Chapter 1

Introduction

One of the basic targets of market research is the investigation of market and the conception of buying behaviour. In previous years, managers could understand consumer behaviour from the daily sellings. Nowadays, the huge number of companies and markets deprive managers of a direct contact with their buyers. Market experts believe that a successful strategy in the promotion of a product can be achieved by taking into consideration customer behaviour. Thus, marketers try to investigate the relation between the irritants of marketing and the customer's response. The following figure illustrates the irritants that affect consumer behaviour.

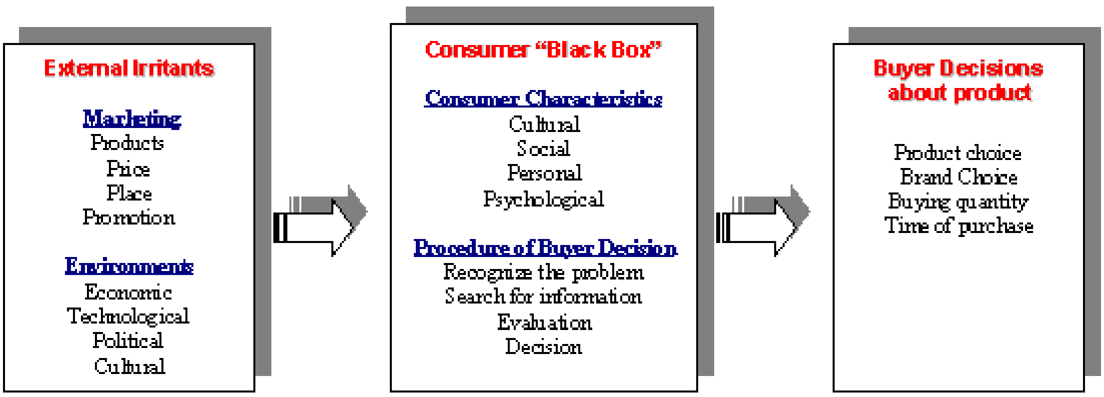

Figure 1: The irritants that affect consumer behaviour

It is obvious that marketing and environments irritants (product, price place, promotion, etc.) were imputed in the "black box" of buyer and determine the decision of the buyer (choice of products, brand choice, etc.).

The above figure (1) gives us the information that buyers decisions are affected by cultural, personal , social and psychological factors. These factors are presented analytically in the following figure.

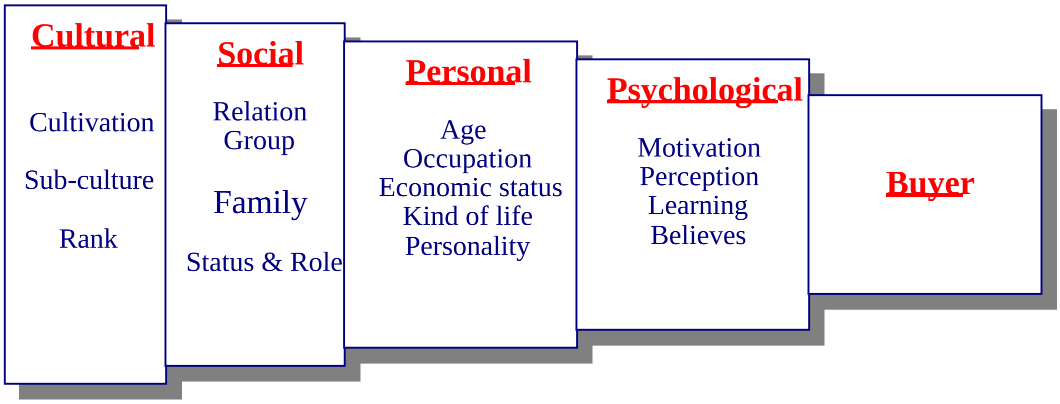

Figure 2: Factors that affect consumer decisions

Most of these factors are not controlled by managers or those who make the decisions concerning the promotion of a product. It is important to take into consideration all these factors in order to apply the most successful strategy. In this part of management in a company, the contribution of statistical science is very important. The science of statistics, by applying appropriate models could help the manager understand what is most probable to have happened in the "black box" of the potential buyer. Specifically, statistical techniques can give answers to questions, such as:

1. which factors (cultural, social, etc.) usually affect the behaviour of a buyer

2. how does the buyer decide on his purchases (definition of decision procedure).

Over the last decade there was a notable change of emphasis in the statistical analysis of survey data. Increasingly, social science researchers are finding it informative to fit probability models to their data. A wide variety of such models are now used routinely including log-linear models, survival models, discrete choice models, latent class models and many others.

Statistical modeling consists of a sequence of procedures. At a first stage, a probability model is proposed in order to describe the process of interest (Model Formulation). This formulation will be guided by substantive theory and it possibly involves a number of assumptions. At the second stage, the probability model is fitted to the observed data. The data are used to quantify the systematic relationships and random variation postulated by the model. This is achieved by estimating the unknown parameters. In the next step we make a criticism of the model. The assumptions and simplifications made to the model formulation stage are put to test. Finally, if the model is deemed to be acceptable, the model interpretation stage concludes the whole process with consideration given to the substantive significance of the results.

1.1 The Problem

An interesting subject in multivariate analysis is to explore the ways in which these techniques can be used by researchers in market research science. Specifically, we will try to investigate the development of such statistical methods in the field of marketing called consumer behaviour. The use of these statistical models in order to represent the purchase processes has been developed extremely in the last decades. These models focus on the ways in which a consumer decides on the brand (or product) he is interested in or best supplies his needs. Moreover they increase the effectiveness of consumer panels as a source of market research data for these products. The contribution of these models in the investigation is significant because they provide structural insight into the ways in which consumers regulate their choice of brands and decisions as far as purchase timing and amounts of frequently purchased products are concerned.

This dissertation focuses on the construction of the most important stochastic models as applied to buying behaviour of frequently purchased products. An attempt is made to describe the main assumptions of such models and their implications in purchasing behaviour. Moreover, brand switching models are dealt with the analysis of this kind of data and specifically via latent class analysis. Lastly, we concentrate on mixture models as applied to buying behaviour.

1.2 Terminology

Before continuing, it is important to determine some basic concepts in order to have a better understanding of this thesis project.

Event An event is defined as a single outcome of realization of the process.

State Space State space is the set of mutually extensive and collectively exhaustive events for the process.

Brand Switchers The market consists of many categories of products. Each category of products (state space) also consists of a variety of items (events). Each consumer chooses a product according to his needs. The consuming society in which a consumer is member of, offers a variety of products that satisfy the same need. Hence, the brand switchers are consumers that choose between a variety of products each time. They may use two or more brands, depending on the situation, the price or may be searching for a brand they consider to be the “best” in its category.

Brand Switching Matrix / Data Much of the information used in a number of popular brand choice models can be extracted from a brand switching matrix. A brand-switching matrix is a cross-tabulation of the number of purchases of one brand which are followed by purchases of the same or another brand. Often, this is expressed either as a conditional probability of repeating or switching brand given purchase of the first brand, or, the unconditional probability of a given sequence of two brand purchases being observed. The data that gives us information about the buying behaviour is called brand switching data.

Brand Switching Models The use of brand switching models can be seen as a suitable method for analyzing brand performance. To built a brand switching model, it is necessary to have detailed consumer panel data. From the data available, an analysis is carried out to discover the loyalty to a particular brand, how consumers are behaving in terms of switching from one brand to another, and the quantity bought per buyer.

Market segmentation This concept describes the division of a market into homogeneous groups which will respond differently to promotions, communications, advertising and other marketing mix variables. Each group, or "segment", can be targeted by a different marketing mix because the segments are created to minimize inherent differences between respondents within each segment and maximize differences between each segment.

There are many good reasons for dividing a market into smaller segments. The primary reason is that it is easier to address the needs of smaller groups of customers, particularly if they have many characteristics in common (e.g. seek the same benefits, same age, gender, etc.). Segmentation can also help avoid sending the wrong message or sending a message to the wrong people, etc.

Latent Class Models Latent variable models provide an important tool for the analysis of multivariate data and provide a way of reducing dimensionality. Bartholomew and Knott (1998) support a classification of latent variable methods. This classification is presented below:

Classification of Latent variable Methods			
	Manifest Variables		
		Metrical	Categorical
Latent Variables	Metrical	Factor Analysis	Latent Trait Analysis
	Categorical	Latent Profile Analysis	Latent Class Analysis

In Latent Class Analysis one tries to find theoretically meaningful discrete latent variables, each having two or more latent categories or classes, that explain the relation among the categorical manifest variable (starting with a multivariate frequency table). The latent class model postulates that the observed dependence between the rows and columns of the table is due to the presence of underlying latent classes where, by the assumption of local independence, the rows and columns are independent within any latent class. More information, about latent class analysis and the way in which this technique is applied in behavioural data, is given in the following chapters.

Heterogeneity Brand choice models have been developed at the market level to describe and predict market phenomena such as switching, market share, penetration and segmentation. This aggregate level analysis raises the issue of how much heterogeneity among households in brand choice affects one's picture of aggregate brand switching. When we say "heterogeneity", we mean differences in long term brand preferences among households, which leads to differences in the relative number of purchases made over time of each brand in a competing set of brands. If households are very heterogeneous, inferences about how much one brand competes with another are likely to be misleading. One brand may compete with another for some of the buyers but not for others.

The lack of sufficient data for many markets makes the measuring of heterogeneity at the household level impossible. More often heterogeneity has been measured instead at the aggregate, or across-household level, using brand switching matrices which have been reconstructed from survey data on only the two most recent brand purchases.

Trivedy and Morgan (1996) developed a model, that formulates a simple and less restrictive representation of consumer heterogeneity allowing a more managerially relevant interpretation and implementation while only requiring aggregate brand switching data. According to the authors heterogeneity with respect to purchase probabilities, implies that θ_i is distributed over the population, so we say that $E(\theta_i) = \theta_i$. If $E(\theta_{ij})$ represents the expectation of switching from i to j in subsequent trials, total switching would be $\sum\sum_{i-j} E(\theta_{ij})$. Thus, $E(\theta_{ij}) = A_{ij}\theta_i\theta_j$, where A_{ij} is a parameter representing heterogeneity in the population.

1.3 Plan of Thesis

As, we also mentioned in the previous pages, in this thesis we will attempt a presentation of some of the most important statistical models applied in order to analyze consumer behaviour. Primarily, we present a brief survey of the development of stochastic models in buying behaviour. Continuing, we attempt to focus our survey in the use of latent class analysis for the analysis of brand switching data. In particular, chapter 2 of this thesis presents some general concepts on statistical modeling of buying behaviour. Moreover, this chapter provides a history of the stochastic models as were developed in order to study consumers behaviour in depth. In chapter 3, we present models which focus on "when" purchase events take place or "how much" products will be purchased in a given time interval. (These models are called purchase incidence models). Chapters 4 deals primarily with brand choice models (Zero Order, Markov and Linear Learning models) in general. Continuing, in chapter 5 we describe the most important concepts as far as the latent class analysis is concerned. Moreover, we deal with latent class models that are applied to buying behaviour. Finally, we present a comparative analysis of the brand choice models.

Chapter 2

Approaches to Stochastic Modeling

The most frequent question that can be asked about consumer choice behaviour, or any human behaviour, is whether that behaviour is at least partially stochastic and/or whether there exist causes and explanations for all behaviour. Freud and many other scientists believed that an explanation exists for all behaviour even if the explanation must be sought in the unconscious. This finding has lead to the development of some empirical studies. Although, these studies have many supporters it is obvious that the incapability of present models to explain a substantial portion of the variance is reason for disappointment. Some evidence suggests that brand choice behaviour is substantially stochastic.

Bass (1974) maintained that the fact that the choice behaviour of individual consumers is substantially stochastic does not mean that it is fruitless to study this behaviour. It is useful to seek the major influences which determine the structure of stochastic preference. According to Bass, attribute studies in particular and the study of the important dimensions of the choice process contribute to an understanding of consumer behaviour and are managerially relevant with respect to product strategy.

2.1 Type of Models and Characteristics of Stochastic Models of Buying behaviour

Market researchers are finding it informative to give emphasis to the statistical analysis of market data by fitting probability models. The models for behavioural phenomena may have Deterministic or Probabilistic character. Another separation, is that the stochastic models may deal with either individual or aggregate behaviour.

When a consumer makes a purchase he makes several simultaneous decisions: deciding what to purchase, when to purchase, and how to purchase. Consequently, stochastic choice models (stochastic models of Buying behaviour) have generally been divided into two classes:

1. purchase - timing (incident) models which focus on "when" purchase events take place or on "how much" will be purchased in a given time interval

2. brand choice models, which focus on "what" to purchase (Stochastic models of Brand Choice)

Thus, the purchase incidence models are used to predict when a purchase of a product will occur or how many purchases will occur in a specific time interval. The brand choice models help market researcher predict which brand (or group of brands) will be purchased by the consumers, given that a purchase does not occur in a particular time.

The distinction between brand choice and purchase incident models is important for several reasons. In order to find out the differences between these types of models we define the following:

Let B_i the purchase of brand i, and P the purchase of any brand of the product class. The probability that an individual will purchase brand i in the interval of time between t and $t+h$ is:

$$\Pr\{B_i \in (t, t+h)\}$$

If we expand this probability according to the rule of conditional probability obtain :

$$\Pr\{B_i \in (t, t+h)\} = \Pr\{B_i|P \in (t, t+h)\} \Pr\{P \in (t, t+h)\} \tag{2.1}$$

This means that the probability of brand i being purchased between (t,t+h) is equal to the conditional probability brand i being purchased between time (t,t+h), given that a purchase occurs, multiplied by the probability of a purchase occurring during this time interval.

In the above equation, the $\Pr\{B_i|P \in (t, t+h)\}$ is a brand choice probability, but $\Pr\{P \in (t, t+h)\}$ is a purchase incident probability. Hence, models that work on predicting $\Pr\{B_i \in (t, t+h)\}$ are called *brand choice models* and models that predict $\Pr\{P \in (t, t+h)\}$ are called *purchase incident models*.

Brand choice models can usually be distinguished according to how they deal with:

1. **Population Heterogeneity**. This refers to differences in long term brand preferences among households, which lead to differences in the relative number of purchases made over time of each brand in a competing set of brands. Many stochastic models of brand choice attempt to take into account population heterogeneity. This can be done by the following ways: a) Certain specific determinants of the purchase probabilities can be identified and built into the model. b) Households can be divided into segments according to their purchases. On the basis of these variables, estimation of the parameters of the model for each segment can be made. c) Assumption that the parameter has a distribution of values in the population (prior distribution) can also be made. Each individual in the sample is assumed to represent a random draw from the prior distribution of possible parameter values. Since heterogeneity has become an important topic in marketing this is referred to this extendedly in the following pages.

2. **Purchase-event feedback**. Some models assume that the purchase of a product has a direct effect on the household's subsequence purchase probabilities. There are

models that assume no purchase-event feedback. These models are called *zero-order models*, because the purchase probability of the brand on the (n+k)th occasion, p_{n+k} is equal to the purchase probability of the brand on the nth occasion p_n. Also, the *first order Markov model* assumes that only the last purchase occasion determines the next purchase. Additionally, the *Linear Learning Models* use the purchase event feedback in order to make a better prediction on buying behaviour.

3. **Time effects.** The effects of time is a significant factor in the use of brand choice models. It is common and sometimes easier to assume that the net effect of such forces on the purchase probabilities can be summed up by a time trend term in the model. The probability diffusion models and some of the market penetration models can be viewed in ths light.

4. **exogenous market factors**. Exogenous factors in consumer behaviour are those factors which existed in the mind of the consumer before he/she encounters the offers of suppliers in the market. In this sense, they influence the plans and ideas of the consumer from outside and determine his/her choice in broad terms. Some of these factors are the following:

    - Consumer characteristics (culture, sub-culture, social class of consumers)
    - Product characteristics (price, product differentiation, etc.)
    - Effect of technology

2.2 Heterogeneity of Consumer Brand Choice

As stressed in the previous section, heterogeneity is one of the most important topics in marketing. In the last decades many researchers have included heterogeneity in studying consumer brand choice processes. This occurs since not taking a notice of heterogeneity may lead to wrong conclusions on the consumers behaviour. Furthermore, ignoring heterogeneity may lead to biased results and wrong inferences concerning marketing strategies that must be followed. According to previous experience, heterogeneity can be distinguished as observed or unobserved. Examples of observed heterogeneity are demographic characteristics, habits etc. whereas unobserved heterogeneity is the heterogeneity that may exist in a segment consisting of consumers with common demographic characteristics, habits, etc.

According to Bemmaor and Schmittlein (1991) the total variance that is present in a purchase behaviour can be defined as the sum of within -household variance and between - household variance (and residual variance). Specifically, the within- household variance is known to us as non- stationarity, but the between- household variance is heterogeneous. Despite the recent advances little is known about the heterogeneity present between households. Most previous research has modeled unmeasured heterogeneity as completely unobservable, or has attempted to reduce variability by introducing terms such brand loyalty, (Guadagni and Little, 1983).

Moreover, other researchers have included heterogeneity in their models by various ways. Thus, we have the following approaches:

1. models in which heterogeneity can be specified inside or outside the likelihood function.

2. models in which heterogeneity can be modeled using a fixed or a random effects specification

3. models in which heterogeneity can be included as a random intercepts or random coefficients

4. models in which heterogeneity can be modeled parametrically or non-parametrically

2.3 Survey of the Literature

If we make a retrospection of the past, we will see that the development and application of stochastic models in the buying behaviour began very early. From the literature, it is obvious that researchers were always interested in extracting meaningful information concerning customer behaviour.

Lipstein (1959) was one of the first who tried to apply a first order Markov model to brand choice behaviour. Kuehn (1962) also adapted the linear learning models to the same problem. Markov and Linear Learning models (LLM) incorporate assumptions about the effects of purchase - event feedback but do not make provisions on time effects and population heterogeneity. The work of Kuehn on linear learning models led Frank (1962) to the development of an alternative learning model. His approach postulated a Bernoulli model for each household but assumed that the brand choice probabilities for different households were likely to differ from one to another.

