Aalto University
School of Science

Juha Nuutinen

A Consumer Choice Model in Retailing of Health Products

Master's thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology.

Helsinki, April 10, 2012

Supervisor: Professor Ahti Salo
Instructor: Pharmacist Jarkko Kiuru

AALTO UNIVERSITY
SCHOOL OF SCIENCE

ABSTRACT OF THE
MASTER'S THESIS

Author:	Juha Nuutinen
Title:	A Consumer Choice Model in Retailing of Health Products
Date:	April 10, 2012 **Number of pages:** 98
Degree Programme:	Degree Programme in Engineering Physics and Mathematics
Major subject:	Systems and Operations Research
Minor subject:	Economics
Supervisor:	Prof. Ahti Salo
Instructor:	M.Sc. (Pharm.) Jarkko Kiuru

In recent years, the amount of detailed information on consumers' purchasing behaviour has increased through the automated recording of sales transactions, the widespread use of customer loyalty cards and the increasing popularity of e-commerce. However, the potential of this data in gaining insight into consumers' purchase behaviour is rarely used due to the scarcity of applications in the literature and challenging modelling environments. This thesis shows that consumer-level point-of-sale data can be easily and successfully analysed using only the standard desktop applications.

The thesis presents two different models in the multinomial logit framework: the stock-keeping unit-based model of Guadagni and Little (1983) and the attribute level-based model of Fader and Hardie (1996). These models are calibrated on point-of-sale data from retail sales of health products to healthcare personnel in 2008-2009 across six product categories. All analyses are performed in Microsoft Excel spreadsheet.

The results demonstrate that preferences toward product attributes (such as brand, package size, form, formula) can be used to predict the share of purchases of new products. Attribute loyalty and previous purchase variables capture heterogeneity across consumers and state dependence in purchase decisions over time. Promotion and discounts are shown to increase substantially the choice probability of a product. The models are able to predict the share of purchases in time for new and existing products in the forecasting period 2010-2011 and for an independent set of customers. The results increase the retailer's understanding of consumers' purchase behaviour, make it possible to forecast the demand of new products and provide the basis for a cost-benefit analysis of marketing actions.

Keywords: consumer choice, multinomial logit, consumer preferences, attribute loyalty, retailing of health products, point-of-sale data, new product sales forecasting

AALTO-YLIOPISTO
PERUSTIETEIDEN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä:	Juha Nuutinen		
Työn nimi:	Kuluttajan valintamalli terveystuotteiden vähittäiskaupassa		
Päivämäärä:	10. huhtikuuta 2012	**Sivumäärä:**	98
Koulutusohjelma:	Teknillisen fysiikan ja matematiikan koulutusohjelma		
Pääaineen nimi:	Systeemi- ja operaatiotutkimus		
Sivuaineen nimi:	Kansantaloustiede		
Työn valvoja:	Prof. Ahti Salo		
Työn ohjaaja:	Prov. Jarkko Kiuru		

Yksityiskohtaisen tiedon määrä kuluttajien ostokäyttäytymisestä on kasvanut viime vuosina myyntitapahtumien automaattisen tallentamisen, kanta-asiakaskorttien yleistyneen käytön ja verkkokaupan kasvaneen suosion ansiosta. Tätä tietoa käytetään kuitenkin harvoin täysimääräisesti hyväksi johtuen sovellusten vähäisestä määrästä kirjallisuudessa ja haastavista mallinnusympäristöistä. Tämä työ osoittaa, että kuluttajakohtaista myyntidataa voidaan analysoida helposti ja tuloksellisesti käyttäen vain yleisiä tietokonesovelluksia.

Työssä kehitetään kaksi multinomiaaliseen logit-luokkaan kuuluvaa mallia: varastoyksikköpohjainen malli (Guadagni ja Little 1983) ja attribuuttipohjainen malli (Fader ja Hardie 1996). Mallit kalibroidaan terveysalan henkilöstön terveystuotteiden vähittäiskaupan myyntidatalla aikavälillä 2008-2009 kuudelle eri tuotekategorialle. Analyysit tehdään Microsoft Excel-taulukkolaskentaohjelmalla.

Tulokset osoittavat, että kuluttajien attribuuttikohtaisia tuotepreferenssejä (kuten brändi, pakkauskoko, olomuoto, kaava) voidaan käyttää uusien tuotteiden myynnin osuuden ennustamiseen. Attribuuttiuskollisuus ja edeltävä ostopäätös kuvaavat kuluttajien heterogeenisuutta ja aikaisempien kulutustottumusten vaikutusta kuluttajan valintaan. Myynninedistäminen ja alennus lisäävät merkittävästi tuotteen valinnan todennäköisyyttä. Mallit selittävät sekä uusien että aiemmin valikoimassa olevien tuotteiden myynnin kehittymistä ennustusajanjaksolla 2010-2011 ja itsenäiselle ryhmälle asiakkaita. Tulokset auttavat yritystä ymmärtämään kuluttajien ostokäyttäytymistä, mahdollistavat uusien tuotteiden menekin arvioimisen ja tarjoavat perustan markkinointitoimenpiteiden kustannustehokkuuden arvioimiselle.

Avainsanat: kuluttajan valinta, multinomiaalinen logit, kuluttajien preferenssit, attribuuttiuskollisuus, terveystuotteiden vähittäiskauppa, myyntidata, uusien tuotteiden myynnin ennustaminen

Acknowledgements

There are many people who deserve my deepest gratitude.

I would like to thank my instructor, Jarkko Kiuru, for guiding this thesis. Your guidance has been invaluable. Your enthusiasm over the sales and marketing of health products has inspired my work. I have truly enjoyed our lively discussions on marketing topics and new business opportunities.

I am very grateful to my director, Lauri Untamo, for supporting my work. Your empowering leadership style has provided me with the freedom to work independently and find my own solutions to the challenges at hand. You gave me an opportunity to work in a stimulating environment on interesting topics.

I wish to express my gratitude to my supervisor, Ahti Salo, for directing my research. Your insightful comments have provided me with clear direction.

