumer Theory", Journal of Political Economy, 24, pp 132-34.

Leeflang P.S., (1974), Mathematical Models in Marketing, Leiden University Press, The Netherlands.

Lehmann D.R., (1971), "Television Show Preference: Application of a Choice Model", Journal of Marketing Research, pp 47-55.
pp 331-34

_____, (1972), "Judged Similarity and Brand-Switching Data as Similarity Measures", Journal of Marketing Research, Vol.9, pp 331-34.

_____, and E.A., Pessemier (1973), "Market Structure Via Clustering and Discriminant Analysis", Working Paper No. 407, Krannert School of Industrial Administration, Purdue University.

Lilien G., (1974), "Application of a Modified Linear Learning Model of Buyer Behavior", Journal of Marketing Research, Vol.11, pp 279-85.

Lindley D.V., (1964), "The Bayesian Analysis of Contingency Tables", Annals of Mathematical Statistics, 35, pp 1622-43.

Lipstein B., (1965), "A Mathematical Model of Consumer Behavior", Journal of Marketing Research, Vol.2, pp 259-65.

Massy W.F., (1960), "Innovation and Market Penetration", unpublished Ph.D. dissertation, Department of Economics, Massachussets Institute of Technology.

_____, and R.E., Frank (1965), "Short Term Price and Dealing Effects in Selected Market Segments", Journal of Marketing Research, Vol.2, (May 1965), pp 171-85.

_____, (1966), "Analysing Product Profiles with Orthogonal Discriminant Function", Proceedings, National Meeting, Business and Economic Statistics Section, A.S.A., pp 144-57.

_____, (1966), "Order and Homogeneity of Family Specific Brand-Switching Processes", Journal of Marketing Research, Vol.III, No.1, pp 48-54.

_____, (1967), "A Stochastic Evolutionary Model for Evaluating New Products", Presented at the American Meeting of the Institute of Management Sciences, Boston, Massachussetts, April 5-7.

_____, (1967a), "Linear Learning Models for Brand Choice", Working Paper No. 133, Graduate School of Business, Stanford University.

_____, D., Montgomery and D., Morrison (1970), Stochastic Models of Buyer Behavior, the M.I.T. Press, Cambridge, Massachussetts.

McConnell D., (1968), Repeat-Purchase Estimation and the Linear Learning Model", Journal of Marketing Research, Vol.V, pp 304-6.

Montgomery D.B., (1966), "A Probability Diffusion Model of Dynamic Market Behavior", Working Paper No. 205-66, A.P., School of Management, M.I.T., May.

_____, (1967), "Stochastic Modeling of the Consumer", Industrial Management Review (Spring), pp 31-42.

_____, (1969), "A Stochastic Response Model with Application to Brand Choice", Management Science: Theory, Vol.XV, No.7, March.

Morrison D.G., (1965a), "Stochastic Models for Time Series with Application in Marketing," Technical Report, No.8, Program in Operations Research, Stanford University.

_____, (1965b), "New Models in Consumer Behavior: Aids in Setting and Evaluating Market Plans", Proceedings: Fall Conference of the American Marketing Association, September 1-3, Washington, D.C.

_____, (1966a), "Interpurchase Time and Brand-Loyalty". Journal of Marketing Research, Vol.III, pp 289-91.

_____, (1966b), "Testing Brand-Switching Models". Journal of Marketing Research, Vol.III, pp 401-09.

Nakanishi M., "Advertising and Promotion Effects on Consumer Responses to New Products", Journal of Marketing Research, 10(August 1973), pp 242-9.

Pessemier E.A., Burger P., Teach R., and Tigert D., (1971), Using Laboratory Brand Preference Scales to Predict Consumer Brand Choice", Management Science, pp B371-85.

_____, (1972), "Multi-attribute Models for Predicting Individual Preference and Choice", Institute Paper No. 346, Krannert Graduate School of Industrial Administration, Purdue University, March.

_____ and W.L., Wilkie (1972), "Multi-Attribute Choice Theory - A Review and Analysis, Working Paper No.372, Krannert Graduate School of Industrial Administration, Purdue University.

Rosenberg M.J., (1960), "A Structural Theory of Attitude Dynamics," Public Opinion Quaterly, 24, pp 319-40.