Morisson (1965 a,b) developed the Compound Bernoulli model. This model is essentially the same as the model proposed by Frank. The advantage of compound Bernoulli model is that it lends itself for a much better statistical analysis. Only short purchase histories of individual consumers would be used (thus, the stationarity assumption of it was not strained too badly). In this approach Morrison used the term "Compound". This term denotes the fact that explicit provision for a distribution of relevant parameter values is included in the model. Morisson also developed the compound Markov model. This model differs from the compound Bernoulli model only in terms of the order of the individuals purchasing process. In the compound Markov model only the last purchase influences the current purchase behaviour. Especially, the compound Markov model allows for the first order behaviour, and compound Bernoulli model allows for zero - order behaviour. Massy (1966), applied the approach followed by Frank to the Markov model in order to test for population heterogeneity effects. He also developed his own version of compound linear learning model.

The contribution of Coleman (1964) to the development of stochastic brand - choice models, combining time effects and population heterogeneity, is also very important. The contributions of Howard (1965) and Montgomery (1966) were also important. Specifically, Howard assumes that households reevaluate the worth of various brands at discrete points in time, that the outcomes of the successive evaluation (drawn from the distribution of probabilities) are independent of each other, and the purchase probability between reevaluations does not change. This is justified by the fact that the process of purchases follows a Bernoulli distribution. In contrast, Coleman's change and response uncertainty model and probability diffusion model postulate that these "reevaluations" occur continuously and the changes in probabilities per unit time are small and are not independent of the past evaluations..

Howard (1963), Telsert (1963) and Lipstein (1965) developed models that include time effects and purchase event feedback but not population heterogeneity. Especially, Howard postulated a Markov model where transition probabilities are related to the time the last purchase occurred. Telsert also developed a Markov model in which the parameters are functions of marketing variables (e.g. advertising). Finally, Lipstein developed a Markov model in which the matrices are estimated from data covering two different time periods.

Duhamel (1966) tried to estimate Telsert's variable Markov model using individual household data. Kuehn and Rohloff (1967) occupied themselves with the development of a learning model where the learning operators are functions of the time since the last purchase. Jones (1969) extended Montgomery's probability diffusion model (extension of cleman's model) so as to include learning characteristics. The common characteristic of the last tree model is that the authors tried to include all four of the characteristics (population heterogeneity, purchase event feedback, time effects and exogenous market factors). A detailed description of the above models can be found in the book of Massy, Montgomery and Morrison, published in 1970.

Jones (1969) also took a major step forward by allowing individuals to differ in the order of the stochastic process that follows as well as in the process parameters. In this model individuals are allocated to Bernoulli, Markov and Linear Learning segments, each segment heterogeneous within itself. Givon and Horsky (1978) developed a model which assumes the individual follows either the Bernoulli, Markov or Linear Learning process as well as have different parameters from other people following the same process. The same authors, in 1979, presented an application of a composite stochastic model that allowed individuals to differ in the order of the stochastic process.

Latent class analysis was proposed by Lazarsfeld (1950). This procedure enabled the researcher to take data from multiway contingency table describing aggregate responses and decompose the tabular frequencies into a set of latent classes or segments that displayed certain characteristics. Despite its promise, latent class analysis had seen little use in the social sciences up to 1968 where Myers and Nicosia made use of the model in marketing. Green, Carmone and Washspress (1976) tried to make a non technical description of latent class analysis, because they judged that the algorithms that had been proposed were complex computationaly and limited in scope.

During the last decades the studies in marketing that used latent class analysis have seen little use. Grover and Dillon (1985) developed a probabilistic model which can be used to test alternate market structures and can be translated in terms of a restricted latent class model. Grover and Srinivasan (1987) presented a new approach to market segmentation. They developed a latent class based approach to analyze market structure by using brand-switching data. Specifically, they defined a market segment to be a group of consumers homogeneous in terms of the probabilities of choosing the different brands in a product class. Colombo and Morrison (1989) proposed a two-class "Hard Core Loyal" and "Potential Switcher" latent model for the analysis of brand switching data. They showed how this simple model can be easily estimated using a standard log-linear approach. Zahorik (1994) tried to generalize models based on latent class analysis by accounting for heterogeneity among consumers and by allowing for brand switching across clusters, in order to depict variety seeking.

With the demonstration of Goodman (1974a,b) on how to obtain maximum likelihood estimates of the model parameters, a major obstacle for the dissemination of the latent class was removed. In chapter 5, we present in more details Goodman's approach.

Eshima, Nobuoki (1993) proved that the Latent Markov Chain (LMC) model and the Latent Mixed Markov Chain (LMMC) model are equivalent, and it is shown that LMC model can be applied to various dynamic latent structure +models based on the Markov chain model.

Shigemasu and Sugiyama (1994) proposed a new model, which combines latent class analysis, and the multinomial logit model. Latent Classes of this model are introduced to explain and predict choice behaviour effectively. Chatterjee and Sudharshan (1994) studied a model for delineating product markets -MARKDEF (market definition). In MARKDEF product market definitions are analytically obtained using consumer perceptions. MARKDEF provides important diagnostic capability. It not only enables one to gain insight into the overall similarity /dissimilarity of products but also the attributes which account for these. In particular, MARKDEF gives a measure called the overall index of deviation, which reflects the overall degree to which two products differ from each other. In addition, MARKDEF provides a measure of deviation on each attribute for the products under consideration, which aids in understanding the attributes responsible for differentiating the products. MARKDEF is supported in the product class investigated, namely the methodology of comparing the results obtained from MARKDEF with those of other standard procedures like multidimensional scaling and cluster analysis and it is demonstrated that MARKDEF has good overall validity.

It is important also to refer to the most important approaches on buying behaviour that are considered significant for the modeling of consumer. In what follows we present some interesting approaches that may affect (positively or negatively) future surveys.

Lattin and McAlister (1985) developed a model of consumer variety-seeking behaviour. This model incorporates brand switching between complementary products to fulfill consumers' desires for variety. The model serves as the basis for developing a technique that allows determination of which products covered by brand switching data may be considered to be substitutable and which should be considered complementary.

Kumar and Sashi (1989) developed a probabilistic model for testing hypothesized hierarchical market structures at the aggregate level using brand switching data. Each hypothesized structure is represented by a directed graph and its parameters are approximated by the use of log-linear modeling techniques. Jain, Bass and Chen (1990) provided an iterative maximum likelihood procedure for estimating parameters of a model that incorporates heterogeneity within segments. Their work is based on the previous work of Grover and Srinivasan.

In 1993, log-linear trees were developed by Novak, and introduced as a model for analyzing market structure in brand switching data. Log- linear trees combine log linear models with a non-spatial graphical representation based on additive trees to represent market structure. Log- linear trees establish connections among previous approaches and afford certain advantages. This can be seen by considering some of the major approaches to analyzing market structure in brand switching data.

Dipak and Ram (1994) used simulated data in order to compare the GS against the JBC model for the switching data with regard to how well they recover the true underlying market structure and the parameters values and also to identify situations, if any, under which one model fits the data better than the other. The robustness of the factor analytic approach for determining the number of segments in a latent class model was also examined.

Mayekawa (1994) worked on the generation of equivalent path models in the linear structural equation models. Equivalent path models are defined as a set of different path models, which give the same value of global fit indices to a data matrix. Therefore, by definition, the data matrix at hand cannot distinguish those models. After deriving some algebraic results, a graphical method is presented, in order to generate a set of equivalent models of a given path matrix. As a result, it is demonstrated that, for most of the four-variable cases, a set of equivalent path models exists. This implies that the cause-effect terminology used in the interpretation of the result of the data analysis using the linear structural equation models in questionable in most of the cases, unless the causal order is determined prior to the analysis on some theoretic basis. Matsuda and Namatame (1995) developed a scaling technique in order to investigate the image integration of consumers, taking advantage of the recent visual and object-oriented programming technology. Moreover, Mayekawa (1996) worked on the derivation and the application of the maximization of the likelihood, which is the product of binomial or multinomial densities under the linear, constrains among the cell probabilities. This type of problem arises when we wish to smooth the histogram, to test the linear trend of proportions, or to compare the means/ variances of number- right test scores across several groups without normality assumption. In the applications of the latent variable models, it is shown that the same method can be used to estimate the distribution of the latent variables.

Chapter 3

Purchase Incidence Models

3.1 The Negative Binomial Distribution Model

As we have already mentioned, purchase incidence models are used in predicting the timing of purchase events. Among a variety of purchase incidence models, the negative binomial distribution (NBD) model is the most significant. According to Ehrenberg (1959), the NBD model assumes that the process which generates purchases at the level of the individual household is Poisson, and that the parameters of these processes are distributed over members of the population according to a Gamma distribution. More specifically, the probability that a consumer makes the next purchase during a period of time interval is independent of the previous purchase time, and the time between two purchases follows an exponential distribution.

In the following, we present the most important issues regarding the Purchase Incident Models.

3.1.1 Description of Negative Binomial Distribution Model

The Gamma-Poisson form of the Negative Binomial distribution (NBD) model generally gives a good fit to many aspects of repeat-buying behaviour for a wide range of frequently bought branded consumer goods. Ehrenberg (1959) was the first to present a well - developed stochastic model for purchase incidence (Negative Binomial Distribution Model). This author presents results showing that the negative binomial distribution tends to fit the frequency histogram for the total number of units bought by members of a consumer panel during a fixed time interval. Moreover, Ehrenberg points out that the negative binomial distribution can be derived by assuming that the process which generates purchases at the level of the individual household is Poisson, and that the parameters of these processes are distributed over members of the population according to a gamma distribution.

The main assumptions for this model are:

1. Purchases of a particular brand by a given consumer in successive equal time periods are independent and follow a Poisson distribution with a constant mean. Namely, the probability that the number of purchases in a period of unit length will be Poisson :

   $P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$, where x = 0, 1, 2, . . . , X = the number of purchases and λ is the purchase rate.

2. Different households have different purchase rate λ. Thus, the purchase rate parameter λ , governing an individual household's purchasing pattern is gamma distributed. Namely, the long average rate of purchasing varies from consumer to consumer and follows a Gamma distribution in the whole population, so we have that $g(\lambda|r,a) = \frac{a^{r}\lambda^{r-1}e^{-a\lambda}}{\Gamma(r)}$, where $\lambda > 0$, and r, α are the parameters of a Gamma Distribution.

3. Stationarity (the probability laws change as a result of consumer controlled actions).

A consequence of this mixed Poisson model is that the frequency distribution of purchases for the whole population in a given time - period follows the Negative Binomial distribution :

$$P(X = x|r,\alpha) = \int_{0}^{\infty} P(X = x|\lambda)g(\lambda|\alpha,r)d\lambda = \binom{x+r-1}{x}\left(\frac{\alpha}{\alpha+1}\right)^{r}\left(\frac{1}{\alpha+1}\right)^{x}$$

for x=0,1,... $\alpha > 0$ and r an integer.

If the period of time is t the scale parameter α is replaced by α/t and we have the following form:

$$P(X = x|r, \alpha/t) = \binom{x + r - 1}{x} \left(\frac{\alpha}{\alpha + t}\right)^r \left(\frac{t}{\alpha + t}\right)^x$$

Useful descriptive measures are the Expected value: $E(X) = E(\lambda) = \frac{r}{\alpha}$ and the variance: $Var(Y) = \frac{r}{\alpha} + \frac{r}{\alpha^2}$

The last equation can be interpreted as follows: The variance in purchasing for an arbitrary individual with the negative binomial distribution is the sum of the within individual variation $Var(\lambda) = \frac{r}{\alpha}$ plus the across individual purchase rate variability.

The major use of NBD model is to predict what would happen and to compare it with what did happen. For example, suppose a brand runs a promotion and sees a sale increase. Did this increase come from (1) non-buyers or (2) from previous heavy buyers?

Morrison and Schmittleing (1988), generalized the Negative - Binomial distribution model for the analysis of purchases.

Extensions developed by replacing some NBD assumptions

Some empirical evidence suggests that purchasing a particular brand size in successive equal time periods tends to be more regular than Poisson. Chatfield and Goodhardt (1973) proposed an alternative model, in which inter-purchase times for a given consumer are described by an Erlang distribution.

These authors dealt with the first assumption of a Negative Binomial Distribution model. This assumption, (individual consumer's purchases in successive equal time should be Poisson) can be criticized because of its implications concerning inter-purchase times. They studied three alternatives to the exponential distribution for describing the inter-purchase times of an individual consumer (Gamma, Weibul and Lognormal distributions). Continuing, they proposed some special cases of the Gamma distribution (or

Pearson type III) mainly because it is mathematically easier to handle than the other alternatives. Moreover, they studied the special case when the exponent (parameter p of Gamma distribution) is a positive integer. This distribution is usually called Erlang distribution. Hence, considering the Gamma as the distribution that describes the inter-purchase time of an individual consumer, these authors proposed an Erlang distribution for interpurchase time, yielding a 'condensed' negative binomial model. The word 'condensed' arises from the fact that the variance is less than the mean for all the values of λ, whereas the Poisson distribution has equal mean and variance.

The 'Condensed' Negative Binomial Distribution Model (CNBD) is an integrated model for the whole population. In this model two properties hold.

1. Purchases of a particular brand-size by a given consumer have Erlang inter-purchase times and the number of purchases in a given period follows the condensed Poisson distribution.

2. The average long-run rate of purchasing varies from consumer to consumer and follows a Gamma distribution in the whole population

Thus, in the given time-period the distribution of purchases in the whole population is obtained by mixing the condensed Poisson distribution with the Gamma distribution. So:

$$\Pr(r\ purchases) = P_{CN}(r) = \int_0^\infty f(\lambda)\, P_C(r) d\lambda, \qquad \text{r =0,1,. . .}$$

Thus,

$$\begin{aligned} P_{CN}(0) &= P_N(0) + \frac{1}{2}P_N(1) \\ P_{CN}(r) &= \frac{1}{2}P_N(2r-1) + P_N(2r) + \frac{1}{2}P_N(2r+1), \qquad \text{r =0,1,. . .} \end{aligned}$$

This distribution is obtained from a NBD by exactly the same weighting procedure with which a condensed Poisson distribution is obtained from the corresponding Poisson distribution. Thus, they called the distribution Condensed Negative Binomial distribution.

Some empirical evidence suggests that purchasing a particular brand size in successive equal time periods tends to be more regular than Poisson.

Morrison and Schmittlein (1982 and 1983) showed that CNBD gives a more regular sales pattern than the NBD model and gives an increasing slope to the zero class (class of individuals who will never buy).

If there exists a class of individuals who never buy, then the proper mixing distribution has a mass point representing that proportion of the population at $\lambda = 0$, and consequently a gamma distribution for the remaining population proportion.

Finally, a most difficult and perplexing problem for all stochastic models is nonstationarity, because it can occur in many different ways.

3.1.2 The Negative Binomial Distribution /Pareto Model

Schmittlein, Morisson and Colombo (1987) developed a model based on the negative binomial distribution that can be used to determine how many of a firm's current customer are 'active' determined by the transaction activity in the last year. The developed model is based on the following assumptions:

1. At the individual customer level: every active customer(when "alive") purchases is Poisson with parameter λ, and also, customers remain alive for a lifetime that is exponentially distributed with parameter μ.

2. At the population level: Gamma heterogeneity for purchasing (λ) and death (μ) rates.

Thus, the Poisson purchasing and gamma heterogeneity assumptions imply a Negative Binomial Distribution model of purchasing behaviour for active customers. Namely:

$$P(X = x|r, \alpha, \tau > T) = \binom{x + r + 1}{x} \left(\frac{\alpha}{\alpha + T}\right)^r \left(\frac{T}{\alpha + T}\right)^x$$

where

x = 0, 1, 2, . . . ,

τ = time of "death"

r, α = gamma distribution parameters for interpurchase time rate.

On the other hand, exponential "death" rates with gamma heterogeneity yield a Pareto distribution of the second kind. Namely:

$$f(\tau|s, \beta) = \frac{s}{\beta}\left(\frac{\beta}{\beta + \tau}\right)^{s+1}, \tau > 0$$

where s, β = gamma distribution parameters for lifetime length.

So, the probability that an individual is "alive" at T, using the Bayes's theorem, is

$$P(\tau > T|\lambda, \mu, X = x, t, T) = \frac{1}{1 + [\mu/ (\lambda + \mu)] \left[e^{(\lambda+\mu)(T-t)}\right]}$$

where

t = time of last purchase

T = current time

X = number of purchases the customer made in (0,T).

Then it is easy to calculate the $P(\text{"}alive\text{"}|parameters\ estimates) = P(\tau > T|r, \alpha, s, \beta, X = x, t, T)$ which is the weighted average over λ and μ of the individual level probabilities.

The utility of this model for a manager is unmeasured since it gives information concerning:

1. the estimation of the number of active customers over time
2. the identification of the P("alive"|parameters) at the individual level
3. prediction of future transaction levels.

3.1.3 Other extensions of NBD model

Brockett, Golden, and Panjer (1996) developed a series of purchase incidence models based upon the NBD model. In this paper, the authors generalized the Poisson process by replacing the exponential distribution assumption with a wide range of other, more sophisticated, distributions. Although better fitting results were presented to conclude the superiority of the new models over the traditional NBD model, the "abstract behaviours" of all the models, including the NBD model, are the same. In particular, as the purchase frequency increases, the NBD model, as well as all others, predict seeing less number of consumers in their experiments measured in a given period of time. Although the actual numbers predicted might be different, the trend of more low-frequency buyers and less high-frequency buyers was consistent among all models.

3.2 Model of Interpurchase Times at the Individual Level

The purchase timing models, such as the negative binomial distribution model (NBD) and the condensed negative binomial distribution model (CNBD) make assumptions on the distribution of interpurchase times at the individual level (exponential interpurchase times or Poisson purchases in the NBD, and Erlang 2 interpurchase times or condensed Poisson purchases in the CNBD) and use a mixing distribution on the scale parameters (gamma distribution in both the NBD and CNBD) to aggregate. In reality however, the timing of when to purchase at the individual level is complex; it depends on how the product is used and how much of the product is at hand.

Kahn (1987) developed a theoretical model of interpurchase times at the individual level that takes into consideration assumptions about how the product is used. Especially, in this model the distribution of interpurchase times is calculated based on factors such as: how the product is used, external market forces, random events etc.