There are many individuals who have greatly contributed to this thesis in my workplace. I wish to thank members of the thesis steering committee for your guidance and helpful comments. I wish to thank people in the Information Management department for providing the valuable point-of-sale data set. People in the Consumer Health unit provided me with valuable marketing and product expertise. Last, and most importantly, people in the Accounting department took me in and offered their friendship. You have made this experience unforgettable!

This thesis could not have been accomplished without the support of a number individuals who are too numerous to be all mentioned here. They are family, friends, co-workers, fellow students and other people who have supported me in many ways. I gratefully acknowledge your contributions to this thesis.

Helsinki, April 10, 2012

Juha Nuutinen

Contents

Abbreviations

CNL	Cross-Nested Logit
DM	Decision Maker
ED	Exogenous Demand
ERP	Enterprise Resource Planning
FH	Fader and Hardie (1996)
GL	Guadagni and Little (1983)
IIA	Independence from Irrelevant Alternatives
iid	Independent and Identically Distributed
LL	Log-likelihood
MLE	Maximum Likelihood Estimator
MMNL	Mixed Multinomial Logit
MNL	Multinomial Logit
MNP	Multinomial Probit
NL	Nested Logit
POS	Point of Sale
RUM	Random Utility Model
SKU	Stock-Keeping Unit

1 Introduction

1.1 Background

Consumer choice modelling has become more important in recent decades as retailers and manufacturers have understood the benefits of learning the determinants of consumers' choice behaviour. Retailers gain by reorganizing their assortments to best serve customer needs and to maximize category profits. Improved knowledge of consumers' response to marketing mix variables helps in designing more efficient and better-targeted marketing programmes. Manufacturers aim to better understand consumers' preferences toward product attributes and characteristics to identify the best product line extension opportunities.

The methodology and theoretical framework to the research of consumer choice was developed by McFadden (1973) in the form of conditional logit analysis. The absence of automated recording of purchases and consumer-level shopping data delayed empirical testing by a decade. The introduction of barcodes and optical scanners in the late 70s made it possible to conduct empirical tests with scanner data (gathered at the point of purchase by an electronic reader). Beginning from the seminal work of Guadagni and Little (1983), the standard approach to modelling product choice has involved the use of the multinomial logit model (MNL). Their pioneering use of scanner panel data showed the usefulness of the MNL model in estimating the impact of marketing mix variables on demand. They also showed that customer loyalty is the single most important factor in predicting the future consumption behaviour of households. Fader and Hardie (1996) proposed the modelling of products through their attributes to achieve a more parsimonious model, managerial insights into consumers' preferences toward product attributes and ability to forecast the demand for new products.

Today, the automated recording of sales transactions, widespread use of customer loyalty cards and careful tracking of marketing actions has greatly increased the opportunities for consumer choice analysis. However, there is a shortage of capable marketing professionals with sufficient data analysis skills who can communicate their results to the boardroom. This thesis seeks to give an overview of recent developments in the field of marketing and show how consumer choice can be modelled in practice.

1.2 Research Objectives

Our main objective is to develop a multinomial logit model to explain consumers' purchase behaviour. Ideally, this model should be easy to implement and use without the need for specialized software or mathematical background. Therefore, we use only standard desktop applications in conducting the analysis presented in this thesis. Furthermore, the model should be able to predict consumers' purchase behaviour based on explanatory variables. This makes it possible to forecast demand for products and assess the cost-effectiveness of marketing actions.

The research objectives for this thesis are as follows.

1) Perform a comprehensive literature review on discrete choice models and their applications in the marketing literature.

2) Formulate a model to explain consumers' purchase behaviour in the retailing of health products.

3) Estimate consumers' preferences toward the attributes of products (e.g., brand, package size, form, formula). Estimate the impact of attribute loyalty and marketing mix variables on the stock-keeping unit's (SKU) share of purchases in a category.

4) Analyse the differences between product categories in consumers' purchase behaviour and marketing response.

5) Evaluate the model by forecasting the SKU's share of purchases in a future period and for an independent set of customers. Evaluate the performance of the model in predicting the share of purchases for new SKUs.

1.3 Scope of the Thesis

The most comprehensive models of purchase behaviour explain consumers' decisions on 1) whether to buy (purchase incidence), 2) what to buy (brand choice), and 3) how much to buy (purchase quantity). This thesis focuses solely on the second question of brand choice. We assume that there is no significant category expansion associated with the introduction of new products or increasing marketing activity. Total category purchases remain the same in our model, making brand shifting and cannibalization the only sources of new demand for the new SKUs. The analysis is limited to six different product categories. The categories analysed include hand sanitizer, magnesium, multivitamin, omega-3, probiotic and vitamin D. We examine each of these categories independently, ignoring possible cross-category relationships.

1.4 Structure of the Thesis

The structure of this thesis is as follows.

- Chapter 2 provides an introduction to discrete choice models. This chapter presents some of the basic concepts in choice modelling, alternative model structures and comparison of models.

- Chapter 3 introduces applications of discrete choice models in marketing. We review some of the most important areas of research and show applications of discrete choice models in the marketing literature.

- Chapter 4 presents the specification of the multinomial logit model, the maximum likelihood estimation method and the techniques used in evaluating the goodness of fit.

- Chapter 5 introduces the empirical data used to calibrate and evaluate the model. We illustrate and analyse the consumer-level point-of-sale data and the historical data on marketing actions.

- Chapter 6 presents the empirical results. We analyse and interpret the parameter estimates. We evaluate the performance of the model by tracking the share of purchases in a future period and for an independent set of customers. We analyse the performance of the model in predicting the share of purchases for new products.
- Chapter 7 concludes with a summary of main findings and proposals for further improvements.

2 Discrete Choice Models

2.1 Basic Concepts of Discrete Choice

Discrete choice models assume that a decision maker chooses a single alternative from a choice set. The choice set is made up of a finite number of mutually exclusive alternatives. The decision rule is the process by which the decision maker evaluates the alternatives in the choice set and makes the choice. In this section, we briefly introduce the basic concepts and the common framework of discrete choice models. See Ben-Akiva and Lerman (1985) for a detailed discussion.