Rosenbrock H.H., (1960), "An Automatic Method for Finding the Greatest or Least Value of a Function", The Computer Journal, 3, pp 1975-84.

Shannon C., and Weaver W., (1949), The Mathematical Theory of Communication, Urbana: University of Illinois Press.

Shepard R.N., (1962), "The Analysis of Proximities: Multi-Dimensional Scaling with an Unknown Distance Function, I", Psychometrica, 27, pp 125-39.

Sheth J.M., and Talarzyck W.W., (1970), "An Investigation of Relationships among Evaluative Beliefs, Affects, Behavioral Intentions and Behavior", Working Paper, University of Illinois.

Simon H.A., (1968), "On Judging the Plausibility of Theories", in Logic, Methodology and Philosophy of Sciences III. Amsterdam: North Holland, Pub. Cy.

Telser L., (1963), "Least Square Estimates of Transition Probabilities", in Measurement in Economics, Christ et al., Stanford, California: Stanford University Press, pp 270-92.

Theil H., (1967), Economics and Information Theory, Chicago: Rand McNally and Company, Amsterdam: North-Holland, Pub. Cy.

Theil H., (1972), Statistical Decomposition Analysis, North-Holland Publishing Company, Amsterdam.

Weiss D.L., Houston F., and P.M., Windal (1976), "The Periodic Pain of Lydia Pinkham", Forthcoming in the Journal of Business.

WilkieDW.L., and E.A. Pessemier (1973), "Issues in Marketing Use of Multi-Attribute Attitude Models", Journal of Marketing Research, 10, pp 428-41.

Wilson A.G., (1970), "The Use of the Concept of Entropy in System Modeling", Operational Research Quaterly, Vol.21, No.2.

Windal P.M., and D.L. Weiss (1976), "Stochastic Response Models - Perspective and Comment", Proceedings of the Annual Conference of the Marketing Division of the Canadian Association of Administrative Sciences, University Laval, Quebec.

APPENDIX A

MOTIVATION FOR THE SIMPLIFYING ASSUMPTION USED IN CHAPTER IV (EQUATION 4.5)

Let

P_{ij} = Proportion of consumers purchasing brands i and j on two successive occasions.

$A = [A_{ij}]$ = n x n similarity matrix whose typical element is A_{ij}.

$B = \{1, \ldots, n\}$ = a set of n brands.

Assume

A_1 $\quad P_{ij} = f_{ij}(A)$ for some function f_{ij}, $i, j \in B$

A_2 $\quad A_{ij} \geqslant 0, \quad \sum_j A_{ij} \neq 0, \quad \sum_i A_{ij} \neq 0, \quad \forall i, j \in B$

A_3 $\quad f_{ij}(A) = 0$ iff $A_{ij} = 0 \quad \forall i, j$

A_4 $\quad f_{ij}(A + \Delta E_{kl}) - f_{ij}(A) = g_{ij}(\Delta) \quad \forall k, l, i, j \in B$
$k, l \neq i, j$

where E_{kl} is a n x n matrix whose only non-zero element is the (kl)th.

Assumption A1 postulates the existence of a functional relationship between the proportion P_{ij} and a similarity matrix A. By introducing this assumption, we need not worry about A provided it satisfies A2. In A2, we require that all entries be non-negative and for any fixed i, that $A_{ij} \neq 0$ for at least one j, j = 1,...,n, and similarly for fixed j, that $A_{ij} \neq 0$ for at least one i, i=1,...,n. Thus, if A_i is the ith row of A, A2 states that A_i is not the zero vector. This rather inconsequential assumption is used to simplify the analysis. Assumption A4 is of critical importance. It states that the change in the proportion of consumers switching from brand i to brand j due to a change Δ in the similarity measure A_{kl} of some pair of brands $k, l \neq i, j$ is a function of Δ but not of k and l.