Assume that the times between purchases depend on three issues:

1. The number of uses per container (There is an assumption about the size of container. Assume either that the sum of uses exceeds the amount of product in the container or is less than the amount of product in this)

2. The distribution of the amount used

3. The distribution of the times between uses

The dependent variable in this model is the number of uses per container. In order to calculate the sum of uses, Kahn tried to estimate how many uses are expected in each container. This author found that the distribution of the number of uses per container, P(n), and the expected number of uses per container, E(n), depend on:

1. The size of the container, and

2. The regularity of the amount used

The number of times between uses which are added up to form the time until the next purchase is a function of the number of uses per container. Thus, the probability density function of the interpurchase times at the individual level is a sum, over the different possible number of uses per container, of P(n) times the n-fold convolution of the interpurchase times distribution.

Thus, Kahn proved that the probability density function for the times between purchases can be written as:

$$h(t) = \sum_{n=1}^{\infty} P(n) g_n(t)$$

where: P(n) is the probability distribution of number of uses per container

$$P(n) = \int_0^K f_{n-1}(y) \int_{K-y}^{\infty} f(x)\, dx\, dy = \int_0^K f_{n-1}(y)[1-F(K-y)]\, dy = F_{n-1}(K)-F_n(K)$$

K is the size of the container, f(x) is the probability density function for the amount x consumed per usage occasion

$$f_n(x) = \int_0^x f(y) f_{n-1}(x-y)\ dy$$

and g(t) is the probability density function for times between the uses

$$g_n(t) = \int_0^t g(t') g_{n-1}(t-t') dt'$$

By deriving this theoretical distribution, Kahn also provided a mean to test the robustness of the distributional assumptions made by other researchers. This author found that, whereas the Erlang or Gamma family of distributions is fairly robust, specific shape parameters are not as robust across different product types.

3.2.1 The Usefulness of models of interpurchase times at the individual level

Chatfield and Goodhardt (1973) argued that practical interest lies more in the behaviour of groups of consumers than in that of single individuals. They found that the NBD model was robust at the individual level, and since it is a relative simple model it is sensible to use it in a frequently - purchased product category where there are many heterogeneous purchases.

It is clear that if a manager is interested in how much of his product will be purchased on the market in a fixed period of time or if he is interested in the expected number of purchases in a future period, given that a certain amount was purchased in an observed period, then the NBD model is adequate. However, there are some situations in which the manager or researcher might be interested in the distribution of interpurchase times at the individual level, and hence the Poisson assumption of the NBD is clearly undesirable.

A manager, or researcher, is interested in the distribution of interpurchase times at the individual level in the following situations:

1. to study the independence assumption between brand choice and interpurchase times. Sometimes managers want to develop methods to make the purchasing habits more or less regular in order to built up loyalty. In order to further investigate this relationship, a theoretical model of the distribution of interpurchase time at the individual level is desirable.

2. to study the effect of promotion on interpurchase times. A way to test the effects of promotions is to examine if the distribution of interpurchase times at the individual level is more irregular when purchases are made on promotion rather than when purchases are made when there is little or no promotion. Thus, in order to test these types of relationships in depth, it is desirable to have a theoretical model, which predicts purchase behaviour in the absence of any external market activity.

3. to study the robustness of the assumption of a Gamma distribution at the individual level.

Chapter 4

Brand-Choice Models

Stochastic models of brand choice can be distinguished by how they deal with purchase-event feedback and the influence of current purchase behaviour on future brand choice probabilities (Lilien, Kotler and Moorthy,1992). We will briefly review three brand-choice models: zero-order models, Markov models, and linear learning models.

4.1 Zero-Order Models

Zero-order models assume that a consumer, regardless of what he and she is exposed to, has a constant purchase probability of buying a brand. In other words, in zero-order models the purchase probability of the brand on the (n+k)th occasion, p_{n+k}, is equal to the purchase probability of the brand on the nth occasion p_n. The zero order models that are presented in this chapter, differ in the assumptions made about consumer preferences and choice and in the number of brands they consider. The most important models used in order to describe consumer behaviour are presented in the following.

4.1.1 Bernoulli Model

The simplest stochastic model for describing consumer behaviour is the Bernoulli process.

Definition 1 *The stochastic process* $\{X_t,\, t \in T\},\; X_t \in R$ *where* $R = (0,1)$*and* $T = 0, 1, 2, ...$ *is a Bernoulli process if and only if*

$$P[X_t = 1 | X_{t-1}, X_{t-2}, ..., X_{t-p}] = p$$

for all $(X_{t-1}, X_{t-2}, ..., X_{t-n}) \in R^n$*, n = 1,2,3,... .*

In the case of buying behaviour, the above implies that the probability p_i is constant over time and independent of the consumer's actual purchase decisions in the past. Thus, at any purchase decision (in a particular product category) the consumer has the same probability p_i of purchasing brand i.

It is important to remember that households usually differ in many ways, and some of these may affect use opportunities or brand preferences for particular products. Similar to other models, the Bernoulli model of buying behaviour attempts to identify, explain, or take account of population heterogeneity.

The heterogeneous Bernoulli model assumes that in a population of customers, each one has a constant probability p_i of buying one of the two brands in the market. Additionally, we don't assume that each consumer has the same p. In the case that we have a heterogeneous population, we assume that p has a Beta distribution over the individuals in the population. So, we have that:

$$f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \qquad \text{for } \alpha > 0,\; b > 0 \text{ and } 0 < p < 1$$

It is obvious that the Beta distribution is an objective measure of heterogeneity in the consumer population. The variance of the Beta distribution will give us a quantification of the heterogeneity in the population.

If we want to estimate the posterior distribution of p given the prior mixing distribution we proceed as follows:

$$f\left(p|\alpha,\beta\right) = k_1 p^{\alpha-1}\left(1-\beta\right)^{\beta-1} \text{ (prior, mixing distribution)}$$

$$p\left(r|n,p\right)=k_2p^r(1-p)^{n-r} \text{ r = 0,1,...,n (binomial likelihood distribution)}$$

Thus, using the Bayes's theorem, the posterior distribution of p is proportional to the prior distribution times the likelihood is:

$$f\left(p|n,r,\alpha,\beta\right)=k_3p^{\alpha+r-1}\left(1-\beta\right)^{\beta+n-r-1}$$

where k_1, k_2, k_3 are appropriate constants. It is obvious that the distribution of p has the same form as the prior. Namely, if $f(p)$ has a beta distribution with parameters α and β and if we observe that an individual makes r purchases of a brand out of n purchase occasions, then the posterior distribution of p is also beta with parameters $\alpha+r$ and $\beta+n-r$.

4.1.2 Simple Multiple-brand Model

Another zero order model is the simple multiple-brand. Ehrenberg (1972), postulated that the joint probability of a consumer purchasing brands i and j on successive purchase occasions is given by

$$p(i,j)=km_im_j$$

where $\{m_i\}$ are the market shares of the respective brands and k an appropriate constant. It easy to show that

$$p(i,i)=m_i-km_i(1-m_i)$$

Then, summing the above equation over brands, we get an equation for k:

$$k=\frac{1-\sum p(i,i)}{1-\sum_i m_i^2}$$

Noting again that $p\left(i\right)=m_i$, from equations $\sum_j p(i,j)=m_i$, $\sum_i p(i|j)=1$, and $p(i,i)=$

$m_i - km_i(1 - m_i)$, we obtain:

$$p(i|j) = km_i, \ \text{j} \neq i$$
$$p(i|j) = 1 - k(1 - m_i), \ \text{j=i}$$

which shows that the conditional probabilities of purchasing brand i are independent of brand j.

Kalwani and Morrison (1977) showed that two assumptions (1) a zero order process applies, and (2) switching is proportional to share, are sufficient to derive the results shown. Alternatively, Kalwani (1979) shows that if the purchase probability density function of consumers in a choice category is given by the Dirichlet distribution, then the equation $p(i, j) = km_im_j$ holds as well. There have been an number of other models and studies dealing with zero order behaviour.

4.2 Markov Models

4.2.1 Basic definitions and Mathematical properties of a first order Markov Chain

A process is called first order Markov when it has a finite number of states and every individual state depends only on its previous states. Markov chains are characterized by many properties but in the case of buying behaviour the basic concepts are: (1) the states of a Markov chain, (2) the transition matrix, (3) the steady state probabilities, and (4) stationarity.

Some basic properties of a first order Markov chain are:

The state of a Markov chain

The state of a stochastic process at time t is the value of the process at time t. Hence, if we have a sequence of purchases in time t, the state of an individual at time t is the brand that he bought at time t.

The Markov Transition Matrix

The transition matrix of a Markov Chain with n states has the following form:

$$
\begin{array}{ccc}
 & & \text{State at time } t+1 \\
 & & \begin{array}{ccc} 1 & j & n \end{array} \\
 & \begin{array}{c} 1 \\ i \\ n \end{array} & \begin{bmatrix} p_{11} & p_{1j} & p_{1n} \\ p_{i1} & p_{ij} & p_{in} \\ p_{n1} & p_{nj} & p_{nn} \end{bmatrix} \\
\text{State at time } t & &
\end{array}
$$

where the transition probability p_{ij} is the conditional probability that the consumer will be in state j at time $t+1$, given that he was in state i at time t.

Steady-State Probabilities

Let $\pi = (\pi_1, \pi_2, ..., \pi_n)$ be the n component row vector that is found by solving the system of equations

$$\pi = \pi P$$

These π_i are called *Steady - State Probabilities,* and can be interpreted as the proportion of time that the Markov chain is in state i when the chain is observed over a long period of time.

Stationarity

Additionally, we assume that the Markov chain is stationary. This means that the transition matrix P(t) is independent of t (probabilities do not change).

4.2.2 First Order Markov Chains

The aspect of "first order" of a Markov chain means that p_{ij} is independent of the individual's state at times $t-1$, $t-2$, . . . Thus, in this case Markov models assume that only the last brand chosen affects the current purchase.

Specifically, there are two properties that characterize a stationary first-order Markov process. Let Y_t denote the brand chosen on the tth purchase occasion and N denote the number of brands. Then the stationary Markov process satisfies the following conditions:

$$p(Y_t = k|Y_{t-1}, Y_{t-2}, ..., Y_0) = p(Y_t = k|Y_{t-1}) \qquad (one - period\ memory)$$
$$p(Y_t = k|Y_{t-1}) = p(Y_1 = k|Y_0) \qquad (stationary\ for\ all\ t, k)$$

Let us denote such a matrix as $P = \{p_{ij}\}$,where p_{ij} is the probability of purchasing j next, given i was last purchased. The stochastic matrix P has the following properties: $0 \leq p_{ij} \leq 1$ and $\sum_i p_{ij} = 1$.

Market Shares Given current market shares, a Markov model can be used to predict how market share changes over time. Suppose we know $\{m_{it}\}$, the market share of brand i at time t, then market shares for all brands at time t+1 can be calculated as:

$$m_{j,t+1} = \sum_{i=1}^{n} p_{ij} m_{it} \ , \ j = 1, 2, ..., n.$$

So, if we apply a Markov model in a brand switching matrix, we obtain some useful information, such as:

1. Using the transition matrix, we can perform forecasting for market shares, and

2. These models show how the effect of a change in market structure can be evaluated.

Remark 1 *As with other stochastic models, several critical assumptions are used, including purchase timing (one purchase per time period), homogeneity and stationarity.*

Incorporating Explanatory Variables

Givon and Horsky (1990) provide a two stage Markov model that incorporates explanatory variables. They define the Markovian transition matrix corresponding to individuals brand switching behaviour among brand A and B (brand B may represent all non-A brands), and then non-stationarity in the transition probabilities is allowed as:

$$\begin{array}{c c} & \begin{array}{cc} A_t & B_t \end{array} \\ \begin{array}{c} A_{t-1} \\ B_{t-1} \end{array} & \begin{pmatrix} P^t_{AA} & P^t_{AB} \\ P^t_{BA} & P^t_{BB} \end{pmatrix} \end{array}$$

where P^t_{XY} is the probability of buying brand Y in period t given that brand X was purchased in the previous period $t-1$.

According to Morrison (1966) and Givon and Horsky (1978) the above transition matrix can be written as:

$$\begin{array}{c c} & \begin{array}{cc} A_t & B_t \end{array} \\ \begin{array}{c} A_{t-1} \\ B_{t-1} \end{array} & \begin{pmatrix} \alpha_t + \beta & 1 - \alpha_t - \beta \\ \alpha_t & 1 - \alpha_t \end{pmatrix} \end{array}$$

where β is the purchase feedback due to experience with the brand and is a non feedback parameter.

Since, the 'first order' consumer only remembers the last experience with the brand and α_t the brand's quality is likely to remain unchanged over time, β will be considered as constant. The non-feedback effect, on the other hand, will be assumed to be impacted by marketing mix activities and therefore α_t is considered to be nonstationary.

The market share of brand A at the time of purchase event t, based on the above matrix, is denoted m_t and has the following form:

$$m_t = m_{t-1}(a_t + \beta) + (1 - m_{t-1})\, a_t = a_t + \beta m_{t-1}.$$

At this point, we specify the nonfeedback effect, a_t, to be:

$$a_t = \alpha + \gamma G_t + \delta R_t + \varepsilon_t$$

Hence, each consumer in the market is assumed to follow the process shown, with β, the "brand feedback" effect constant across customers, and the effect of controllable variables α_t, is going to model as follows:

$$\alpha_t = \alpha_0 + \gamma(a_t + \lambda a_{t-1} + \lambda^2 a_{t-2} + ...+) + \delta R_t + \varepsilon_t$$

where $\alpha_0, \gamma, \lambda, \delta$ are constants

a_t is the advertising share of brand A in period t,

G_t is its goodwill

R_t is the (relative) price of brand A at time t

ε_t is the random error term (random events which impact α_t).

Thus, the relative information contained in ads is assumed to be perfectly remembered, at the point of purchase, if it occurred in the past last purchase period and to be forgotten in a pattern of geometric decay, if it occurred in prior periods. The fraction remembered from one period to the next is λ. Unlike advertising share which contains information which may be remembered, relative price is considered to embody no such lasting information, and only the current relative price is assumed to affect the brand's market share.

Combining the previous equations we have:

$$m_t = \alpha_0 + \gamma(a_t + \lambda a_{t-1} + \lambda^2 a_{t-2} + ...+) + \delta R_t + \beta m_{t-1} + \varepsilon_t.$$

Transforming the last equation by first considering $m_t - \lambda m_{t-1}$ and then $m_t - \beta m_{t-1}$ yields

$$m_t = \alpha_0(1-\lambda) + (\beta + \lambda)m_{t-1} - \beta m_{t-2} + \gamma X_t + \delta R_t + +\delta\lambda R_{t-1} + \mu_t$$

where $\mu_t = \varepsilon_t - \lambda\varepsilon_{t-1}$.

Remark 2 *This model does not deal with population heterogeneity or with purchase timing.*

Remark 3 *More complete stochastic models, incorporating timing, choice, heterogeneity, marketing variables, feedback and so on, face two problems:*

1. *The inclusion and linking of these additional phenomena leads to models that are analytically complex and difficult to communicate.*

2. *These models require more data and more subtle estimation procedure than are often available for practical applications.*

4.3 Linear Learning Models (LLM)

Linear Learning models provides one possible way to accommodate adaptive behaviour into stochastic models of consumer brand choice. Kuehn (1958) developed linear learning models on switching patterns for frozen orange juice. The basic idea which led him to the development of such a model is that consumers are affected by feedback from past brand choices. Hence, the act of purchasing a particular brand is assumed to affect the probability that this brand will be selected the next time (purchase event feedback).

The linear learning models offers a set of hypotheses about the way in which a purchase event feeds back on the post-purchase probabilities. The most important assumption is that the post-purchase probability is always a linear function of the pre-purchase probability. This linear function has the following form:

$$p_t = \alpha + \beta p_{t-1} \tag{4.1}$$

for all the values of p_t and all t, and $\alpha, \beta \in [0, 1]$. The right -side part $(\alpha + \beta p_{t-1})$ is called feedback operator. Additional assumptions contained in linear learning models are the following:

- The model assumes quasi-stationarity in the sense that the parameters of the change operators do not change over short periods in time.
- The model assumes that all households exhibit adaptive behaviour that can be described by feedback operators with, at least approximately, the same parameters. Hence, the parameters in the learning model are assumed to be the same for all households.

4.3.1 A brief description of LLM (Linear Learning Model)

Assume, that we have two brands in a market. Given this condition, the purchase history of each household is represented by a series of zeros and ones. So, we have:

$$Y_t = \begin{cases} 1, & \text{if the brand of interest is purchased on occassion } t \\ 0, & \text{otherwise} \end{cases} \tag{4.2}$$

Let $p_t^{(1)}$ is the probability of purchasing the brand on occasion t. Then $p_t^{(0)} = 1- p_t^{(1)}$, and the basic equations of simple linear learning model are a pair of operators called acceptance and rejection operator, respectively. These equations have the following form:

$$p_t^{(1)} = \alpha + \beta + \lambda p_{t-1}, \quad \text{if brand 1 is purchased at } t \text{ (acceptance operator)} \tag{4.3}$$

$$p_t^{(0)} = \alpha + \lambda p_{t-1}, \quad \text{if brand 0 is purchased at } t \text{ (rejection opperartor)} \tag{4.4}$$

where λ presents the slope of the model in every case.

At this point, it is important to remember that all households are assumed to have the same values for the parameters α, β and λ, (according to previous assumptions), even though this process leads to different values of p_t for different households. Because of $0 \leq p_t \leq \lambda$ for all t, this implies that $\alpha + \beta + \lambda$ must also be constrained to lie in the interval $[0, 1]$.