Decision Maker

The decision maker (DM) is an entity that chooses among a set of alternatives. The unit that makes a decision can be an individual person, a group of individuals such as a family or a household, a company or an organization. Decision makers may have different socio-demographic attributes (e.g., age, gender, income) and heterogeneous preferences deriving from different needs, tastes and habits. These attributes may have an impact on the choice.

Alternatives and Choice Set

The choice is made among a set of alternatives (e.g., product, travel mode). The set of all possible alternatives is called the universal set. A decision maker considers only a subset of the universal set, called the choice set. The choice set includes only those alternatives that are feasible, available, known and considered by the decision maker during the decision process. Alternatives in the universal set that do not meet these requirements are excluded from the choice set. The choice set contains a finite number of alternatives that can be explicitly listed. The alternatives must be mutually exclusive so that only a single alternative from the choice set is chosen. The choice set must be exhaustive so that it includes all possible alternatives.

Attributes of Alternatives

Each alternative is characterized by its attributes (e.g., price, brand, size). Attributes may be common to all alternatives or alternative specific. The attractiveness of an alternative is evaluated in terms of values of the set of attributes that are specific to the alternative. Decision makers can attach diverse values for the same attribute of the same alternative, deriving from the heterogeneity of preferences.

Decision Rule

The decision rule guides the choice of the decision maker. It describes the internal mechanisms used by the decision maker to process the available information and arrive at a unique choice. A variety of rules can be employed, including dominance and elimination by aspects. However, most discrete choice models are based on the utility theory. This class of decision rules assumes that the attractiveness of an alternative characterized by its attributes can be captured by the utility index. The decision maker selects the alternative in the choice set with the highest utility. The probabilistic choice theory based on the assumption of utility maximizing behaviour of the decision makers is described in the following section.

2.2 Random Utility Theory

In random utility theory, a decision maker is always assumed to choose the alternative with the highest utility. However, the utility cannot be completely observed. This is due to the incomplete information and the inherent randomness involved in the choice behaviour. Manski (1977) argues that the randomness in choice behaviour is caused by unobserved attributes of alternatives, unobserved attributes of decision makers, measurement errors and functional misspecification. The observed inconsistencies in choice behaviour are taken to be the result of observational deficiencies on the part of the observer (Ben-Akiva and Lerman 1985). As a result, the utility is presented as a random variable. It can be divided into an observed (deterministic) component and an unobserved (random or error) component. Total utility associated with alternative j for decision maker i at time t is given as

$$TU_j^i = U_j^i + \varepsilon_j^i = U\left(x_j^i, s^i, t^i\right) + \varepsilon_j^i. \tag{1}$$

The deterministic component of utility U_j^i can be expressed as a function of x_j^i (vector of observed attributes of alternative j as faced by decision maker i), s^i (vector of socio-demographic attributes of decision maker i) and t^i (vector of parameters representing the tastes of decision maker i, to be estimated from the data).

Assumption of the distribution of error terms ε_j^i defines the discrete choice model. Thus, various discrete choice models can be generated based on alternative distributional assumptions on the error terms. Assumption that the error terms follow a joint normal distribution results in the multinomial probit (MNP) model. Assumption of independently and identically distributed (iid) extreme value error terms results in the multinomial logit (MNL) model. Assumptions of the error terms of the MNL model are considered in detail in the next section, because relaxing some of these assumptions leads to important generalizations of the MNL model.

2.3 Assumptions of the Multinomial Logit Model

The multinomial logit (MNL) model, introduced more specifically in the next section, is the most widely used discrete choice model in the marketing literature. It imposes several restrictive assumptions about the structure of error terms. These include the assumption of independent and identically distributed error terms, response homogeneity and error variance-covariance homogeneity. Together they lead to the simple and elegant closed form expression of choice probabilities. More general models of choice are obtained by relaxing some of these assumptions. The description of restricted assumptions underlying the MNL model presented in this section is based on Bhat (1997) and Bhat et al. (2000).

The first assumption of independent and identically distributed (iid) error terms assumes that error terms follow type I extreme value (or Gumbel) distribution. The assumption of independence of error terms across alternatives implies that there are no common unobserved factors affecting the utilities of the alternatives. This assumption is violated if unobserved factors related to one alternative are similar to those related to another alternative. For example, the decision maker might assign a higher utility to all alternatives that share the same brand or a lower utility to all alternatives that share the same size. The assumption of independence also prevents the dependence of choices over time. The relaxation of this assumption permits covariance in error terms, resulting in more realistic substitution patterns between alternatives. The assumption of identically distributed error terms across alternatives implies that the extent of variation in unobserved factors affecting the utilities of the alternatives is the same across all alternatives, i.e., error terms are homoscedastic across alternatives. In general, there is no theoretical reason to believe that this is the case. This assumption is violated if the values of unobserved variables (e.g., brand image) vary considerably across alternatives, resulting in different variances for the error terms of different alternatives. Assumption of identically distributed error terms prevents a treatment of random variations in tastes across decision makers.

The second assumption of response homogeneity maintains homogeneity in responsiveness to attributes of alternatives across decision makers. More specifically, this assumption denies sensitivity or taste variations to an attribute (e.g., brand, price, promotion) due to unobserved characteristics of the decision maker. However, unobserved characteristics of the decision maker clearly do affect responsiveness. For example, each consumer has her own preferences toward product attributes and each consumer responds differently to marketing stimuli. Ignoring the heterogeneity of decision makers can lead to biased parameter estimates and choice probability estimates (Chamberlain 1980). Response heterogeneity can be accommodated by allowing the coefficients to vary across decision makers, or by assigning decision makers to segments and estimating the coefficients for each segment separately.

The third assumption of error variance-covariance homogeneity states that the error variance-covariance structure is identical across decision makers. As with the assumption of identical error terms, the assumption of identical error variance across decision makers is violated if the values of unobserved variables vary considerably across alternatives. This may be the case if decision makers have different abilities to perceive accurately the overall utility offered by the alternative (Swait and Adamowicz

1996). The assumption of identical error covariance across decision makers is not appropriate if the amount of substitutability among alternatives differs across decision makers.

2.4 Overview of Discrete Choice Models

In this section, we give a brief introduction to the choice models that are most frequently used in the marketing literature. For a more extensive review of discrete choice models, see Ben-Akiva and Lerman (1985), Anderson et al. (1992) and Train (2003). For applications of advanced discrete choice models, see Hess (2005) and Arunotayanun (2009).