<u>Proposition</u>

If $P_{ij} = f_{ij}(A)$ for some function f_{ij} and some matrix $A = [A_{ij}]$, i,j = 1,..., n, satisfying assumptions A3 through A4, then

$$P_{ij} = f_{ij}\ (A_{ij},\ \{\alpha_i\},\{\beta_j\})$$

Where $\{\alpha_i\} \doteq (\alpha_1,\ldots,\alpha_n) = \sum_j v_j A_{ij} \qquad j \in B$

$$\{\beta_j\} \doteq (\beta_i,\ldots,\beta_n) = \sum_i w_i A_{ij} \qquad i \in B$$

for some $w_i \geq 0$ and $v_j \geq 0$ and not all zero.

Proof

Consider the set

$$S = \{A: A_{11} \text{ is constant and } \sum_i w_i A_{ij} = \beta_j,$$

$$\sum_j v_j A_{ij} = \alpha_i, \text{ for some } w_i \geq 0,\ v_j \geq 0$$

$$\text{and not all zero, } i, j = 1,\ldots, n\}$$

The non-negativity restriction on the weights w_i and v_j together with assumption A2 ensure that the elements of the sets $\{\alpha_i\}_{i=1}^n$ and $\{\beta_j\}_{j=1}^n$ are themselves non-negative and not all identically zero. Let $\bar{A}$, $\bar{\bar{A}} \in S$ and $\bar{A} \neq \bar{\bar{A}}$ component-wise.

Then it will be shown that $f_{11}(\bar{A}) = f_{11}(\bar{\bar{A}})$ from which it may be concluded that $f_{ij}(A)$ is a function only of A_{ij} and the sets $\{\alpha_i\}_i$, $\{\beta_j\}_j$.

Let $A^o = \min(\bar{A}, \bar{\bar{A}})$ taken component-wise and E_{ij} be a matrix whose only non-zero element is the (ij)th one, i.e.

$$E_{kl} = \begin{cases} 0 & \text{if } k \neq i \text{ and/or } l \neq j \\ 1 & k = i \text{ and } l = j \end{cases}$$

Then if b_o is defined as the smallest non-zero element of the two matrices:

$(\bar{A} - A^o, \bar{\bar{A}} - A^o)$, $\exists$ some i, j, k, l such that we can define

$$\bar{A}^1 = A^o + b^o E_{ij}, \quad \bar{A}^1 \leqslant \bar{A}$$

$$\bar{\bar{A}}^1 = A^o + b^o E_{kl}, \quad \bar{\bar{A}}^1 \leqslant \bar{\bar{A}}$$

where either $\bar{A}^1_{ij} = \bar{A}_{ij}$ or $\bar{\bar{A}}^1_{kl} = \bar{\bar{A}}_{kl}$

By assumption A4 $f_{11}(\bar{A}^1) = f_{11}(\bar{\bar{A}}^1)$.

Now define b^1 as the minimum non-zero element of

$(\bar{A} - \bar{A}^1, \bar{\bar{A}} - \bar{\bar{A}}^1)$ and form

$$\bar{A}^2 = \bar{A}^1 + b^1 E_{ij} \text{ for some } i,j, \quad \bar{A}^2 \leqslant \bar{A}$$

$$\bar{\bar{A}}^2 = \bar{\bar{A}}^1 + b^1 E_{kl} \text{ for some } k, l, \quad \bar{\bar{A}}^2 \leqslant \bar{\bar{A}}$$

Again, by A_4 $f_{11}(\bar{A}^2) = f_{11}(\bar{\bar{A}}^2)$.

Since the number of zero elements of $(\bar{A} - A^{-k}, \bar{\bar{A}} - \bar{\bar{A}}^k)$ increases by at least one at each iteration of this procedure, and

$$f_{11}(A^k) = f_{11}(A^k) \quad \forall k,$$

we have:

$$\bar{A}^k = \bar{A} \text{ and } \bar{\bar{A}}^k = \bar{\bar{A}} \text{ for } k \geqslant n^2 - 1.$$

Thus $f_{11}(\bar{A}) = f_{11}(\bar{\bar{A}})$ as required, establishing the claim that the function $h_{ij}(A)$ is constant over the set S and hence depends only upon the quantities A_{ij}, $\{\alpha_i\}$ and $\{\beta_j\}$. So we can write:

$$f_{ij}(A) = f_{ij}(A_{ij}, \{\alpha_i\}, \{\beta_j\}).$$