According to Hermiter and Howard (1964) learning has a non-negative influence. This means that it is assumed that the customer's probability of purchasing brand 1 in period t is greater if brand 1 has been purchased the previous time than if some other brand has been purchased, $p_t^{(11)} > p_t^{(01)}$. Also, the parameters of the model (α, β, λ) are heavily restricted by the standard probabilistic constrains:

$$0 \leq p_t^{(01)} \leq 1$$
$$\sum p_t^{(01)} = 1$$

Hence, the above assumptions lead to the following set of conditions which have to be satisfied:

$$\lambda_j = \mu_i = \lambda \qquad \text{for all } i, j = 1, 2, 3..., n$$
$$\beta_i + \sum \alpha_j = 1 - \lambda \qquad \text{for all } i, j = 1, 2, 3..., n$$
$$-1 \leq \max_{1 \leq j \leq n}(-\alpha_j) \leq \lambda \leq \min_{1 \leq j \leq n}(1 - \alpha_j - \beta_j) \leq 1$$

This means that the equality of the slope coefficients λ in the purchase and rejection operators is an assumption that may be relaxed in the two-brand case. If more than two brands are considered, the slopes must be equal in order to ensure that the brand-choice probabilities will always lie in $[0, 1]$. The above are illustrated in the following figure:

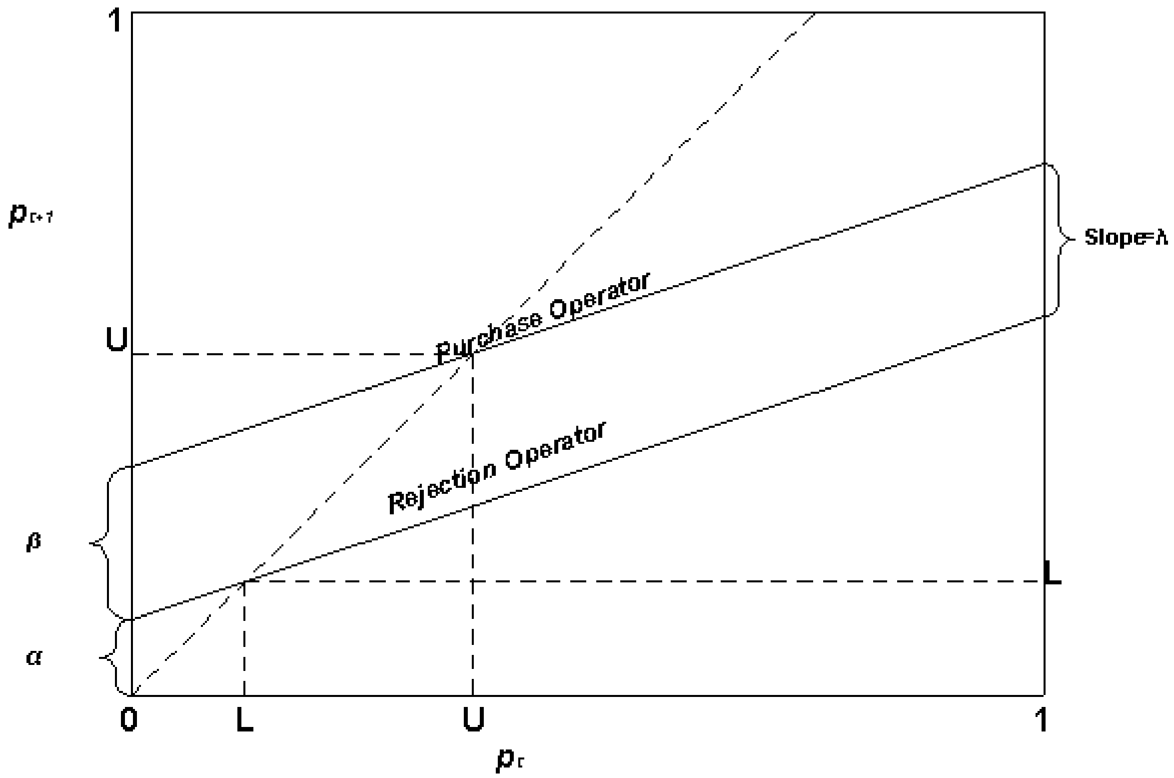

Figure 3: Graphical representation of the Linear Learning Model

The horizontal coordinate presents the probability of choosing brand 1 in period t and the vertical coordinate presents the probability of choosing brand 1 in period $t + 1$. There is a positively slope 45^0 line that is used as a norm. The two positively sloped lines present the acceptance and rejection operators. The two points U and L, give us the information as far as the purchase is concerned. In particular, the points show us where the acceptance and rejection operators cross the 45^0 line, respectively.

It easy to prove, that if a consumer continues to buy brand 1, the probability of buying brand 1 approaches 0.87 as a limit. This upper limit, given the intersection of the purchase - operator and the 45^0 line, presents a phenomenon known as **incomplete habit formation.** On the other hand, if a consumer does not buy brand 1 for a long time, the probability of buying this brand falls continuously but it never becomes zero. This is the phenomenon of **incomplete habit extinction.**

Chapter 5

Latent Class Models applied to buying behaviour

5.1 Restrictive and Unrestrictive Latent Class Model

Goodman (1974 a) dealt with the relationships among m polytomous variables, i.e. with the analysis of an m-way contingency table. These m variables were manifest variables in that, for each observed individual in a sample, his class with respect to each of the m variables was observed. In the same paper, Goodman considered polytomous variables that are latent in the sense that an individual's class with respect to these variables was not observed. The classes of a latent variable was called **latent classes**.

5.1.1 The Latent Class Model Unrestricted

Consider a 4-way contingency table which cross-classifies a sample of n individuals with respect to four manifest polytomous variables A, B, C and D. Suppose also that the manifest polytomous variables A, B, C and D consist of I, J, K and L classes, respectively. Let π_{ijkl} denote the individuals probability of being at level (i, j, k, l), with respect to the joint variable (A, B, C, D) $\{i = 1, ..., I; j = 1, ..., J; k = 1, ..., K; l = 1, ..., L\}$. Suppose that there is a latent polytomous variable X, consisting of T classes, that can explain the relationships among the manifest variables (A, B, C, D). This means that π_{ijkl} can be expressed as:

$$\pi_{ijkl} = \sum_{t=1}^{T} \pi_{ijklt}^{ABCDX} \tag{5.1}$$

where

$$\pi_{ijklt}^{ABCDX} = \pi_t^X \pi_{it}^{\bar{A}X} \pi_{jt}^{\bar{B}X} \pi_{kt}^{\bar{C}X} \pi_{lt}^{\bar{D}X} \tag{5.2}$$

denotes the individuals probability of being at level (i, j, k, l, t) with respect to the joint variable (A, B, C, D, X). Here π_t^X denotes the individuals probability of being at level t with respect to variable X and $\pi_{it}^{\bar{A}X}$ denotes the probability that an individual will be at level i with respect to variable A, given that he is at level t with respect to variable X, and finally $\pi_{jt}^{\bar{B}X}$, $\pi_{kt}^{\bar{C}X}$ and $\pi_{lt}^{\bar{D}X}$ denote similar conditional probabilities. Formula (5.1) states that the individuals can be classified into T mutually exclusive and exhaustive latent classes whereas formula (5.2) states that within the t*th* latent classes the manifest variables (A,B,C,D) are mutually independent (t=1,...,T). In order for equations (5.1) and (5.2) to be verified, the following constrains must hold:

$$\sum_{t=1}^{T} \pi_t^X = 1,\ \sum_{i=1}^{I} \pi_{it}^{\bar{A}X} = 1,\ \sum_{j=1}^{J} \pi_{jt}^{\bar{B}X} = 1,\ \sum_{k=1}^{K} \pi_{kt}^{\bar{C}X} = 1,\ \sum_{l=1}^{L} \pi_{lt}^{\bar{D}X} = 1 \tag{5.3}$$

$$\pi_t^X = \sum_{i,j,k,l,} \pi_{ijklt}^{ABCDX} \tag{5.4}$$

$$\pi_t^X \pi_{it}^{\bar{A}X} = \sum_{j,k,l,} \pi_{ijklt}^{ABCDX} \tag{5.5}$$

It is straightforward to see that by substituting $\pi_{it}^{\bar{A}X}$ by $\pi_{jt}^{\bar{B}X}$,$\pi_{kt}^{\bar{C}X}$, $\pi_{lt}^{\bar{D}X}$, formulas analogous to (5.5) can be obtained. In addition:

$$\pi_{ijklt}^{ABCD\bar{X}} = \pi_{ijklt}^{ABCDX} | \pi_{ijkl}^{ABCD} \tag{5.6}$$

where $\pi_{ijklt}^{ABCD\bar{X}}$ denotes the conditional probability that an individual is in latent class t, given that he was at level (i, j, k, l) with respect to the joint variable (A, B, C, D). Using (5.6) we can rewrite (5.4) and (5.5) as:

$$\pi_t^X = \sum_{i,j,k,l,} \pi_{ijkl}\pi_{ijklt}^{ABCD\bar{X}} \tag{5.7}$$

$$\pi_t^X \pi_{it}^{\bar{A}X} = \left(\sum_{j,k,l,} \pi_{ijkl}\pi_{ijklt}^{ABCDX}\right) | \pi_t^X \tag{5.8}$$

In the same way, if $\pi_{it}^{\bar{A}X}$ is substituted by $\pi_{jt}^{\bar{B}X}$,$\pi_{kt}^{\bar{C}X}$, $\pi_{lt}^{\bar{D}X}$, then equations analogous to (5.8) are obtained.

The maximum likelihood estimates of the corresponding parameters in the latent-class model for all the above are:

$$\hat{\pi}_{ijkl} = \sum_{t=1}^{T} \hat{\pi}_{ijklt}^{ABCDX} \tag{5.9}$$

where

$$\hat{\pi}_{ijklt}^{ABCDX} = \hat{\pi}_t^X \hat{\pi}_{it}^{\bar{A}X} \hat{\pi}_{jt}^{\bar{B}X} \hat{\pi}_{kt}^{\bar{C}X} \hat{\pi}_{lt}^{\bar{D}X} \tag{5.10}$$

We can also find that,

$$\hat{\pi}_{ijklt}^{ABCD\bar{X}} = \hat{\pi}_{ijklt}^{ABCDX} \; | \; \hat{\pi}_{ijkl}^{ABCD} \tag{5.11}$$

If p_{ijkl} denotes the observed proportion of individuals at level (i, j, k, l) with respect to the joint variables (A,B,C,D), standard methods prove that the maximum likelihood estimates satisfy the following system of equations:

$$\hat{\pi}_t^X = \sum_{i,j,k,l} p_{ijkl} \, \hat{\pi}_{ijklt}^{ABCD\bar{X}}$$

$$\hat{\pi}_{it}^{\bar{A}X} = \sum_{j,k,l} p_{ijkl} \, \hat{\pi}_{ijklt}^{ABCD\bar{X}} \; | \; \hat{\pi}_t^X$$

$$\hat{\pi}_{jt}^{\bar{B}X} = \sum_{i,k,l} p_{ijkl}\, \hat{\pi}_{ijklt}^{ABCD\bar{X}} \,|\, \hat{\pi}_t^{X}$$

$$\hat{\pi}_{kt}^{\bar{C}X} = \sum_{i,j,l} p_{ijkl}\, \hat{\pi}_{ijklt}^{ABCD\bar{X}} \,|\, \hat{\pi}_t^{X}$$

$$\hat{\pi}_{lt}^{\bar{D}X} = \sum_{i,j,k} p_{ijkl}\, \hat{\pi}_{ijklt}^{ABCD\bar{X}} \,|\, \hat{\pi}_t^{X}$$

Let π denote the vector of parameters $\left(\pi_t^{\bar{X}}, \pi_{it}^{\bar{A}X}, \pi_{jt}^{\bar{B}X}, \pi_{kt}^{\bar{C}X}, \pi_{lt}^{\bar{D}X}\right)$ in the latent class model, and $\hat{\pi}$ denote the corresponding maximum likelihood estimate of the vector. In order to calculate $\hat{\pi}$, the following iterative procedures are applied:

1. Start with an initial trial value for $\hat{\pi}$, $[\pi(0) = \{\pi_t^{\bar{X}}(0), \pi_{it}^{\bar{A}X}(0), \pi_{jt}^{\bar{B}X}(0), \pi_{kt}^{\bar{C}X}(0), \pi_{lt}^{\bar{D}X}(0)\}]$

2. Use the formula $\hat{\pi}_{ijklt}^{ABCDX} = \hat{\pi}_t^{X}\hat{\pi}_{it}^{\bar{A}X}\hat{\pi}_{jt}^{\bar{B}X}\hat{\pi}_{kt}^{\bar{C}X}\hat{\pi}_{lt}^{\bar{D}X}$ to obtain a trial value for $\hat{\pi}_{ijklt}^{ABCDX}$ replacing the terms on the right-hand side of this equation by the corresponding components in $\pi(0)$.

3. Then, use equation (5.9) to obtain a trial value for $\hat{\pi}_{ijkl}$ and equation (5.11) in order to obtain a trial value for $\hat{\pi}_{ijklt}^{ABCD\bar{X}}$. Similarly, use the rest of the equations to obtain new trial values for $\hat{\pi}_t^{X}$, $\hat{\pi}_{it}^{\bar{A}X}$, $\hat{\pi}_{jt}^{\bar{B}X}$, $\hat{\pi}_{kt}^{\bar{C}X}$ and $\hat{\pi}_{lt}^{\bar{D}X}$. Thus a new trial value for the vector $\hat{\pi}$ is obtained.

4. Repeat the procedure starting with the new trial value to obtain the next trial value for $\hat{\pi}$.

5. The procedure will converge to a solution for the system of equations and to a corresponding likelihood.

Remark 4 *In this iterative procedure a latent class is deleted if the corresponding estimates tend to zero.*

We compare the solutions obtained using the iterative procedure and see which one minimizes the chi-squared (X^2) statistic

$$X^2 = 2\sum_{ijkl} f_{ijkl}\log(f_{ijkl}|\ \hat{F}_{ijkl})$$

$$f_{ijkl} = np_{ijkl}\ ,\ \ \hat{F}_{ijkl} = n\ \hat{\pi}_{ijkl}$$

where $\hat{\pi}_{ijkl}$ is obtained from (5.9). The maximum likelihood estimate for $\hat{\pi}$ minimizes the equation of X^2.

Remark 5 *If* $\hat{\pi}$ *is uniquely determined by* $\hat{\pi}_{ijkl}$, *then it is said to be* identifiable.

Remark 6 *If* $\hat{\pi}$ *is uniquely determined by* $\hat{\pi}_{ijkl}$, *within some neighborhood of* π, *then it is said to be* locally identifiable.

5.1.2 Some Restricted Latent Structures

The estimation procedure presented in the previous section can be modified in a way to accommodate models that possess what is called T-class restricted latent structure.

1. Models in which a condition of the following type is imposed upon the parameters:

$$\pi_{i1}^{\bar{A}X} = \pi_{i2}^{\bar{A}X}\quad (i = 1, ..., I)$$

2. More generally, models in which the T latent classes can be partitioned into α mutually exclusive and exhaustive subsets $\mathcal{T}_1^A, ..., \mathcal{T}_\alpha^A$ where $\alpha \leq \mathcal{T}$, and /or into β mutually exclusive and exhaustive subsets $\mathcal{T}_1^B, ..., \mathcal{T}_\alpha^B$, where $\beta \leq \mathcal{T}$, such that

$$\pi_{it}^{\bar{A}X} = \pi_{it'}^{\bar{A}X}\ \ ,\ \left(t, t' \in \mathcal{T}_\alpha^A\right)\ \text{and}\ \pi_{it}^{\bar{B}X} = \pi_{it'}^{\bar{B}X}\ \ ,\ \left(t, t' \in \mathcal{T}_b^B\right) \tag{5.12}$$

where $\alpha = 1, ..., \alpha$; $b = 1, ..., \beta$; $i = 1, ..., I$; $j = 1, ..., J$.

3. Models in which, in addition to condition (5.12) the following kind of condition is satisfied for certain specified pairs of subscripts, say (α, α^*) and/or (b, b^*):

$$\pi_{it}^{\bar{A}X} = \pi_{jt'}^{\bar{B}X}, \ (t \in \mathcal{T}_{\alpha}^{B}, t' \in \mathcal{T}_{b}^{B}),$$

$$\pi_{it}^{\bar{A}X} = \pi_{i^*t^*}^{\bar{A}X} \ \left(t \in \mathcal{T}_{\alpha}^{A}, t^* \in \mathcal{T}_{\alpha^*}^{A}\right); \pi_{jt}^{\bar{B}X} = \pi_{j^*t^*}^{\bar{B}X} \ \left(t \in \mathcal{T}_{b}^{B}, t^* \in \mathcal{T}_{b^*}^{B}\right)$$

$i = 1, ..., I$, where there is a one-to-one correspondence between i and j, between i and i* , and between j and j*.

4. Finally, models which were expressed in terms pertaining to variables A and B, are extended to other subsets of the m manifest variables in the m-way contingency table.

To determine whether the estimated parameters in a restricted latent structure are locally identifiable, we can use a modified form. Before we refer to this modified form it is useful to give a sufficient condition for local identifiability.

Goodman (1974a) proposed a method to study whether π is locally identifiable and/or whether $\overset{a}{\pi}$ is locally identifiable. If we combine equations (5.1) and (5.2) our model becomes:

$$\pi_{ijkl} = \sum_{t=1}^{T} \pi_t^X \pi_{it}^{\bar{A}X} \pi_{jt}^{\bar{B}X} \pi_{kt}^{\bar{C}X} \pi_{lt}^{\bar{D}X} \tag{5.13}$$

This formula describes a set of $IJKL$ functions that transform the parameters

$$\left(\pi_t^X, \pi_{it}^{\bar{A}X}, \pi_{jt}^{\bar{B}X}, \pi_{kt}^{\bar{C}X}, \pi_{lt}^{\bar{D}X}\right)$$

into the π_{ijkl} . Since the restrictions (5.3) hold, we need only consider $T - 1$ of the π_t^X only $i - 1$ of the $\pi_{it}^{\bar{A}X}$ etc.. Thus, we need only consider

$$T - 1 + (I + J + K + L - 4)T = (I + J + K + L - 3)T - 1$$

parameters. Similarly , since $\sum_{ijkl} \pi_{ijkl} = 1$ we need only consider $IJKL - 1$ of the π_{ijkl} for $(i, j, k, l) \neq (IJKL)$. When $IJKL < (I + J + K + L - 3)T$ the number of parameters in the basic set exceeds the corresponding number of π_{ijkl} , and so the parameters will not be identifiable in this case. If the number of parameters in the basic set does not exceed the corresponding number of π_{ijkl} then, for each π_{ijkl} in the basic set, we calculate the derivative of the function π_{ijkl} (equation 5.13) with respect to the parameters in the basic set. Thus, we obtain a matrix consisting of $IJKL - 1$ rows and $(I+J+K+L-3)T-1$ columns. For example, in the column pertaining to the derivative with respect to π_t^X ,

$$\frac{\partial \pi_{ijkl}}{\partial \pi_t^X} = \pi_{it}^{\bar{A}X} \pi_{jt}^{\bar{B}X} \pi_{kt}^{\bar{C}X} \pi_{lt}^{\bar{D}X} - \pi_{iT}^{\bar{A}X} \pi_{jT}^{\bar{B}X} \pi_{kT}^{\bar{C}X} \pi_{lT}^{\bar{D}X}$$

for t = 1, ... , T-1; in the column pertaining to the derivative with respect to $\pi_{it}^{\bar{A}X}$,

$$\frac{\partial \pi_{ijkl}}{\partial \pi_{st}^{\bar{A}X}} = \begin{cases} \pi_t^X \pi_{jt}^{\bar{B}X} \pi_{kt}^{\bar{C}X} \pi_{lt}^{\bar{D}X} & (i = s) \\ -\pi_t^X \pi_{jt}^{\bar{B}X} \pi_{kt}^{\bar{C}X} \pi_{lt}^{\bar{D}X} & (i = i) \quad for\ s = 1, ..., i-1 \\ 0 & otherwise \end{cases}$$

Replacing the π's by the corresponding $\hat{\pi}$'s, we have the maximum likelihood estimates of the parameters in the model. These estimates are locally identifiable.