2.4.1 Multinomial Logit (MNL)

The multinomial logit (MNL) model is the most commonly used discrete choice model in econometrics (McFadden 1973), transportation (Ben-Akiva and Lerman 1985) and marketing (Guadagni and Little 1983). The MNL model can be derived from the random utility model by assuming that the error term is independently and identically distributed (iid) extreme value (McFadden 1973). This distribution is also known as Gumbel distribution and type I extreme value distribution. The iid assumption results in a closed form expression for choice probabilities which does not involve probabilistic error terms. As a result, the model can be easily estimated without the use of simulation.

However, the MNL model suffers from several limitations. One of the major limitations of the MNL model is its Independence from Irrelevant Alternatives (IIA) property which follows from the assumption of independence of error terms. It states that the relative choice probability of any two choice alternatives is independent of any other alternative or its attributes. This implies that the choice probability of a missing alternative transfers to other alternatives in proportion to their popularity. This substitution pattern may be unrealistic in situations where some alternatives are better substitutes for each other than for other alternatives (Greene and Zhang 2003), as famously illustrated by the red bus-blue bus paradox (Debreu 1960). Furthermore, the

MNL model does not accommodate random taste variation and heteroscedasticity of error terms. Due to its popularity, many extensions to the standard MNL model have been developed to overcome its limitations. Some of the most widely used extensions to the MNL model are presented in the following sections 2.4.2–2.4.4.

2.4.2 Nested Logit (NL)

The nested logit (NL) model, introduced by Ben-Akiva (1973), was designed to overcome the restrictive IIA property of the MNL model by relaxing the assumption of independent error terms. In the NL model, alternatives are partitioned into mutually exclusive subsets, or nests, so that similar alternatives lie in the same nest. IIA property holds within each nest but not across nests. For two alternatives in different nests, the ratio of probabilities can depend on other alternatives or their attributes.

The choice is modelled as a hierarchical process, which can be illustrated with a tree diagram. Each branch denotes a subset of alternatives for which IIA holds. Every leaf on each branch denotes an alternative. IIA does not hold across branches. The decision is made sequentially at each level of this decision tree. The final choice probability of the alternative is obtained by multiplying all conditional probabilities along the sequential process. As with the MNL model, the NL model has a closed form expression and can be estimated efficiently without the use of simulation (Train 2003). However, the use of the NL model requires the knowledge of key attributes and their hierarchy.

In the basic NL model each alternative can only belong to one nest. The cross nested logit (CNL) model is an extension of the NL model where each product can belong to multiple nests (Vovsha 1997, Vovsha and Bekhor 1998). It can be used when there is heightened substitution between alternatives A and B, as well as between alternatives A and C without heightened substitution between alternatives B and C (Hess 2005). However, the CNL model is less tractable and has higher computational requirements than the basic NL model.

2.4.3 Multinomial Probit (MNP)

The multinomial probit (MNP) model is the main alternative to the MNL and NL model structures in discrete choice analysis. The model is based on the assumption that the error terms follow a joint normal distribution with zero mean and covariance matrix Ω. The MNP model with a full covariance matrix accommodates heteroscedasticity and any degree of correlation between the single error terms. This makes it possible to model very complex substitutions patterns.

The choice probabilities of the MNP model are written in a form of integral, which does not have a closed form analytical form. Estimation is carried out using numerical integration or simulation. Generally, the MNP model has greater computational requirements and interpretational challenges than the MMNL model. A major restriction of the MNP model is its reliance on the assumptions of a normal distribution. The assumption that error terms are normally distributed with mean zero may be unreasonable in some situations, most notably when theoretical considerations suggest the density of predicted coefficients to be only on one side of zero. The identification of the MNP model can be challenging (Train 2003), especially the choice of an appropriate covariance structure.

2.4.4 Mixed Multinomial Logit (MMNL)

The mixed logit model (MMNL) has been considered to be the most promising of discrete choice models (Hensher and Greene 2003). The error term can be divided into two parts. One contains all correlation and heteroscedasticity, and the other is iid extreme value. The first part can follow any distribution, which makes the MMNL highly flexible. In fact, it can approximate arbitrarily closely any random utility model (McFadden and Train 2000).

The MMNL model obviates the three limitations of the MNL and NL models by allowing for random taste variation, unrestricted substitution patterns and correlation in error terms over time (Train 2003). The MMNL model does not generally have a closed form solution, making the model computationally less tractable. The MMNL model is usually estimated with the use of simulation.

2.4.5 Exogenous Demand (ED)

The exogenous demand (ED) model is the most commonly used demand model to describe consumers’ substitution behaviour in the assortment planning literature (Kök et al. 2009, Hübner 2011). It is not based on the random utility theory, unlike the other models presented in this thesis. Instead, the demand is specified exogenously.

The ED model specifies directly what the demand for each alternative is and what an individual does when her favourite alternative is not available (Kök et al. 2009). It assumes that the consumer first chooses her favourite alternative. If the favourite alternative is not available, with probability δ she chooses a second favourite and with probability $1 - \delta$ she chooses not to purchase. This is repeated until an alternative that is available is chosen or the decision not to purchase is made.

The modelling approach is very flexible, allowing for any substitution structure. Assortment planning literature differentiates between stockout substitution (the favourite alternative is temporarily out-of-stock) and assortment substitution (the favourite alternative is not carried by the particular store). The ED model is able to treat different types of substitution differently. The model is usually estimated with nonlinear optimization techniques. With the ED model, many parameters are needed to describe consumers’ choice behaviour, which can lead to computational challenges.

2.5 Comparison of Models

Table 1 and Table 2 present a comparison of discrete choice models in their limitations and capabilities, respectively. The MNL model and the NL model have several restrictive assumptions about the structure of error terms. They are incapable of capturing response heterogeneity and error variance/covariance heterogeneity in their standard forms. The NL model is less restrictive because the IIA property holds within each nest but not across nests. These models have a closed form expression for choice probabilities, which makes it easier to estimate them.

The MNP, MMNL and ED models have some structural advantages over MNL and NL models. These models relax the restrictive assumptions of the MNL and NL models about the structure of error terms. All of these models obviate the limitations of the MNL and NL models by allowing for random taste variation (differences in tastes that cannot be linked to observed characteristics), unrestricted substitution patterns (no IIA property) and correlation in unobserved factors of utility over time (correlated error terms). However, these models are computationally less tractable and more complex than the MNL and NL models. Generally, they do not have closed form expressions for choice probabilities, and thus the estimation of these models requires simulation or numerical integration.

Table 1: Limitations of discrete choice models

	MNL	NL	MNP	MMNL	ED
Independent error terms	Yes	-	-	-	-
Identical error terms	Yes	Yes	-	-	-
IIA property	Yes	Within nests	-	-	-
Response homogeneity	Yes	Yes	-	-	-
Error variance homogeneity	Yes	Yes	-	-	-
Error covariance homogeneity	Yes	Yes	-	-	-
Restricted to Gumbel distribution	Yes	Yes	-	-	-
Restricted to normal distribution	-	-	Yes	-	-
Numerical estimation	-	-	Yes	Yes	Yes

Table 2: Capabilities of discrete choice models

	MNL	NL	MNP	MMNL	ED
Closed form choice probabilities	Yes	Yes	-	-	-
Correlation in unobserved factors over alternatives	-	Yes	Yes	Yes	Yes
Correlation in unobserved factors over time	-	-	Yes	Yes	Yes
Random taste variation	-	-	Yes	Yes	Yes
Unrestricted substitution patterns	-	-	Yes	Yes	Yes

3 Choice Models in Marketing

3.1 Early Research

McFadden (1973) laid the theoretical foundation for discrete choice theory by developing the random utility model. He showed that under certain assumptions on consumer behaviour and error terms, the random utility theory produces the multinomial logit model that can be easily estimated. More theoretical breakthroughs followed, e.g., nested logit model by Ben-Akiva (1973), Williams (1977) and Daly and Zachary (1978), generalized extreme value model by McFadden (1978), multinomial probit model by Hausman and Wise (1978) and Daganzo (1979), and mixed multinomial logit model by Boyd and Mellman (1980) and Cardell and Dunbar (1980).

The seminal work of Guadagni and Little (1983) showed the power of the MNL model and popularized it in the marketing literature. They showed how scanner panel data can be used to analyse consumers' purchase behaviour. They introduced the concepts of brand and size loyalty, formulated as an exponentially smoothed average of past purchases. Their approach also demonstrated the impact of marketing mix variables such as price, promotion and discount on share.

3.2 Heterogeneity and State Dependence

One of the shortcomings of the MNL model is that it treats consumers as a homogeneous group. Structurally, the MNL model cannot capture heterogeneity across consumers or state dependence of choice probabilities. Heterogeneity represents the differences in consumers' preferences and responses to marketing stimuli whereas state dependence is the influence of observed past purchases on the current choice probabilities. However, they can be incorporated into the MNL model as explanatory variables. These explanatory variables summarize the observed past choice behaviour into a single variable, given as a function of observed past purchases.

Loyalty variables are the most widely used explanatory variables in the marketing literature to incorporate heterogeneity across consumers and purchase-to-purchase dynamics into the choice model. Guadagni and Little (1983) define loyalty as an exponentially smoothed average of a household's past purchases of the given brand or size. The loyalty variables of Guadagni and Little are widely used in the marketing literature (e.g., Lattin 1987, Gupta 1988, Tellis 1988, Kalwani et al. 1990, Chiang 1991). There are also many alternative formulations of loyalty variables. Fader and Lattin (1993) separate the heterogeneity and nonstationarity components of the loyalty measure by using a Dirichlet distribution over purchase occasions. Other researchers define loyalty as the ratio of purchases of a particular brand or size to total purchases (e.g., Tellis 1988, Lattin and Bucklin 1989, Bucklin and Gupta 1992). Fader and Hardie (1996) extend the loyalty variables to several other product attributes, including taste, form and formula.

Many structural approaches to capturing consumer heterogeneity are developed, most importantly the latent class model (Kamakura and Russell 1989, Chintagunta et al. 1991, DeSarbo et al. 1995, Bucklin and Gupta 1992) and the random coefficients formulation of the mixed multinomial logit model (Boyd and Mellman 1980, Cardell and Dunbar 1980). The latent class model assumes that consumers belong to one of several classes, or segments. The model assigns different coefficient values of preferences and choice behaviour to different segments. Consumers are assigned to these segments either deterministically or probabilistically depending on observed choice history, demographics or other covariates. Kamakura and Russell (1989) use observed purchase histories of households to segment them probabilistically to homogeneous segments with different intrinsic brand utilities and sensitivities to price changes. Chintagunta et al. (1991) use a semiparametric random effects specification to estimate the underlying probability distribution across households for each brand by a finite number of support points. The random coefficients formulation of the mixed multinomial logit model exploits the error structure to permit random taste heterogeneity. The model estimates unknown parameters governing the heterogeneity in the population. See Revelt and Train (1998), Revelt and Train (2000) and Sándor and

Wedel (2002) for applications of the MMNL model in marketing and Hensher and Greene (2003) for a review of the state of practice.

State dependency can also be incorporated in the MNL model by introducing lagged purchase indicators (e.g., Jones and Landwehr 1988, Chintagunta 1993, Erdem 1996, Ailawadi and Neslin 1998, Seetharaman et al. 1999, Sun et al. 2003). Positive values of lagged purchase indicators imply habit persistence (or inertia) and negative values imply variety seeking. Habit persistence is the positive effect of past consumption of a product, brand or attribute on the consumer's current propensity to consume it. The existence of habit persistence motivates marketers' employment of promotional schemes such as advertisement and free sampling in the hope that in the long term costs are outweighed by the benefits of getting households hooked and used to consuming certain products. On the other hand, the existence of variety seeking motivates the lengthening of product lines by manufacturers in the hope that households' variety-driven brand switching benefits their franchise (Seetharaman 2004). However, the lagged purchase indicator can only capture the effect of the previous purchase so that the purchases made before the previous purchase do not affect the current choice probability. Allenby and Lenk (1994) propose a model with autocorrelated errors and consumer heterogeneity to capture the carry-over in the utilities from one purchase occasion to another. Roy et al. (1996) and Keane (1997) model state dependence, habit persistence and unobserved heterogeneity in a single framework, aiming to disentangle their impacts. Seetharaman (2004) provides a comparison of random utility models of state dependence in the marketing literature and proposes a model of brand choice where the preference parameters are allowed to vary as a function of past marketing actions.