We now describe some simple restrictions that would make the parameters in the latent structure unidentifiable. First, we consider the case where the T-class model is such that

$$\pi_{i1}^{\bar{A}X} = \pi_{i2}^{\bar{A}X}, \ \pi_{j1}^{\bar{B}X} = \pi_{j2}^{\bar{B}X}, \ \pi_{k1}^{\bar{C}X} = \pi_{k2}^{\bar{C}X}, \ \pi_{l1}^{\bar{D}X} = \pi_{l2}^{\bar{D}X}$$

In this case, the formula which describes the set of IJKL functions and transforms the parameters $\pi_t^X, \pi_{it}^{\bar{A}X}, \pi_{jt}^{\bar{B}X}, \pi_{kt}^{\bar{C}X}, \pi_{lt}^{\bar{D}X}$ into π_{ijkl} becomes:

$$\pi_{ijkl} = \sum_{t=2}^{T} \Theta_t^X \pi_{it}^{\bar{A}X} \pi_{jt}^{\bar{B}X} \pi_{kt}^{\bar{C}X} \pi_{lt}^{\bar{D}X}$$

where $\Theta_t^X = \begin{cases} \pi_1^X + \pi_2^X \ , (t=2) \\ \pi_t^X, \ (t = 3, 4, ..., T) \end{cases}$. Thus, we can collapse latent classes (1) and (2) to obtain an equivalent latent structure having T-1 classes rather that T classes.

A second case of restrictions is the one when the T-class model is such that

$$\pi_{j1}^{\bar{B}X} = \pi_{j2}^{\bar{B}X}, \pi_{k1}^{\bar{C}X} = \pi_{k2}^{\bar{C}X}, \pi_{l1}^{\bar{D}X} = \pi_{l2}^{\bar{D}X}$$

then we have

$$\pi_{ijkl} = \sum_{t=2}^{T} \Theta_t^X \Theta_{it}^{\bar{A}X} \pi_{jt}^{\bar{B}X} \pi_{kt}^{\bar{C}X} \pi_{lt}^{\bar{D}X}$$

where Θ_t^X is defined as before and $\Theta_{it}^{\bar{A}X} = \begin{cases} \left(\pi_1^X \pi_{i1}^{\bar{A}X} + \pi_2^X \pi_{i2}^{\bar{A}X} \right) / \Theta_2^X, \ (t=2) \\ \pi_{it}^{\bar{A}X}, \ (t = 3, 4, ..., T) \end{cases}$. Thus, we can also collapse latent classes (1) and (2), and the parameters $\pi_1^X, \pi_2^X, \pi_{i1}^{\bar{A}X}, \pi_{i2}^{\bar{A}X}$, will not be identifiable unless additional restrictions are imposed upon them.

A last case of restrictions is the one when the T-class model is such that

$$\pi_{k1}^{\bar{C}X} = \pi_{k2}^{\bar{C}X}, \pi_{l1}^{\bar{D}X} = \pi_{l2}^{\bar{D}X}.$$

Then we have

$$\pi_{ijkl} = \sum_{t=2}^{T} \Theta_t^X \Theta_{it}^{\bar{A}X} \Theta_{jt}^{\bar{B}X} \pi_{kt}^{\bar{C}X} \pi_{lt}^{\bar{D}X}$$

where Θ's are equal to the corresponding $\pi's$ for t = 3, ...,T, and where

$$\sum_{t=1}^{2} \Theta_t^X \Theta_{it}^{\bar{A}X} \Theta_{jt}^{\bar{B}X} = \sum_{t=1}^{2} \pi_t^X \pi_{it}^{\bar{A}X} \pi_{jt}^{\bar{B}X}$$

for i = 1, ...,I; j = 1, ... ,J. When T>2 the above equation imposes IJ restrictions on the Θ's. Because of (5.3) the number of Θ's that we need to consider is $2(I + J - 2) + 2 = 2(I + J - 1)$.

5.1.3 The EM algorithm for Latent Class Models

The estimation problem of the latent class models is solved via two methods. The first is the Newton-Raphson approach as used by Haberman (1988). Specifically, in this approach, latent class analysis is conceived as a log-linear model that describes the relation between some manifest variables and a latent variable. This method requires a relative small number of iterations, allows for the implementation of various types of constraint on the parameters, and finds asymptotic covariances for the estimates as a by-product.

In the second method, the estimation problem of latent class models can be handled as an estimation problem with missing data (namely the observations on the latent variable are missing). Specifically, we derive maximum likelihood estimates of the parameters using the EM algorithm (Dempster,Laird and Rubin 1977). In this section we study the estimation of parameters of LCM via the EM algorithm. Further, we present a review of the estimation of the unconstrained Latent Class Model using the EM algorithm and try to find out how it works when there are constraints on the parameters.

The EM Algorithm for the Unconstrained Latent Class Models

As we can see, the EM algorithm maximizes the likelihood function with a distinction between the observed and missing data. The observed data are the scores of the manifest categorical variables, and the missing data are the scores of the latent variable. Before continuing, it is important to note, that latent class analysis assumes a categorical latent variable with some categories, such that given a level x of this latent variable the manifest variables are independent.

The estimation method of EM algorithm consists of two steps:

1. The E (expectation) step: In this step we compute the expectations of the unobserved complete data conditional on the observed incomplete data matrix and the current parameters estimates.

2. The M (maximization) step: In this step, we maximize the expected log-likelihood of the incomplete data matrix as a function of the unknown model parameters.

Assuming a multinomial distribution, the Kernel likelihood of the complete data can be written as:

$$L^c = \prod_{s,k} (\pi_{s,k})^{n_{s,k}}$$

where:

$\pi_{s,k}$ denotes the unobserved probabilities of falling simultaneously in the categories denoted by vector s (vector of the length of the manifest variables) and the latent class k , and

$n_{s,k}$ denotes the number of subjects in the sample who have pattern s and fall in class k.

E-step: In this step the conditional expectation of $n_{s,k}$ has to be formulated. According to the Bayes' theorem we have that

$$\pi_{k|s} = \pi_k \pi_{s|k} / \sum_t \pi_k \pi_{s|t}, \text{ where } \pi_{s|t} = \prod_u \pi_{u,s(u)|k}$$

$\pi_{v,s(v)|k}$ is the probability of category $s(v)$ for variable v conditional on class k.

Consequently, the conditional expectation of $n_{s,k}$ is

$$n_s \pi_{k|s} = \frac{n_s \pi_k \pi_{s|k}}{\left(\sum_t \pi_t \pi_{s|t}\right)}$$

M-step: In this step the complete -data log likelihood is maximized with respect to the unknown model parameters π_k and $\pi_{u,s(u)}$, with $n_{s,k}$ replaced by the conditional expectation (from E-step), denoted as $n^+_{s,k}$. So, we maximize the following:

$$\log(L^c) \;=\; \sum_{s,k} n^+_{s,k} \log(\pi_{s,k}) = \sum_{s,k} n^+_{s,k} \log\left(\pi_k \prod_v^V \pi_{v,s(v)|k}\right) =$$

$$\sum_{s,k} n_{s,k}^{+} \log\left(\pi_k\right) + \sum_{s,k} n_{s,k}^{+} \sum_{v}^{V} \log\left(\pi_{v,s(v)|k}\right) \tag{5.14}$$

Optimization of $\log(L^c)$ is achieved using Lagrange multipliers.

Estimation of π_k: In this case the Lagrangian can be written as

$$f_1 = \sum_{s,k} n_{s,k}^{+} \log\left(\pi_k\right) - \alpha \left(\sum_{k} \pi_k - 1 \right) \tag{5.15}$$

If we want to maximize equation (5.14) over π_k, we should maximize the equation (5.15). Then by taking derivatives, putting them equal to zero, and solving for α, we get

$$\frac{\partial f1}{\partial \pi_k} = \sum_{s} \frac{n_{s,k}^{+}}{\pi_k} - \alpha = 0 \Leftrightarrow \alpha = N$$

Thus, we have that the estimation of π_k is:

$$\pi_k = \sum_{s} \frac{n_{s,k}^{+}}{\alpha} = \sum_{s} \frac{n_{s,k}^{+}}{N}.$$

Estimation of $\pi_{v,s(v)|k}$: It is easy to verify the following:

$$\sum_{s,k} n_{s,k}^{+} \sum_{v}^{V} \log\left(\pi_{v,s(v)|k}\right) = \sum_{v}^{V} \sum_{i}^{I_v} \sum_{k}^{K} n_{v,i,k}^{+} \log \pi_{v,i|k}$$

where

$\pi_{v,i|k}$ is the unknown conditional probability, for which the side condition $\sum_i \pi_{v,i|k} = 1$, and

$n_{v,i,k}^{+}$ is defined as the number of subjects in category i of variable v and simultaneously, in class k (updated value of $n_{v,i,k}$ from the E-step).

So, the function to be optimized becomes

$$f_2\left(\pi_{v,i|k}\right) = \sum_{v}^{V} \sum_{i}^{I_v} \sum_{k}^{K} n_{v,i,k}^{+} \log \pi_{v,i|k} - \sum_{v}^{V} \sum_{k}^{K} \beta_{v,k} \left(\sum_{i}^{I_v} \pi_{v,i|k} - 1 \right)$$

where the $\beta_{v,k}$ is a Lagrange multiplier.

Then by taking derivatives, putting them equal to zero, and solving for β, we obtain

$$\frac{\partial f_2\left(\pi_{v,i|k}\right)}{\partial \pi_{v,i|k}} = \frac{n^+_{v,i,k}}{\pi_{v,i|k}} - \beta_{v,k} = 0 \Leftrightarrow \beta_{v,k} = \sum_i n^+_{v,i,k}$$

Thus, we have that the estimation of $\pi_{v,i|k}$ is :

$$\pi^+_{v,i|k} = \frac{n^+_{v,i,k}}{\sum_i n^+_{v,i,k}}.$$

The EM Algorithm in Constrained Latent Class Analysis

In this case, we estimate the probabilities when some of them are constrained. Two types of constraints are considered:

- Fixed value constraints for one or more conditional probabilities $\pi_{v,i|k}$ (these constraints are used to assess whether an estimate of a parameter is significantly different from some value of theoretical interest)
- equality constraints for one or more sets on conditional probabilities (these constraints are used to assess whether the estimate of two or more parameters are different).

According to Mooijaart and Heijden (1992) the function that is going to be maximized, for estimation of the conditional probability $\pi_{v,i|k}$, is :

$$f\left(\pi_{v,i|k}\right) = \sum_{v,i,k}^{V,I_v,k} n_{v,i,k} \log(\pi_{v,i|k}) \tag{5.16}$$

This function is to be maximized over the unknown parameters π_{E_l} and the unknown free parameter $\pi_{v,i|k}$.

Before going to the estimation procedure, it is important to make some useful notation. Thus, a set of elements $\pi_{v,i|k}$ is constrained to

1. fixed values (F)

2. free values (G), and

3. equality values (L), denoted by E_l ($l = 1, 2, ..., L$), { E_l consists of elements $\pi_{v,i|k}$ that are constrained to be equal}.The union of the sets E_l is denoted as E, and the elements $\pi_{v,i|k}$ in set E_l are equal to π_{E_l}.

Hence, the equation (5.16) can be rewritten as

$$f\left(\pi_{E_l}, \pi_{v,i|k}\right) = \sum_{l}^{L} n_{E_l} \log \pi_{E_l} + \sum_{\substack{v,i,k, \\ v,i,k \in G}} n_{v,i,k} \log \pi_{v,i|k} + \sum_{\substack{v,i,k, \\ v,i,k \in F}} n_{v,i,k} \log \pi_{v,i|k} \qquad (5.17)$$

We are going to maximize this equation (5.17) using Lagrange multipliers. There are VK different variables and latent classes (indexed by v and k). The restrictions that hold for each variable - latent class combination are:

$$\left(\sum_{l}^{L} d_{l,v,k} \pi_{E_l} + \sum_{\substack{i \\ v,i,k \in G}} \pi_{v,i|k}\right) = \left(1 - \sum_{\substack{i \\ v,i,k \in F}} \pi_{v,i|k}\right) \equiv c_{v,k} \qquad (5.18)$$

The constrained function, including the Lagrange multipliers, can now be written as

$$\begin{aligned} f^*\left(\pi_{E_l}, \pi_{v,i|k}\right) &= \sum_{l}^{L} n_{E_l} \log \pi_{E_l} + \sum_{\substack{v,i,k, \\ v,i,k \in G}} n_{v,i,k} \log \pi_{v,i|k} - \\ &\quad \sum_{v,k}^{V,K} \alpha_{v,k} \left(\sum_{l}^{L} d_{l,v,k} \pi_{E_l} + \sum_{\substack{i \\ v,i,k \in G}}^{I_v} \pi_{v,i|k} - c_{v,k}\right) \end{aligned}$$

The derivatives of this function with respect to the unknown parameters are

$$\frac{\partial f^*\left(\pi_{E_l}, \pi_{v,i|k}\right)}{\partial \pi_{E_l}} = \frac{n_{E_l}}{\pi_{E_l}} - \sum_{v,k}^{V,K} d_{l,v,k} \alpha_{v,k}$$

$$\frac{\partial f^*\left(\pi_{E_l}, \pi_{v,i|k}\right)}{\partial \pi_{v,i|k}} = \frac{n_{v,i,k}}{\pi_{v,i|k}} - \alpha_{v,k}, \qquad i,k \in G.$$

Equating the above derivatives to zero and solving for the parameters, gives

$$\pi_{E_l} = \frac{n_{E_l}}{\sum_{v,k}^{V,K} d_{l,v,k}\alpha_{v,k}} \tag{5.19}$$

$$\pi_{v,i|k} = \frac{n_{v,i,k}}{\alpha_{v,k}}, \qquad i,k \in G \tag{5.20}$$

So, by solving for th VK Lagrange multipliers, $\alpha_{v,k}$ we obtain

$$\sum_{l}^{L} \frac{d_{l,v,k} n_{E_l}}{\sum d_{l,w,t}\alpha_{w,t}} + \frac{n_{G_{v,k}}}{\alpha_{v,k}} = c_{v,k} \tag{5.21}$$

which defines VK equations for VK unknown parameters $\alpha_{v,k}$. Solving for these parameters and substituting into (5.19) and (5.20) gives the solutions of the unknown model parameters. So, (5.21) is the basic formula that has to be solved.

In some cases there does not exist an explicit solution for $\alpha_{v,k}$ from (5.21), and thus there are not explicit solutions for the unknown parameters, in others, an iterative procedure has to be used for solving $\alpha_{v,k}$ from (5.21).

Case 1: No equality constraints for the probabilities in the variable-latent class combination of variable v in class k.

Case 2: Probabilities in different variable-latent class combinations are not constrained to be equal.

Case 3: Probabilities in different variable-latent class combinations are constrained to be equal, and $d_{l,vk} = d_{lc_v,k}$.

Case 4: Probabilities in different variable-latent class combinations are constrained to be equal.

5.2 Latent Class models applied to Brand Switching Data

5.2.1 Mixed Markov and Latent Markov Modeling

Markov chains have been proposed as possible models for the analysis of brand switching data many years ago, but this was not a popular approach because of the poor fit to the data. The reason for this can be found in two assumptions. First, all consumers are assumed to follow the same change process, and second it is presumed that the state space can be measured directly without error.

The first of these assumptions is relaxed in the mixed Markov model by allowing for heterogeneity in individual transition probabilities. Poulsen (1990), proposed a Mixed Markov model that follows from the partial segmentation view of buyer heterogeneity. According to him the buyer population consists of an unknown number of segments that follow separate first order, non-stationary choice processes. This model is a generalization of the latent class model which assumes zero-order, non stationary choice models. In the sequel, we present the Mixed Markov model and some restrictions of this.

Mixed Markov Model

In order to describe this model it is necessary to assume the following buyer structure:

1. The buyer market consists of a finite but unknown number of buyer segments (S).
2. A common choice set of alternative brands $cC = \{A, B, C, ...\}$ correspond to all the buyers for each purchase occasion.
3. Each buyer belongs exclusively in one segment.

4. Within a segment s at purchase occasion w, each buyer chooses among $cC = \{A, B, C, ...\}$ with probability vector

$$\delta_{is}^{(w)} = \left(\delta_{A|is}^{(w)}, \delta_{B|is}^{(w)}, \delta_{C|is}^{(w)}, ...\right). \tag{5.22}$$

5. Each buyer makes his choice independently of any other buyer.

In order to avoid computational difficulties, we consider three consecutive purchase occasions. We have a three-dimensional contingency table, that contains only the counts of buyer with choice sequence (i,j,k). In this case, we have that:

π_s = the probability a buyer belongs to segment s (segment size)

$\pi_{ijk|s}$ = the joint probability for a buyer in segment s to make the sequence (ijk) of purchases at three occasions

π_{ijk} = the marginal or overall probability of observing a buyer with that particular sequence. Thus, it is obvious that:

$$\pi_{ijk} = \sum_s \pi_s \pi_{ijk|s}$$

Additionally, if we take into consideration equation (5.22) we have that:

$$\pi_{ijk} = \sum_s \pi_s \delta_{i|s}^{(1)}, \delta_{j|is}^{(2)}, \delta_{k|js}^{(3)} \tag{5.23}$$

This is the fundamental equation of the Mixed Markov model, where π_s represents the size of segment, $\delta_{i|s}^{(1)}$is the initial choice probabilities, and $\delta_{j|is}^{(w)}$ represents the transition matrices $\{i \in cC, s = 1, 2, ...S\ , w = 2, 3, 4, ..., W\}$.

It is important to mention that some familiar models used to examine the choice behaviour are restriction of the Mixed Markov model. Some of them are the following:

- The *homogeneous Markov model* is constrained in the mixed model when the number of segments is equal to one, namely when S=1.

- The *latent class model* is also a special case of the mixed Markov model if the equality constrains of no feedback effects is imposed: $\delta_{j|is}^{(w)} = \delta_{j|s}^{(w)}$ $\forall i, j \in cC$, $s = 1, 2, ..., S$. Hence, the zero-order assumption of the latent class model can be tested within the present model hierarchy of mixed Markov models.

Latent Markov Model

In this kind of Markov model a different structure philosophy has been developed. Specifically, the basic idea is to define a set of behavioural models as states in a Markov model. These states will be stationary, zero-order models. Shifts between these states will take place according to a stationary transition matrix.

Neither the states nor the transition matrix of the Markov models are directly observable, but must be inferred from manifest choices. The entire model is called Latent Markov model. This kind of model consists of two parts:

- a probabilistic relation between the observed response (manifest variable) and the unobservable true state (latent variable), and
- the change process operating on the latent states (described by a Markov transition matrix).

The basic structural assumptions for this model are:

1. There is a finite and unknown number S of latent buyer states
2. Each s is described by a stationary probability vector over the choice set $cC = \{A, B, C, ...\}$, $\delta_s = \left(\delta_{A|s}, \delta_{B|s,}\delta_{c|s,......}\right)$. This vector captures the structural relation between the manifest choices and latent states
3. At any purchase occasion each buyer is in a particular state. Over time, buyers are allowed to change position according to a stationary transition matrix $T = (\tau_{t|s})$, denoting the conditional probability of moving to state t, given the present state s $(s,\, t = 1, 2, 3, ..., S)$.

As in the previous model, for computational reasons we consider three consecutive purchase occasions. The manifest choices are (ijk) and assume for a moment that the corresponding states (stu) of each buyer are directly observable. Thus, we can write:

$$\pi_{(ijk)(stu)} = \pi_s^{(1)} \delta_{i|s} \tau_{t|s} \delta_{j|t} \tau_{u|t} \delta_{k|u}$$

where $\pi_{(ijk)(stu)}$ is the joint probability of an individual starting s to choose i at the first purchase occasion, then move to state t and choose j at the next purchase, and finally move to state u and choose k. Because the states are not directly observable, only the marginal π_{ijk} have empirical counterparts, so the Latent Markov model has the following form:

$$\pi_{ijk} = \sum_s \sum_t \sum_u \pi_s^{(1)} \delta_{i|s} \tau_{t|s} \delta_{j|t} \tau_{u|t} \delta_{k|u}$$

where:

$\pi_s^{(1)}$ is the initial state probability, $(s = 1, 2, 3, ...S)$

$\delta_{i|s}$ is the choice probabilities, $i \in$, $(s = 1, 2, 3, ...S)$

$T = (\tau_{t|s})$ is the transition probabilities, s, $t = 1, 2, 3, ..., S$.

5.2.2 Grover and Srinivasan Model (GS Model)

Grover and Srinivasan (1987) developed an innovating method to extract market structure from brand switching data by using latent class analysis. To capture the heterogeneity in household's choice probabilities, they divide the total population of households into brand loyal segments and switching segments. Households within each switching segment are assumed to follow a zero-order switching behaviour and also are assumed to be homogeneous in their choice probabilities. Moreover, they discuss how within - segment heterogeneity in choice probabilities can be incorporated in the latent class models.

In this model the main assumption is the one that refer to stationarity. This assumption gives the possibility to approximate the stochastic brand choice behaviour of consumers as a zero order process. Thus, a consumer's brand choice probabilities at a purchase occasion are assumed to be unaffected by the previous purchase history. Also, the assumption of stationarity implies that the consumer's choice probabilities can be taken to be constant over the time period.

Let n denote the number of brands in the product class of interest. It is assumed that the heterogeneity can be captured adequately by:

1. n brand loyal segments corresponding to the n brands, $l = 1, 2, \ldots, n$
2. m switching segments, k = 1, 2, . . . , m.

A consumer that belongs to a brand loyal segment l will always purchase brand l with probability one during the time period. So, in this case, there is no within-segment heterogeneity. Considering segment l, the proportion S_{ijl} of consumers who buy brand i on one purchase occasion and brand j on another purchase occasion is given by:

$$S_{ijl} = \begin{cases} 1 & \text{if } i = j = l \\ 0 & \text{otherwise} \end{cases}$$

that is, segment l consumers will purchase only brand l on both purchase occasions.

For switching segments, let p_{ik} denote the probability of choosing brand i . Thus, p_{ik} is the market share of brand i in segment k. It follows that $p_{ik} \geq 0$ for $i = 1, 2, 3, \ldots, n$; $k = 1, 2, 3, \ldots, m$ and $\sum_{i=1}^{n} p_{ik} = 1$ for $k=1,2,\ldots,m$. Because the choice process is assumed to be zero order and stationary, the within -segment k probability of buying brand i on one purchase occasion and a different brand j on another occasion is given by: $S_{ijk} = p_{ik}p_{jk}$ for $i \neq j$. The within-segment k probability of buying the same brand i on two purchase occasions is given by: $S_{iik} = p_{ik}p_{ik} = p_{ik}^2$ for $i = 1, 2, \ldots, n$.

Let S_{ij} be the theoretical proportion of consumers who buy brand i on one occasion and different brand j on another occasion in the entire market. Then we have that :

$$S_{ij} = \sum_{l=1}^{n} V_l S_{ijl} + \sum_{k=1}^{m} W_k S_{ijk}, \text{ for i} \neq \text{j}$$

where:

V_l (the proportion of consumers in the total market who are loyal to brand l) ≥ 0 for $l = 1, 2, ..., n$

W_k (the proportion of consumers belonging to switching segment k) ≥ 0 for $k = 1, 2, ..., m$, and

$$\sum_{l=1}^{n} V_l + \sum_{k=1}^{m} W_k = 1.$$

According to the above assumptions, it follows that $S_{ijl} = 0$ since $i \neq j$. This means that there is no brand switching by loyal consumers. Moreover,the proportion in the entire market who buy the same brand i on both purchase occasions can be computed. In this case we have:

$$S_{ii} = V_i + \sum_{k=1}^{m} W_k p_{ik}^2 \text{ , for } i = 1, 2, ..., n \tag{5.24}$$

Further, the proportion of consumers in the entire market who buy brand i on one purchase occasion and brand j on another purchase occasion can be computed. Thus, we have:

$$S_{ij} = \sum_{k=1}^{m} W_k p_{ik} p_{jk} \text{ , for } i \neq j \tag{5.25}$$

Using the weights (relatives sizes of the segments) $\{V_i\}$ and $\{W_k\}$, we can get the aggregate market share:

$$MS_i = V_i + \sum_{k=1}^{m} W_k p_{ik}, \text{ for } i = 1, 2, ..., n.$$

The above equations (5.24) and (5.25) are the best approximations to real data because of assumptions such as stationarity and zero order and because of sampling fluctuations in obtaining the empirical proportion. In order to present these equations more compactly, we adopt the following notation:

$h = 1, 2, . . ., n + m$: denotes the segments

β_h = is the proportion of all consumers who belong to segment h

q_{ih} = is the probability with which segment h consumers buy brand i.

5.2.3 Jain, Bass and Chen Model (JBC Model)

Jain, Bass and Chen (1990) developed another method in order to estimate a model that explicitly incorporates within-segment heterogeneity in choice probabilities. They use a zero order household purchase behaviour model which is similar to that used by Grover and Srinivasan but it operationalizes household heterogeneity in choice probabilities in a different manner. First, they develop a stochastic model of household purchasing behaviour at the segment level under the following assumptions:

- All households follow a zero-order choice process in each segment,
- The choice probabilities of each household are constant over the purchase occasions, and
- The choice probability vector of households is Dirichlet distributed over the segment.

In a second step, they aggregate the model to the market level by summing over all the segments. It is important to note that this aggregate model cannot be estimated by using the conventional latent class model estimation procedure. They propose an iterative estimation procedure to estimate the parameters of such a model by using brand switching data.

In order to incorporate the heterogeneity into a stochastic brand choice model, it is necessary to make some distributional assumptions.

1. The probability distribution that describes the multi-brand buying behaviour, is the Dirichlet compound multinomial distribution. This distribution has the following form:

$$Multinomial\ (\bar{P_1}, \bar{P_2}, ..., \bar{P_n}) \cap Dirichlet\ (\alpha_1, \alpha_2, ..., \alpha_n)$$

where $\cap$ denotes a compounding operator. This multinomial distribution refers to the distribution of purchases of an individual household that selects from among n brand with probability vector $\bar{P}= (\bar{P}_1,\bar{P}_2, ..., \bar{P}_n)$ on each purchase occasion. The probability vector $\bar{P}$ across the heterogeneous population of households follows a Dirichlet distribution with parameters $(\alpha_1, \alpha_2, ..., \alpha_n)$.

2. For an individual household the probabilities $\bar{P}_i$, $i = 1, ..., n$ are constant on each purchase occasion and therefore the probability vector does not vary over the purchase occasion.

$$E(P_i) = P_i,\ E(\bar{P}_i, \bar{P}_j) = (1-\theta)P_iP_j$$
$$Var(\bar{P}_i) = \theta P_i(1-P_i)$$
$$Cov(\bar{P}_i, \bar{P}_j) = -\theta P_iP_j$$

where $P_i = \alpha_i/\alpha$, $\theta = 1/(1-\alpha)$, and $\alpha = \sum_{i=1}^{n} \alpha_i$.

From the assumption of Dirichlet distribution for the probability vector $\bar{P}= (\bar{P}_1,\bar{P}_2, ..., \bar{P}_n)$ we obtain the following results:

- The expression $E(\bar{P}_i, \bar{P}_j)$ indicates that the expected switching from brand i to brand j on two consecutive purchase occasions is proportional to the product of their market shares.
- The constant of proportionality $(1-\theta)$ is a number between 0 and 1 that is independent of the brands in the probability vector.
- Rewriting the variance of $\bar{P}_i$ in the form $\theta = var(\bar{P}_i)/P_i(1-P_i)$, it can be seen that θ can be viewed as a measure of relative heterogeneity in that it expresses the validity $\bar{P}_i$ as a proportion of the total variation $P_i(1-P_i)$ under the complete heterogeneity.

Suppose that the population of N households can be grouped into L segments. Consider the stochastic process $\{X_t, t \in T\}$ where X_t denotes the brand purchased on occasion t by a randomly drawn household. Define $\bar{P}_{ij}$ as

$$\bar{P}_{ij} = P(X_t = i, X_{t+1} = j) \; i, j = 1, 2, ..., n.$$

The event $P(X_t = i, X_{t+1} = j)$ can happen in L mutually exclusive ways. The randomly drawn household may belong to any of the L segments and then choose brands j and l on the two consecutive purchase occasions. Let w_k be the probability that the household belongs to S_k, the k^{th} segment (k = 1, 2, . . . , L) and define $\bar{P}_{ij}^{(k)} = P(X_t = i, X_{t+1} = j/S_k)$. Then, $\bar{P}_{ij} = \sum_{k=1}^{L} \bar{P}_{ij}^{(k)} w_k$ where $w_k \geq 0$ for all k and $\sum_{k=1}^{L} w_k = 1$ $\{w_k$ is the relative size of the k^{th} segment$\}$. Taking expectations on both sides over all households in the population, we obtain:

$$P_{ij} = \sum_{k=1}^{L} P_{ij}^{(k)} w_k \tag{5.26}$$

where P_{ij} and $P_{ij}^{(k)}$ denote the expected values of $\bar{P}_{ij}$ and $\bar{P}_{ij}^{(k)}$, respectively.

Using the assumptions that the households follow a zero order purchase behaviour in each segment and that the choice probability vector is Dirichlet distributed over the households in each segment, we obtain the following equations:

$$P_{ij}^{(k)} = (1 - \theta_k) P_i^{(k)} P_j^{(k)}, \text{ for all } i \neq j \tag{5.27}$$

$$P_{ii}^{(k)} = \theta_k P_i^{(k)} + (1 - \theta_k) P_i^{(k)} P_i^{(k)}, \text{ for all } i \neq j \tag{5.28}$$

where θ denotes the heterogeneity parameter for the k^{th} segment.

Substituting equations (5.27) and (5.28) in equation (5.26) we obtain the following results:

$$\begin{aligned} P_{ij} &= \sum_{k=1}^{L} (1 - \theta_k) P_i^{(k)} P_j^{(k)} \\ P_{ii} &= \sum_{k=1}^{L} \left[\theta_k P_i^{(k)} + (1 - \theta_k) P_i^{(k)} P_i^{(k)} \right] w_k = \sum_{k=1}^{L} \theta_k P_i^{(k)} w_k + \sum_{k=1}^{L} (1 - \theta_k) P_i^{(k)} P_i^{(k)} w_k. \end{aligned}$$

The parameters in the above equations satisfy the following conditions:

1. $P_i^{(k)}, \theta_k, w_k \geq 0$ for all $k = 1, 2, 3, ..., L$ and $i = 1, 2, ..., n$

2. $\sum_{k=1}^{L} w_k = 1$ and $\sum_{k=1}^{L} P_i^{(k)} = \sum_{k=1}^{L} P_j^{(k)} = 1$ for all k.

The above equations and the restrictions represent our model.

Remark 7 *When $\theta_k = 0$ for all k, the model becomes:*

$$P_{ij} = \sum_{k=1}^{L} w_k P_i^{(k)} P_j^{(k)}$$

This is the Grover and Srinivasan model excluding the brand loyal segments. Also, this equation implies that all households within a segment are homogeneous in their choice probabilities.

Remark 8 *This model is a general type of latent class model. In fact it can be viewed as an extension of equation $P_{ij} = \sum_{k=1}^{L} w_k P_i^{(k)} P_j^{(k)}$ where the parameters $P_i^{(k)}$ and $P_j^{(k)}$ are not fixed constants but random variables distributed over the population.*

Estimation Procedure: Jain, Bass and Chen (1990) proposed an alternative maximum likelihood method for estimating the parameters of the model. They define $b_k = \theta_k w_k$, and substitute for $b_k, \theta_k w_k$. Thus, the model can be transformed as

$$P_{ij} = \sum_{k=1}^{L} (w_k - b_k) P_i^{(k)} P_j^{(k)} \qquad \text{for } i \neq j \tag{5.29}$$

$$P_{ii} = \sum_{k=1}^{L} b_k P_i^{(k)} + \sum_{k=1}^{L} (w_k - b_k) P_i^{(k)} P_i^{(k)} \tag{5.30}$$

From equations (5.29) and (5.30) $w_k, b_k, P_i^{(k)}$ and $P_j^{(k)}$ can be estimated. As the estimations of w_k and b_k are obtained, it is easy to estimate the parameter θ_k. The procedure for estimating $w_k, b_k, P_i^{(k)}$ and $P_j^{(k)}$ consists of two stages.

1. Derivation of the maximum likelihood estimates of the parameters $w_k, P_i^{(k)}$ and $P_j^{(k)}$, for fixed b_k's.

2. Performing a search over the b_k's to find the best b_k's that maximize the likelihood function.

Starting with a different set of initial values for the various parameters, we compare the solutions obtained by this iterative procedure to seek the solution that gives the minimum value of the chi-square statistic. $XX^2 = 2\sum_i \sum_j f_u \log \left(f_u / \hat{F}_u \right)$ where

$f_{ij}/ = N r_{ij}$, $\hat{F}_{ij} = N\ \hat{P}_{ij}$

r_{ij} = the observed values of the proportion of households in the $(i, j)^{th}$ cell of the contingency table

$\hat{P}_{ij}$= the estimated proportion values that are obtained from the iterative procedure, and

N =the total number of households in the sample.

Then, we minimize the equation of X^2 through a search for possible values of b_k's. The solution that minimizes the above equation of chi-square, yields the maximum likelihood estimates for the parameters.

The chi-square statistic in the above equation can be used to test the null hypothesis (H_o) that the L- class latent structure is true. When H_0 is true and the parameters of the model are locally identifiable, the statistic is asymptotically distributed as chi-square with $(n^2 - 1) - L(L - 1) - (n - 1)L = n^2 - (n + 1)L$ degrees of freedom.

Determining the number of segments: In latent structure analysis, one must prespecify the number of segments before estimating the parameters of the model. A way to determine the number of segments is to use a factor analytic model. There is a resemblance between this model and the latent structure model.

Let $\alpha_k = w_k - b_k$. By replacing b_k with $w_k - \alpha_k$ equation (5.30) can be rewritten as

$$P_{ii} = \sum_{k=1}^{L} w_k P_i^{(k)} + \sum_{k=1}^{L} \alpha_k P_i^{(k)} (P_i^{(k)} - 1) \tag{5.31}$$

The first term on the right side of this equation represents the aggregate market share of brand i, P_i. Therefore, this equation can be rewritten as:

$$P_i - P_{ii} = \sum \alpha_k P_i^{(k)}(P_i^{(k)} - 1) \tag{5.32}$$

We have assumed that the choice probability vector $(\bar{P}_1^{(k)}, \bar{P}_2^{(k)}, ..., \bar{P}_n^{(k)})$ in segment k follows a Dirichlet distribution. Using the expressions $Var(\bar{P}_i^{(k)}) = \theta P_i^{(k)}(1 - P_i^{(k)})$, $Cov(\bar{P}_i^{(k)}, \bar{P}_j^{(k)}) = -\theta P_i^{(k)} P_j^{(k)}$ the following results are obtained:

$$P_{ij} = \sum(-\alpha_k/\theta_k) Cov(\bar{P}_i^{(k)}, \bar{P}_j^{(k)}) \tag{5.33}$$

$$P_i - P_{ii} = \sum(\alpha_k/\theta_k) Var(\bar{P}_i^{(k)}) \tag{5.34}$$

By defining $t_k = \alpha_k/\theta_k$ the last equation can be rewritten as:

$$P_{ij} = \sum t_k[-Cov(\bar{P}_i^{(k)}, \bar{P}_j^{(k)})] \tag{5.35}$$

$$P_i - P_{ii} = \sum t_k Var(\bar{P}_i^{(k)}) \tag{5.36}$$

If we consider a matrix T with diagonal elements $T_{ii} = P_i - P_{ii}$ and the off-diagonal elements $T_{ii} = P_{ij}$, it is obvious that T is the weighted sum of L variance -covariance matrices. This procedure is analogous to principal components analysis, where a variance-covariance matrix is factored into a set of additive matrices so that this set of matrices best approximates the original variance-covariance matrix.