3.3 Attribute-Based Approaches to Consumer Choice

Attribute-based approaches use additional information on product attributes to achieve a deeper understanding of the preferences of consumers. Fader and Hardie (1996) were among the first to fully acknowledge the importance of attributes in explaining consumers' preferences toward SKUs. Rather than model consumers' preferences toward SKUs, they express preferences for a given SKU as an additive function of preferences for the underlying attributes (e.g., brand, form, formula, size). They show that the attribute-based approach results in a more parsimonious model with managerial insights on consumers' underlying preferences toward product attributes.

Products in a category with many hundreds of SKUs generally contain fewer unique attribute levels (i.e., values for attributes). An attribute-based approach results in a more parsimonious model as only the preferences toward attribute levels need to be estimated instead of preferences toward all SKUs in a category. The number of parameters increases with the number of additional levels of underlying attributes rather than with the number of SKUs. This approach makes it possible to estimate the demand for new products based on consumers' preferences toward attribute levels that already exist in the market.

Attribute-based approaches are frequently used by industry and marketing consultants, e.g., "The Mind of the Customer" by Information Resources, Inc. (Guadagni and Little 2008). However, there are only a few papers employing the attribute-based approach in the marketing literature. Ho and Chong (2003) develop a model based on learning where consumers reinforce chosen and non-chosen options in their patterns of SKU selection. In their model, the required number of parameters does not depend on the number of SKUs in a category or the number of attribute levels. Bell et al. (2005) propose an attribute level model where they recover the SKU-level parameters by calculating using the estimated attribute level parameters. This circumvents the need to estimate the more complex SKU-level model. Chintagunta and Dubé (2005) develop a model that combines household panel data and store data. Household panel data captures the heterogeneity across consumers whereas the store data is abundant and gives the mean effects of marketing activities. Trinh et al. (2009) find that different product attributes and variants appeal to different demographic segments of consumers. The results suggest that segmentation of consumers would prove valuable in attribute-based experiments. Decker and Scholz (2010) propose an attribute-based Poisson regression model which can be applied to POS scanner data, which is automatically recorded in most supermarkets today. They show that their attribute-based model is able to predict the sales of new SKUs very accurately.

At present, an attribute-based approach to consumer choice has been introduced to assortment planning and optimization applications. Optimal assortment models require the correct characterization of demand for each product and the substitution patterns across products (Chong and Ho 2009). However, there is a large number of SKUs in a category and a huge number of substitution patterns. If there are N products in a store, one must estimate $N(N-1)/2$ substitution patterns. The complexity can be substantially reduced if one represents products through their attributes and estimates the demand for SKUs based on their attribute levels. Chong et al. (2001) consider brand level assortment decisions in a joint purchase incidence and brand choice model with a no-purchase option. Fisher and Vaidyanathan (2009) use aggregate sales history of the SKUs currently carried by the retailer to estimate the demand for attribute values. See Kök et al. (2009) for a review of literature and industry practice of assortment planning.

3.4 Joint Modelling of Purchase Decisions

Joint modelling of purchase incidence, brand choice and quantity decisions consider the questions of whether to buy, what to buy and how much to buy within a single framework (e.g., Chiang 1991, Chintagunta 1993). An alternative to the purchase incidence model is the purchase timing model, considering the question of when to buy instead of whether to buy (e.g., Gupta 1988). Joint models of purchase incidence and brand choice are able to capture both the effect of category expansion and brand switching, whereas brand choice models are only able to recognize brand switching and cannibalization as the sole source of new demand. This makes it possible to calculate the sales response to marketing activities, as a marketing campaign is likely to increase total category sales (Guadagni and Little 1998). The inclusion of quantity decision permits more accurate forecasting of sales value and quantity (Krishnamurthi et al. 1992). Additionally, some models consider the store choice decision of consumers in a context where data is gathered from several nearby store locations (e.g., Bell and Lattin 1998).

Table 3: Comparison of research on joint modelling of purchase decisions in single categories

	Purchase Incidence	Brand Choice	Purchase Quantity
Guadagni and Little (1983)	-	Yes	-
Guadagni and Little (1998)	Yes	Yes	-
Krishnamurthi and Raj (1988)	-	Yes	Yes
Gupta (1988)	Yes (timing)	Yes	Yes
Bucklin and Lattin (1991)	Yes	Yes	-
Jain and Vilcassim (1991)	Yes (timing)	-	-
Chiang (1991)	Yes	Yes	Yes
Krishnamurthi et al. (1992)	-	Yes	Yes
Buckling and Gupta (1992)	Yes	Yes	-
Chintagunta (1993)	Yes	Yes	Yes
Dillon and Gupta (1996)	Yes	Yes	-
Mela et al. (1997)	-	Yes	-
Bucklin et al. (1998)	Yes	Yes	Yes
Arora et al. (1998)	-	Yes	Yes
Ailawadi and Neslin (1998)	Yes	Yes	Yes
Mela et al. (1998)	Yes	-	Yes
Bell et al. (1999)	Yes	Yes	-
Jedidi et al. (1999)	-	Yes	Yes
Pauwels et al. (2002)	Yes	Yes	Yes
Allenby et al. (2004)	-	-	Yes
Zhang and Krishnamurthi (2004)	Yes	Yes	Yes
Chib et al. (2004)	Yes	Yes	-