Therefore, we can apply factor analysis methods in order to obtain the number of segments for the latent class model. Also, from the above approximation, we conclude that a latent class model can be viewed as a discrete version of a factor analytic model.

5.3 Other Approaches with Latent Class Analysis

5.3.1 Grover and Dillon model

Grover and Dillon (1985) developed a probabilistic model which provides a general, flexible framework which can be used to test a hypothesized hierarchical market structure. The general probabilistic model can be easily implemented since it can be translated in terms of a restricted latent class models.

The basic concept of the general probabilistic model is simple. Each household can be classified into one of a set of mutually exclusive and exhaustive latent aspect classes. The latent aspect classes reflect the different evaluation strategies that are being used and define the likely pattern of switching that can occur since they determine the specific set of alternatives that a household considers when making a purchase decision. Another property of the model is that within a latent aspect class the purchase-to purchase transition obey the law of the local independence.

Before presenting the model we adopt some useful notation:

N_1 = the number of brands from which the household can choose at the first purchase occasion

N_2 = the number of brands from which the household can choose at the second purchase occasion

M = the number of latent aspect class types

1= a response vector with N1 elements consisting of 1's and 0's which indicates purchase or non purchase behaviour with respect to the first purchase occasion.

2 = a response vector with N2 elements consisting of 1's and 0's which indicates purchase or non purchase behaviour with respect to the second purchase occasion.

$O_{1,2}$ = a response vector of $N1 + N2$ dichotomously scored elements that reflect the household's observed purchase behaviour across the two purchase occasions.

$E_{1,2}^{(m)}$= a response vector of $N1 + N2$ dichotomously scored elements that reflect the household's expected purchase behaviour across the two purchase occasions for a household belonging to the mth latent aspect class.

Thus, the general model has the following form:

$$P(O_{1,2}) = \sum_{m=1}^{M} P(O_{1,2}/E_{1,2}^{(m)})P(m)$$

where:

$P(m)$ denotes the probability that a household is using the aspect strategy associated with the $m-th$ latent aspect class .

$P(O_{1,2}/E_{1,2}^{(m)})$ is a conditional probability. All these link the observed purchase history vectors to the expected purchase vectors which are governed by the latent aspect classes. A general representation for these conditional probabilities is:

$$P(O_{1,2}/E_{1,2}^{(m)}) = \prod_{t=1}^{2} \prod_{i=1}^{N_t} P(i_t/m)^{a_{it}}(1 - P(i_t/m))^{1-a_{it}}$$

where:

$P(i_t/m)$ is the conditional probability that the i-th brand is purchased at the t-th purchase occasion (t=1 or 2) given that the household belongs to the m-th latent aspect class.

$1 - P(i_t/m)$ is the conditional probability that the $i-th$ brand is not purchased at the t-th purchase occasion ($t = 1\ or\ 2$) given that the household belongs to the $m-th$ latent aspect class.

$$a_{it} = \begin{cases} 1, & \text{if the } i-th \text{ element found on } O_{1,2} \text{ is non-zero} \\ 0, & \text{otherwise} \end{cases}$$

Note: The conditional probabilities can be used to compute transition probabilities. Also the $P(i_t/m)$ are not the transition probabilities since the conditioning is with respect to the latent aspect classes and not to the brand purchased at a given purchase occasion.

So, the transition probabilities can be obtained by noting that:

$$P(i_t/m) = (P(i_t/m)/P(i_t))P(m)$$

where $P(i_t)$ is the marginal probability of purchasing the $i-th$ brand at the $t-th$ ($t = 1 or\ 2$) purchase occasion. The transition probabilities, denoted by $t_{i_1 i_2}$, can now be obtained from the following relationship: $t_{i_1 i_2} = \sum_{m=1}^{M} P(m|i_1)P(i_2|m)$.

5.3.2 Zahorik's Approach

Zahorik (1994), developed an aggregated model of market structure based on brand switching data which illustrates the market as a set of overlapping clusters of substitutable brands. During this trial, he generalized earliers models based on latent class analysis (e.g. Grover and Srinivasan model, Jain, Bass and Chen model) by accounting for heterogeneity among consumers and by allowing for brand switching across clusters a way of depicting variety seeking .

According to Zahorik, the term sequence refer to any two consecutive choices of brands in a product category, while the term 'switch' specifically refers to sequences of two different brands. Moreover, the model assumes that all brands switches are of two types:

1. Consistency Switches (CSs) which involve two brands perceived to be highly substitutable, including successive purchases of the same brands, and
2. Variety Switches (VSs) which involve brands perceived to differ on some important attribute.

There are many reasons why consumers might switch brands, even when the intended usage occasions are quite similar. In particular, the consumer may switch because:

1. He considers the two brands highly substitutable for the intended use.

2. He finds the favorite brand out of stock.

3. He switches to a brand which is being offered on deal.

4. He is bored or satiated with some aspect of the last brand purchased and may consciously seek a brand which will provide some variety.

5. He experiences a change in needs or circumstances, such as caused by the onset of cold weather, changes in the household membership, a change or diet, etc.

6. He tries an unfamiliar brand to learn about it.

7. He has undergone a change of perceptions of the relative merits of brands in the market due to new marketing programs, social influences, changes in personal values, or other reasons.

The models that were developed by Zahorik were extensions of the procedure described by Rao and Sabavala (1981) in which market structure for an individual was derived from switching patterns using cluster analysis. In their model (RS), a consumer perceives market structure as a hierarchy, with branching nodes representing the different levels of the major attributes on which brands are classified. Those brands grouped together at the end of the branches are perceived as essentially equivalent and substitutable to the consumer.

The individual model assumes that a consumer's history of purchase sequences consists of strings of zero-order CSs within PACs, separated by occasional (higher-order) VSs across PACs, as in the "leapfrog" models. The zero-order assumption within PACs is made under the following rationale: Once a consumer has decided upon a particular PAC for a specific choice occasion, by definition the brands within the PAC differ on only unimportant attributes, so past purchase history should not alter one's usual brand choice probabilities. The probability that a given purchase sequence is from brand i to brand j, is

$$\phi_{ij} = \gamma_{ij} + v_{ij} + error$$

where γ_{ij} is the probability that a given sequence is A CS from brand i to brand j, and v_{ij} is the probability that a given sequence is a VS from brand i to brand j.

The aggregate model follows from the individual model by taking expectations across the consumers. In order to achieve a workable parsimonious model, some assumptions about the joint distribution of individual probabilities would be made:

1. Brand switching within PACs is proportional to the brands' shares within the PAC.
2. Variety seeking switching between brands in different PACs is proportional to the brands' shares of sales within the PACs.
3. One's likelihood of switching between two specific PACs for variety is uncorrelated with the brands one chooses in those PACs.

The extensions of the model to include within PAC heterogeneity and cross-PAC switching preclude the use of latent class analysis for estimating the parameters.

Chapter 6

A Comparative Analysis for Brand Choice Models

6.1 Linear Learning, Markov and Bernoulli Models

According to the previous chapters, the most known models of brand choice are Linear Learning Models, Markov and Bernoulli models.

Givon and Horsky (1978) developed a model that describes the individual's brand choice behaviour. More specifically a brand choice behaviour of an individual consumer is assumed to be described by either Bernoulli, Markov or Linear Learning models. The Linear Learning model is the more general of the three models since the other two are constrained versions of it (zero-order Bernoulli and the first order Markov models).

An individual following the Linear Learning model in his brand choice behaviour will choose brand A at time t+1 out of a choice set A and $\bar{A}$ with conditional probability

$$\Pr(A_{t+1}|X_t) = \alpha + \beta X_t + \lambda P_t,\ \alpha, \beta, \lambda \geq 0; \alpha + \beta + \lambda \leq 1 \tag{6.1}$$

where A_t is the event of choosing brand A at time t,

P_t is the probability of that event, and

$$X_t = \begin{cases} 1 \text{ if } A_t \text{ occurs} \\ 0 \text{ otherwise} \end{cases}.$$

The probability of choice (6.1) is updated with each purchase by a non-feedback constant, α, by a feedback parameter β which represents the last experience with the brand, and by a fraction λ of the previous probability P_t. The magnitude of λ is related to the consumer's learning rate. If he learns fast, λ will be small and vice versa. If $\lambda = 0$, the linear learning model is equivalent to a first-order Markov process with purchase decisions as states having transition matrix:

$$P(j|i) = \begin{array}{c} \\ 1 \\ \\ 0 \end{array} \begin{array}{c} \begin{array}{cc} 1 & 0 \end{array} \\ \begin{bmatrix} \alpha + \beta & 1 - \alpha - \beta \\ & \\ \alpha & 1 - \alpha \end{bmatrix} \end{array}$$

In Markov model all households are going to have the same parameters as in the case of Linear Learning Models.

If $\alpha = \beta = 0$ and $\lambda = 1$, then $P_{t+1} = P_t$, which is the zero-order Bernoulli process. It can be proven that the feedback parameters β and λ are constrained to be nonnegative and equal for both brands. This precludes the possibility that an individual may have different experiences with different brands and even a negative experience with a certain brand. These limitations can be avoided if the parameters for both brands are allowed to differ. The individual's probability revision operators for the two brands will now be:

$$\Pr(A_{t+1}|X_t) = \alpha_1 + \beta_1 X_t - \beta_2 \bar{X}_t + \lambda_1 P_t - \lambda_2 \bar{P}_t$$

$$\Pr(\bar{A}_{t+1} \,|X_t) = \alpha_2 + \beta_1 X_t - \beta_2 \bar{X}_t + \lambda_1 P_t - \lambda_2 \bar{P}_t, \;\; \alpha_2 = 1 - \alpha_1$$

where the feedback parameters $\beta_1, \beta_2, \lambda_1, \lambda_2$ may be either positive or negative.

Population Heterogeneity

When the entire population, rather than just the individual, is considered, the parameters of the individual Linear Learning model will vary in their values across individuals. The parameters α, β, λ are assumed to be jointly distributed according to a Dirichlet distribution with a mass point at $\alpha = \beta,\ \lambda = 1$ and a mass plane at $\lambda = 0$. Thus, $f(\alpha, \beta, \lambda) =$

$$= \begin{cases} w_1 \text{ , if } \alpha = \beta, \lambda = 1 \\ w_2 \frac{\Gamma(v_3+v_4+v_6)}{\Gamma(v_3)\Gamma(v_4)\Gamma(v_6)} \alpha^{v_3-1} \beta^{v_4-1} (1-\alpha-\beta)^{v_6-1} \text{ , if } \lambda = 0, \alpha, \beta > 0, \alpha+\beta \le 1 \\ (1-w_1-w_2) \frac{\Gamma(v_3+v_4+v_5+v_6)}{\Gamma(v_3)\Gamma(v_4)\Gamma(v_5)(\Gamma(v_6)} \alpha^{v_3-1} \beta^{v_4-1} \lambda^{v_5-1} (1-\alpha-\beta-\lambda)^{v_6-1} \text{ ,} \\ \text{if } \alpha, \beta, \lambda > 0, \alpha+\beta+\lambda \le 1; \;\; w_1, w_2 \ge 0; w_1 + w_2 \le 1; v_3, v_4, v_5, v_6 > 0 \end{cases}$$

The weights w_1, w_2 and $(1 - w_1 - w_2)$ give the proportions of the population following the Bernoulli, Markov and Linear Learning processes, respectively. The initial probability P is assumed to be distributed independently of α, β, λ according to a Beta distribution with a mass point at P = 0:

$$f(p) = \begin{cases} w_0, \text{ if } \quad p = 0 \\ (1-w_0) \frac{\Gamma(v_1+v_2)}{\Gamma(v_1)\Gamma(v_2)} p^{v_1-1} (1-p)^{v_2-1}, \text{ if } 0 < p < 1; \;\; 0 \le w_0 \le 1; \;\; v_1, v_2 > 0 \end{cases}$$

The amount of consumer heterogeneity to be expected under different market conditions may be predictable. The diversity of individual experiences with brand A is likely to be related to the proportion of product class consumers purchasing the brand at least some of the time. If such a proportion is high, it is indicative of positive experiences by most consumers; if it is low, the diversity of experiences is probably large, few consumers have a positive experience and many have either no experience or a negative experience.

Market Share Models

The market share process will be derived through the aggregation of the individual consumers.

Heterogeneous Market Share Models

For an individual consumer with process parameters $\theta = (\alpha, \beta, \lambda, P)$ and a choice of X_0 at time $t = 0$, the probability of choosing brand A at time t equals

$$\Pr(A_t|X_0, \theta) = (\beta + \lambda)^{t-1}(\beta X_0 + \lambda P) + \alpha \frac{1 - (\beta + \lambda)^t}{1 - \beta - \lambda} \tag{6.2}$$

Given an initial market share m_0 the predicted heterogeneous market share of brand A at time t, m_t^r, depends on the expectation of the individual transition probabilities:

$$m_t^r = m_0 \underset{\theta}{E}[\Pr(A_t|A_0, \theta)] + (1 - m_0) \underset{\theta}{E}\left[\Pr(A_t| \bar{A}_0\ \theta)\right] \tag{6.3}$$

Using equation (6.3) and substituting $\Pr(A_t|X_0 = 1, \theta)$ for $\Pr(A_t|A_0, \theta)$ and $\Pr(A_t|X_0 = 0, \theta)$ for $\Pr(A_t| \bar{A}_0, \theta)$ the last equation becomes:

$$m_t^r = m_0 E\left[(\beta + \lambda)^{t-1}\beta\right] + E\left[(\beta + \lambda)^{t-1}\lambda\right] E(P) + E\left[\alpha \frac{1 - (\beta + \lambda)^t}{1 - \beta - \lambda}\right] \tag{6.4}$$

For the nested models Markov and Bernoulli, the predictions of market share m_t^r are, respectively,

$$m_0 E(\beta^t) + E\left[\alpha \frac{1 - \beta^t}{1 - \beta}\right] \text{ and } E(P).$$

6.2 Brand Switching Analysis via Latent Class Models

In the community of Marketing, a type of latent models are mixture models. These are relevant when "measurements are available from experimental units which are known to belong to a set of classes, but whose individual class-membership are unavailable" (Titterington, Smith and Markov, 1985). One special case of latent structure analysis, called latent class models, has attracted several marketing researchers. Lazarfeld (1950) saw latent class analysis as a measurement model, suitable for categorical or qualitative data and similar to factor analysis of quantitative data. Nevertheless, the interpretation of the model within marketing has been somewhat broader. Poulsen (1982) shows that the latent class model applied to panel data on brand choice, represents a mixture of zero-order, non-stationary choice processes. Close in spirit to the application in Poulsen are the works of Grover and Srinivasan (1987, 1989), where a latent class model was considered. Grover and Srinivasan also referred to the applicability of these models which provide a segmentation of the buyers in the market and insights into the competitive structure of the brands. Poulsen (1990) generalizes the latent class model in two separate directions, both relevant to the analysis of brand choice. Firstly, the assumption of zero-order behaviour, inherent in the latent class model, is relaxed by allowing first or higher order processes, leading to a mixture of Markov chain models. Secondly, he formulates a Markov model with states representing behavioural models that allow latent change processes.

A distinct advantage of latent structure analysis is that it simultaneously addresses two important issues facing a product manager: market structure analysis and market segmentation (Grover and Srinivasan, 1987). Also, it provides managerially useful information as estimates of the segment sizes and within segment brand shares inturn bear on the design of effective marketing strategies. Jain, Bass, Chen (1990) investigated issues concerning latent structure analysis as applied to the market structure.

In order to compare the previous models (Grover and Srinivasan model {GS model} and Jain, Bass, Chen model {JBC model}) we will illustrate the proposed approaches with a common application to the instant coffee market. Both approaches used the following cross-classification matrix (Grover and Srinivasan, 1987, pp146) :

Instant Coffee Cross – Classification Matrix

			HP	**TC**	**TC**	**FL**	**MH**	**S**	**S**	**MX**	**N**	**N**	**B**
			D	**C**	**D**	**C**	**C**	**D**	**D**	**C**	**C**	**D**	**D**
			R	**FD**	**FD**	**R**	**R**	**R**	**FD**	**FD**	**R**	**R**	**FD**
HP	**D**	**R**	93	7	17	19	18	43	1	4	6	7	10
TC	**C**	**FD**	9	80	12	11	24	7	4	2	6	3	3
TC	**D**	**FD**	9	14	46	3	7	7	4	2	2	0	9
FL	**C**	**R**	19	18	4	82	29	14	0	4	9	2	6
MH	**C**	**R**	26	11	6	35	184	24	3	11	18	6	6
S	**D**	**R**	15	13	8	13	28	127	4	3	3	8	8
S	**D**	**FD**	2	0	3	2	1	7	17	3	0	1	4
MX	**C**	**FD**	4	3	4	3	6	5	2	27	1	0	4
N	**C**	**R**	5	3	2	4	16	4	0	1	46	9	2
N	**D**	**R**	6	1	4	1	5	9	0	0	11	15	22
B	**D**	**FD**	10	4	4	4	2	10	2	2	5	2	27

Number of households buying row brand on the first purchase occasion and column brand on the second purchase occasion.

HP= High Point, TC = Taster's Choice, FL = Folgers, MH = Maxwell House, S = Sanka, MX = Maxim, N = Nescafe, B = Brim

D = Decaffeinated, C = Caffeinated

FD = Freeze dried, R = regular (spray dried)

****

This table contains data that come from the MRCA panel data (1981). In particular, this dataset consists of 4657 households that made at least two purchases of one or more of the 11 brands during a 12-month period. In order to ensure statistical independence of the observations in the cross-classification table, Grover and Srinivasan used only the data of the first two purchases of a household (among the set of 11 brands). This is the table that the authors used in both cases.

In this case, Jain and Rao (1994), compared the two proposed models by Grover and Srinivasan (1987) and Jain, Bass nad Chen (1990), respectively. They concluded the following:

1. The Latent Class Analysis is a method that gives important information about buyers behaviour. Specifically, it gives information on market structuring and market segmentation. It recovers the underlying structure in a product structure. Even if the market consists of overlapping segments, it will uncover such a structure.

2. By comparing the two proposed models, it is obvious that both GS and JBC approaches account for heterogeneity in households preferences. The choice of a particular approach will therefore be dictated by a prior knowledge of the market. If there exist brand loyal segments within the product category then the GS approach is preferable since JBC approach cannot provide estimates of the sizes of such segments. If the population consists of heterogeneous segments but not of brand loyal types, then the JBC is the appropriate approach.

3. Finally, JBS's procedure of using a factory analytic approach to determine the number of segments in a latent class models works very well. Hence, it should be considered for specifying the number of segments in latent structure analysis.

Chapter 7

Conclusions

The successful design and development of a product strategy is determined by the market researcher ability to obtain knowledge of the demands and expectations of his target group. Such knowledge requires a deep investigation as far as consumer behaviour is concerned. Although, in previous years results on consumer behaviour were easily extracted based on daily sellings, nowadays due to the complexity of market and huge number of available products, this is impossible. Thus, the contribution of the science of statistics in the estimation of the basic parameters, which constitute the consumer profile, is considered essential.

In this thesis, a review of stochastic consumer behaviour models has been attempted. Firstly, emphasis has been given to the statistical analysis of market data by fitting a variety of stochastic models classified according to "when", "what" and "how" purchase occurs. Thus, we referred to purchase incident models usually used to predict "when" and "how much" purchases occur in a specific time interval. A class of purchase incident models dealt with extensively in this thesis is the NBD models as well as extensions of this. Moreover, the model of interpurchase times at the individual level is referred too. As far as brand choice models are considered, we focused on zero-order models, Markov models and linear learning models. Finally, latent class models applied to consumer behaviour were dealt with.

It is obvious from the above, that several approaches have been suggested in the literature in order to cope with the needs of market research. The analysis of market using latent variables has grasped the interest of many marketing researchers, due to several factors. Progress has been made in methodological as well as computational capabilities. Moreover, the increase in the supply of single source and other types of panel data has improved the availability and quality of the basis for applying these methods.

Latent models offer a framework for analysing and interpreting buyer behaviour by focussing on the underlying structure such as segment-membership and segment profile. By segmenting buyers according to their behaviour one manages to identify buyers that differ significantly and thus seek for explanations regarding their buying behaviour. Understanding the underlying reasons for an observed behaviour will increase the ability to predict future behaviour, which is of central importance to the marketer.

Bibliography

[1] **Agresti, A. (1990). Categorical Data Analysis**. Wiley - Inter-science Publication.

[2] **Aickin, M. (1983)**. **Linear Statistical Analysis of Discrete Data**. John Wiley & Sons.

[3] **Bartholomew and Knott (1998). Latent Variable Models and Factor Analysis**. Oxford University Press Ins., New York.

[4] **Bass, F. (1974)**. The Theory of Stochastic Preference and Brand Switching, **Journal of Marketing Research**, XI, 1-20.

[5] **Bass, F., Jeuland, A. and Wright, G. (1976)**. Equilibrium Stochastic Choice and Market Penetration Theories: Derivation and Comparisons. **Management Science**, 22,1051-1063.

[6] **Bass, F. M., Givon, M.M., Kalwani, M.V., Reibstein, D.and Wright, G.P. (1984)**. An Investigation into the Order of the Brand Choice Process. **Marketing Science**, 267-287.

[7] **Bass, F.M. and Chen, Y. (1989)**. Incorporating Heterogeneity of Choice Probabilities in Latent Class Models: An application to Market Structure Analysis. Working paper, School of Management, **University of Texas at Dallas**.

[8] **Bemmaor A. C. and Schmittlein D.C**. (1991). Models of Purchase Timing and Models of Brand Choice: Outlook and Issues. **International Journal of Research in Marketing**, 8, 3, 163-8.

[9] **Brockett, P. L., Golden L. L. and Panjer H. H. (1996)**. Flexible Purchase Frequency Modeling. **Journal of Marketing Research**,Vol.XXXIII, 94-107.

[10] **Bucklin, E. R. and Gupta, S. (1992)**. Brand Choice, Purchase incidence, and Segmentation: An Integrated Modeling Approach. **Journal of Marketing Research**, XXIX, 201 - 215.

[11] **Bye, V. B. and Schechter, S. E. (1986)**. A Latent Markov Model Approach to the Estimation of Response Errors in Multiwave Panel Data. **Journal of the American Statistical Association**, 81, 393, 375-380.

[12] **Chatfield, C. and Goodhardt, J. G. (1973)**. A Consumer Purchasing Model with Erlang Inter- purchase Times. **Journal of the American Statistical Association**, 68, 828-835.

[13] **Chatterjee, S. C. and Sudharshan, D. (1994)**. MARKDEF: A Methodology for Understanding Competitive Spaces Based on Consumer Perceptions. **Behaviormetrika**, 97-119.

[14] **Clogg, C. C. and Goodman, A. L. (1984)**. Latent Structure Analysis of a set of Multidimensional Contingency Tables. **Journal of the American Statistical Association**, 79, 388, 762-771.

[15] **Coleman, J.S. (1964)**. **Models of Change and Response Uncertainty**. Englewood Cliffs, New Jersey: Prentice Hall, Inc.

[16] **Colombo, A. R. and Morrison, G. D. (1989)**. A Brand Switching Model with Implications for Marketing Strategies. **Marketing Science**, 3, 89-99.

[17] **Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977)**. Maximum Likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society**, Series A, 39, 1-38.

[18] **Duhammel, W. F. (1966)**. The use of Variable Markov Processes as a partial basis for the Determination and Analysis of Market Segments. Unpublished Ph.D disserattion at the Graduate **School of Business Stanford University**.

[19] **Ehrenberg, A.S.(1972)**. **Repeat-Buying: Theory and Applications**. Amsterdam: North-Holland.

[20] **Ehrenberg, A.S. (1959)**. The Pattern of Consumer Purchase. **Applied Statistics**, VIII, 26-41.

[21] **Eshima, N. (1993)**. Dynamic Latent Structure Analysis through the Latent Markov Chain Model. **Behaviormetrika**, 151-160.

[22] **Everit, B. S. (1976). The Analysis of Contingency Tables**. Chapman and Hall.

[23] **Frank, R.,E. (1962)**. Brand Choice as a Probability Process. **Journal of Business**, XXXV, 43-56.

[24] **Formann, K. A. (1992)**. Linear Logistic Latent Class Analysis for Polytomous Data. **Americal Statistical Association**, 87, 418, 476-486.

[25] **Givon, M. and Horsky, D. (1990)**. Untangling the Effects of Purchase Reinforcement and Advertising Carryover. **Marketing Science**, 9, 2, 171-187.

[26] **Givon, M and Horsky, D. (1979)**. Application of a Composite stochastic Model of Brand Choice. **Journal of Marketing Research**, XVI, 258-267.

[27] **Givon, M. and Horsky, D. (1978)**. Market Share Models as Approximators of Aggregated Heterogeneous Brand Choice Behavior. **Management Science**, 24, 13, 1404- 1416.

[28] **Goodman, Leo, (1974a)**. Explanatory Latent Structure Analysis using both identifiable and unidentifiable models. **Biometrika**, 61, 215-231.

[29] **Goodman, Leo, (1974b)**. The Analysis of Systems of Qualitative Variables when some of the variables are Unobservable. Part I - Modified Structure Approach. **American Journal of Sociology**, 79, 5, 1179 - 1259.

[30] **Green, E. P., Carmone, J. F. and Wachspress, P. D. (1976)**. Consumer Segmentation via Latent Class Analysis. **Journal of Consumer Research**, 3, 170- 174.

[31] **Grover, R. and Dillon W. (1985)**. A Probabilistic Model for testing Hypothesized Hierarchical Market Structures. **Marketing Science**, l.4, 312-335.

[32] **Grover, R. and Srinivasan, V. (1987)**. A simultaneous Approach to Market Segmentation and Market Structuring. **Journal of Marketing Research**, XXIV, 139-153.

[33] **Grover, R. and Srinivasan, V. (1989)**. An Approach for Tracking Within-Segment Shifts in Market Shares. **Journal of Marketing Research**, 230-236.

[34] **Guadagni, P. and Little J. D.C. (1983)**. A Logit Model of Brand Choice Calibrated on Scanner Data. **Marketing Science** 2 (3), 203-238.

[35] **Haberman, S. J. (1978)**. **Analysis of Qualitative Data**. Vol. 1. New York: Academic Press.

[36] **Haberman, S. J. (1988)**. A Stabilised Newton - Raphson Algorithm for log - linear Models for Frequency Tables Derived by Indirect Observation. In Sociological Methodology 1988, ed. C.C. Clogg, Washington: **American Sociological Association**, 193- 211.

[37] **Herniter, D. J and Howard, A. R. (1964)**. **Stochastic Marketing Models. Progress in Operations Research**, vol. II. New York: John Wiley & Sons, Inc.

[38] **Howard, R. A. (1963)**. Stochastic Process Models of Consumer Behavior. **Journal of Advertising Research**, 3, 3 (September), 35 - 42.

[39] **Howard, R. A. (1965)**. Dynamic Inference. **Journal of the Operations Research Society of America**, 13, 2, 712-733.

[40] **Jain, D., Bass, F. M. and Chen, Y. (1990)**. Estimation of Latent Class Models with Heterogeneous Choice Probabilities: An Application to Market Structuring. **Journal of Marketing Research**, XXVII, 94-101.

[41] **Jain, D. (1986)**. Studies in Stochastic Choice Models with Implications for Market Structure. Unpublished doctoral dissertation, **University of Texas at Dallas**.

[42] **Jain, D. and Rao, C.R. (1994)**. Latent Class Models to Infer Market Structure: A Comparative Analysis. **European Journal of Operational Research**, 331-343.

[43] **Jeuland, A, Bass, F and Wright, G. (1980)**. A Multibrand Stochastic Model Compounding Heterogeneous Erland Timing and Multinomial Choice Processes. **Operations Research**, 28, 2, 255-277.

[44] **Jobson, J. D. (1992)**. **Applied Multivariate Data Analysis** (Volume II: Categorical and Multivariate Methods). (Springer)

[45] **Jones, J. M.(1969)**. A Non-Stationary Probability Diffusion Model of Consumer Brand Choice Behavior. Unpublished doctoral dissertation in the department of Operations Research, **Standford University**.

[46] **Kahn, E. B. (1987)**. A Theoretical Model of Interpurchase Times. **Applied Stochastic Models and Data Analysis**, 3, 93- 109.

[47] **Kalwani, M.U. (1979)**. **An Information Processing Theory of Consumer Choice**. Addison-Wesley Publishing Co.,Inc. Reading, MA.

[48] **Kalwani, M. U. and Morrison, G.D. (1977)**. Estimating the Proportions of 'Always Buy' Consumers: A likelihood Ratio Test with Sample Size Implications. **Journal of Marketing Research**., 14, 601-606.

[49] **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**.

[50] **Kuehn, A. A. (1962)**. Consumer Brand- Choice - A Learning Process? **Journal of Advertising Research**, II, 10-17.

[51] **Kuehn, A. A. and Rohloff, A.C. (1967). Consumer Response to Promotions**. (In Patrick J. Robinson et al., Promotional Decisions Using Mathematical Models), Boston: Allyn and Bacon, Inc.

[52] **Kumar, A. and Sashi, C. M. (1989)**. Confirmatory Analysis of Aggregate Hierarchical Market Structures: Inferences from Brand Switching Behavior. **Journal of Marketing Research**, 444-453.

[53] **Lattin, J. M. and McAlister,L. (1985)**. Using a Variety-Seeking Model to Identify Substitute and Complementary Relationships Among Competing Products. **Journal of Marketing Research**, 330-339.

[54] **Lazarsfeld, P. F. (1950)**. **The logical and Mathematical Foundation in Latent Structure Analysis**, in Studies in Social Psychology in World W. II (vol. IV: Measurement and Prediction), Princeton NJ: Princeton University Press, 362 - 412.

[55] **Leeflang, S. H. P. and Boonstra, A. (1982)**. Some Comments on the Development and Application of Linear Learning Models. **Management Science**, 28, 11, 1233- 1246.

[56] **Lilien, G., Kotler, P. and Moorthy, S. (1992)**. **Marketing Models**. Prentice-Hall International Editions.

[57] **Lipstein, B. (1965).** A Mathematical Model of Consumer Behavior. **Journal of Marketing Research**, 2, 259-265.

[58] **Lipstein, B. (1959)**. The Dynamics of Brand Loyalty and Brand Switching, in Better Measurements of Advertising Effectiveness: The Challenge of the 1960's. **Proceedings of the fifth Annual Conference of the Advertising Research Foundation**, New York.

[59] **Massy, W.F. (1966)**. Order of Homogeneity of Family Specific Brand - Switching Processes. **Journal of Marketing Research**, III, 1, 48 - 54.

[60] **Massy, W. F., Montgomery, D. B. and Morrison,D. G. (1970)**. **Stochastic Models of Buying Behavior**, Cambridge, MA: The MIT Press.

[61] **Matsuda, N. and Namatame, M. (1995)**. Interactive Measurements of Hierarchically Related Consumers' Images. **Behaviormetrika**, 129-143.

[62] **Maxwell, A. E. (1979)**. **Multivariate Analysis in Behavioral Research**. Chapman and Hall.

[63] **Mayekawa, S. I. (1994)**. Equivalent Path Models in Linear Structural Equation Models. **Behaviormetrika**, 79-96.

[64] **Myers and Nicosia F. M. (1968)**. On the Study of Consumer Typologies. **Journal of Marketing Research**, V, 34- 58.

[65] **Mayekawa, S. I. (1996)**. Maximum Likelihood Estimation of the Cell Probabilities under Linear Constrains. **Behaviormetrika**, 111-128.

[66] **Meiser, T. and Ohrt, B. (1996)**. Modeling Structure and Chance in Transitions: Mixed Latent Partial Markov - Chain Models. **Journal of Educational and Behavioral Statistics**, 21, 2, 91- 109.

[67] **Morrisson, D.G. (1965 a)**. Stochastic Models for Time Series with Applications in Marketing. **Techniqual report No. 8, Program in Operations Research, Stanford University**.

[68] **Morrisson, D.G. (1965 b)**. New Models of Consumer Behavior: Aid in Setting and Evaluating Marketing Plan. **Conference of the American Marketing Association**, September 1 - 3 , Washington, D.C.

[69] **Morrison, D.G. (1966).** Testing Brand-Switching Models. **Journal of Marketing Research**, 3, 401-409.

[70] **Morrison, D.G. and Schmittlein, D.C. (1981)**. Predicting Future Random Events Based on Past Performance. **Management Science**, 27, 9, 1006-1023.

[71] **Morrison, D.G. and Schmittlein, D.C. (1988)**. Generalizing the NBD Model for Customer Purchases: What Are the Implications and is it Worth the Effort? (with Discussion and Rejoinder). **Journal of Business and Economic Statistics**, 6, 2, 145-166.

[72] **Montgomery, D.,B. (1966)**. A probability Diffusion Model of Dynamic Market Behavior. Working Paper No. 205 - 66, Alfred P. Sloan School of Management, **Massachusetts Institute of Technology**, March.

[73] **Mooijaart, A. and Van der Heijden, P.G.M. (1992)**. The EM Algorithm for Latent Class Analysis with Equality Constraints. **Psychometrika**, 57, 2, 261-269.

[74] **Novac, T. P. and Stangor, C. (1987)**. Testing Competitive Market Structures: An Application of Weighted Least Squares Methodology to Brand Switching Data. **Marketing Science**, 82-97.

[75] **Novac, T. P. (1993).** Log-linear Trees : Models of Market Structure in Brand Switching Data. **Journal of Marketing Research**, 267-287.

[76] **Oakes, M. (1986)**. **Statistical Inference: A Commentary for the Social and Behavioral Sciences**. John Wiley & Sons.

[77] **Popkowski, L.T. L. and Bass, M. F. (1998)**. Determining the Effects of Observed and Unobserved Heterogeneity on Consumer Brand Choice. **Applied Stochastic Models and Data Analysis**, 14, 95 - 115.

[78] **Poulsen, S. C. (1990)**. Mixed Markov and Latent Markov modeling Applied to Brand Choice Behavior. **International Journal of Research in Marketing,** 7, 5-19.

[79] **Rao V.R. and Sabavala, D.J. (1981)**. Inference of Hierarchical Choice Processes from Panel Data. **Journal of Consumer Research** 8, 85-96.

[80] **Schafer, J. L. (1997)**. **Analysis of Incomplete Multivariate Data**. Chapman and Hall.

[81] **Schmittlein, D.C. and Morrison, D.G. (1983)**. Prediction of Future Random Events with the Condensed Negative Binomial Distribution. **Journal of the American Statistical Association**, 78, 382, 449-456.

[82] **Schmittlein, D, Morisson, D and Colombo, R. (1987)**. Counting Your Customers: Who are They and What will They Do Next? **Management Science**, 33, 1, 1-24.

[83] **Shigemasu, K., Sugiyama, N. (1994)**. Latent Class Analysis of Choice Behavior. **Behaviormetrika**, 49-60.

[84] **Tabachnick, B. G. , Fidell, L. S. (1996)**. Using Multivariate Statistics. Harper Collins College Publishers.

[85] **Telster, L. (1963). Least Square Estimates of Transition Probabilities. In measurement in Economics**, Christ et al., Stanford, California: Stanford University Press, 270-292.

[86] **Titterington, D.M., Smith, A.F.M and Markov**,**U.E. (1985). Statistical Analysis of Finite Mixture Distributions**. New York, NY: Wiley.

[87] **Trivedi, M. and Morgan, S. M. (1996)**. Brand - Specific Heterogeneity and Market - Level Brand Switching. **Journal of Product and Brand Management**, 5, 29-39.

[88] **Vilcassim J. N. and Jain, C. D. (1991)**. Modeling Purchase-Timing and Brand Switching Behavior Incorporating Explanatory Variables and Unobserved Heterogeneity. **Journal of Marketing Research,** XXVIII, 29-41.

[89] **Zahorik, A. J. (1994)**. A Non-Hierarchical Brand Switching Model for Inferring Market Structure. **European Journal of Operational Research**, 344-358.