Table 3 provides an overview of research on joint modelling of purchase decisions in single categories. The work of Guadagni and Little (1983) introduced the MNL model in the brand choice application. This research was followed by an application of the nested logit model of purchase incidence and brand choice from the same authors in 1987, first distributed as a working paper and later published (Guadagni and Little 1998). Joint models of purchase incidence, brand choice and quantity decisions followed. Gupta (1988) was first to consider all three decisions simultaneously using a separate statistical model for each decision. Chiang (1991) and Chintagunta (1993) set all three decisions within a single utility maximization problem. Since this early research on joint models of purchase decisions, the emphasis in the marketing literature has been on studying the effects of heterogeneity and marketing response in joint models of purchase incidence, brand choice and purchase quantity decisions. Dillon and Gupta (1996) and Ailawadi and Neslin (1998) study the effect of promotion to category expansion and brand switching. Mela et al. (1997), Mela et al. (1998) and Pauwels et al. (2002) determine the long-term impact of marketing mix variables to choice. Bucklin et al. (1998) determine latent segments of households on the basis of their response to marketing activities simultaneously with their purchase incidence, brand choice and purchase quantity decisions. Jedidi et al. (1999) study the effect of marketing mix variables on long-run profitability. Chib et al. (2004) include the no-purchase option to account for correlation between the no-purchase and brand choice decisions. Zhang and Krishnamurthi (2004) account for consumer heterogeneity to customize marketing actions to households.

The availability of rich market basket data and the recent developments in numerical methods and computing power have made it possible to analyse data on multi-category choice behaviour of consumers. This research focuses on joint modelling of purchase decisions simultaneously across multiple product categories. Traditional single-category choice models are computationally more tractable but they ignore the possible dependencies between consumers' purchase decisions across product categories. Ignoring these dependencies can lead to biased understanding of the determinants of consumer choice. Well-known examples of closely related product categories include complementary products (e.g., toothbrush and toothpaste, printer and ink) and substitutable products (e.g., potato chips, tortilla chips and popcorn). Song and Chintagunta (2007) and Mehta (2007) propose frameworks to simultaneously study purchase incidence and brand choice decisions of households in a multi-category setting. See Seetharaman et al. (2005) for an extensive review of multi-category choice.

4 A Multinomial Logit Choice Model

In this chapter, we create a model to describe the consumers' choice of stock-keeping units (SKUs). We apply two different multinomial logit specifications in modelling consumers' choice behaviour. The first model is the standard SKU-specific model of Guadagni and Little (1983) which is the most widely used brand choice model in the marketing literature. The second model is based on the attribute level-specific approach of Fader and Hardie (1996) which describes SKUs through their attributes, such as brand, size, form, formula and taste. Both models use explanatory variables, such as attribute loyalty, promotion, regular price, discount, previous promotional purchase and previous non-promotional purchase to explain the probabilities of the consumers' in-category SKU choice decisions. We also present the maximum likelihood estimation and the model evaluation methods.

4.1 Indices, Variables, Parameters and Attributes

In this section, we present the full notational convention used in this study. We present our indexing and naming convention, as well as the list of variables arranged in categories.

All indices are presented with lowercase letters. We write consumer index i in superscript to highlight its importance in segregating different consumers from each other. All other indices are written in subscript. The indices are presented in Table 4.

Table 4: Indices

Index	Description
t	shopping occasion
i	consumer
j	alternative
k	attribute
l	level of attribute

Different types of notation are used for different categories of variables. Capital letters are used for components of utility, independent variables and other variables. The Greek alphabet is used for stochastic variables and parameters that are estimated from the data. Components of utility are presented in Table 5, independent variables in Table 6, other variables in Table 7 and parameters in Table 8.

Table 5: Components of utility

Variable	Description
TU_{jt}^{i}	Total utility
U_{jt}^{i}	Deterministic component of utility
ε_{jt}^{i}	Stochastic component of utility
V_{jt}^{i}	Attribute-specific component of deterministic utility
W_{jt}^{i}	SKU-specific component of deterministic utility

Table 6: Independent variables

Variable	Description
LOY_{lkt}^{i}	Attribute loyalty – exponentially weighted average of past purchases treated as binary variables indicating whether or not the household purchased the specific brand, size, form or formula on previous shopping occasions
PPP_{jt}^{i}	Previous promotional purchase – 1 if the consumer's previous purchase was a promotional purchase of attribute level l, 0 otherwise
$PNPP_{jt}^{i}$	Previous non-promotional purchase – 1 if the consumer's previous purchase was a non-promotional purchase of attribute level l, 0 otherwise
$PROM_{jt}^{i}$	Promotion – 1 if the alternative j is promoted at shopping occasion t, 0 otherwise
$PRICE_{jt}^{i}$	Regular price – undiscounted gross price of the alternative j at shopping occasion t denoted as gross price / wholesale price $=$ 1 + mark-up %
$DISC_{jt}^{i}$	Discount – Promotional price decrease of the alternative j at shopping occasion t as a percentage (≥0) of regular price

Table 7: Other variables

Variable	Description
I_{lkj}	Attribute indicator describing which attribute levels are characterized by the alternative j
S_{lkt}^{i}	Share of purchases-variable describing certain attribute level's share of purchases on a shopping occasion
$PREF_{lkt}^{i}$	Attribute preference toward an attribute level
A_{jt}	Available indicator – 1 if the alternative j is available at shopping occasion t, 0 otherwise
$PURCH_{jt}^{i}$	Purchase indicator – 1 if the alternative j is purchased at shopping occasion t, 0 otherwise

Table 8: Parameters

Parameter	Description
α_{0lk}	Attribute level-specific intercept term (FH model only)
α_{1k}	Attribute-specific coefficient of LOY_{lkt}^{i} variable
γ_k	Carry-over constant of attribute k
β_{0j}	SKU-specific intercept term (GL model only)
β_{1j}	SKU-specific coefficient of PPP_{jt} variable
β_{2j}	SKU-specific coefficient of $PNPP_{jt}$ variable
β_{3j}	SKU-specific coefficient of $PROM_{jt}$ variable
β_{4j}	SKU-specific coefficient of $PRICE_{jt}$ variable
β_{5j}	SKU-specific coefficient of $DISC_{jt}$ variable

4.2 Multinomial Logit Model Specification

Total Utility

Consider a consumer i confronted with a choice from the choice set with alternatives $j \in J$ at shopping occasion t. Here the alternatives are SKUs in a certain category. We assume that the consumer has already made a decision to purchase an SKU from the category. The consumer chooses an SKU from the choice set so that her total utility is maximized. The total utility TU_{jt}^{i} is divided into two parts and can be expressed as

$$TU_{jt}^{i} = U_{jt}^{i} + \varepsilon_{jt}^{i}, \tag{2}$$

where

U_{jt}^{i} = deterministic component of utility of alternative j for consumer i at shopping occasion t, and

ε_{jt}^{i} = stochastic component of utility.

The stochastic component of utility varies from one choice occasion to the next. It can be thought to represent the unobserved variables affecting consumer choice or heterogeneity of consumers' preferences. The stochastic components are independently distributed Gumbel stochastic variables. They are characterized by the double exponential distribution

$$P(\varepsilon_j \leq \varepsilon) = e^{-e^{-\varepsilon}}. \tag{3}$$

The distribution has a mean which corresponds to Euler's constant (0.57722). A more general form of Equation (3) would include a location parameter to set the distribution mean to zero and a scale parameter that implies the degree of heterogeneity among the consumers. However, we can omit these parameters without any loss of generality. The location parameter is common to all alternatives and therefore does not have an effect to choice probabilities. Scaling of parameters to obtain optimal model fit is built into the estimation step making a scaling parameter unnecessary.

Deterministic Component of Utility

The deterministic component of utility U_{jt}^i contains all information on observed variables affecting the decision. The observed variables are modelled as attributes of the alternative j. They capture the information on products, consumers and marketing environment. We assume that there are no interactions between the observed variables which lead to an additive utility function. Also, we assume linear relationships between dependent and independent variables. Thus the deterministic component of utility for consumer i for alternative j can be written as

$$U_{jt}^i = \sum_k b_{kj} x_{kjt}^i, \tag{4}$$

where

b_{kj} = utility weight of attribute k of alternative j, and

x_{kjt}^i = observed value of attribute k of alternative j for consumer i at shopping occasion t.

This is a general representation of the deterministic component of utility. Most marketing studies present the deterministic component of utility as a function of SKU-specific intercept terms. This approach implicitly assumes that the consumers' preferences are maintained toward SKUs themselves instead of their attributes. Instead of using SKU as the fundamental unit of analysis, we use the attribute-specific approach of Fader and Hardie (1996) to model consumer preferences over the attributes that describe the SKUs. We still maintain the SKU-specific intercept terms for the marketing mix variables because marketing mix components, such as promotion and discount, are clearly directed at individual SKUs rather than their attributes.

In our approach it is convenient to break the attributes into two classes: attribute-specific attributes and SKU-specific attributes. The attribute-specific component captures consumers' preferences toward attributes of alternatives. The SKU-specific component captures the marketing mix effects on consumer choice. We can therefore express the deterministic component of utility as

$$U_{jt}^{i} = V_{jt}^{i} + W_{jt}^{i}, \tag{5}$$

where

V_{jt}^{i} is the attribute-specific component of utility, and

W_{jt}^{i} is the SKU-specific component of utility.

Attribute-Specific Component of Utility

The attribute-specific component of utility V_{jt}^{i} can be expressed as a linear function of the attributes of the alternative j. These are common to all alternatives. Examples include brand, size, form and formula of an alternative. The attribute-specific component of utility can be written as

$$V_{jt}^{i} = \sum_{k \in K} \sum_{l \in L} I_{lkj} PREF_{lkt}^{i}, \tag{6}$$

where

I_{lkj} = binary variable, which has the value 1 if the alternative j has the l'th level of attribute k, 0 otherwise, and

$PREF_{lkt}^{i}$ = preference of l'th level of attribute k at shopping occasion t for consumer i.

$PREF^i_{lkt}$ consists of attribute-specific intercept terms, loyalty variables, previous promotional purchase variables and previous non-promotional purchase variables. The expression can be written as

$$PREF^i_{lk} = \left[\alpha_{0lk} + \alpha_{1k} LOY^i_{lkt}\right], \tag{7}$$

where

α_{0lk} = attribute level-specific intercept term (parameter, FH model only),

α_{1k} = attribute level-specific coefficient of LOY^i_{lkt} variable (parameter),

LOY^i_{lkt} = attribute level-specific loyalty variable, which is defined as exponentially weighted average of past purchases treated as binary variables indicating whether or not the household purchased an alternative with attribute level l of attribute k,

The parameters α_{0lk} and α_{1k} are estimated from the data.

Attribute Level-Specific Intercept Term

α_{0lk} is the attribute level-specific intercept term which captures the consumers' base preference toward an attribute level. These are additive constants specific to attribute levels. One of the intercept terms is constrained to zero for each attribute k to avoid singularity in the maximum likelihood estimation. The resulting intercept terms describe the uniqueness of an alternative that is not captured by the other explanatory variables.

Attribute Loyalty

LOY^i_{lkt} variable describes the consumers' loyalty toward attributes of alternatives, e.g., brand, size, form and formula. The loyalty variable is defined in a similar fashion to Guadagni and Little (1983) and Fader and Hardie (1996). In addition to their definition of loyalty, we take into account that multiple SKUs in a category can be bought simultaneously on each shopping occasion. The loyalty variable is defined as the exponentially weighted average of past purchases of the attribute level, treated as binary variables indicating whether or not the household purchased an SKU with the attribute on earlier shopping occasions. To emphasize the shopping occasion index t, we write $LOY^i_{lk}(t) = LOY^i_{lkt}$ and $S^i_{lk}(t) = S^i_{lkt}$. The loyalty variable is expressed with

$$LOY_{lk}^{i}(t+1) = \gamma_k LOY_{lk}^{i}(t) + (1-\gamma_k) S_{lk}^{i}(t), \tag{8}$$

where

γ_k = carry-over constant for attribute k, $\gamma_k \in [0,1]$, and

S_{lkt}^{i} = share of purchases on shopping occasion t that have attribute level l of attribute k.

The carry-over constant γ_k determines the share of LOY_{lkt}^{i} variable carried over to the next shopping occasion $t+1$. It is a parameter which is to be estimated for each attribute k separately so that it offers the model the best fit to the data. The sum of loyalties across attributes equals 1.

Multiple SKUs can be bought on each shopping occasion. Therefore, we must introduce variable S_{lkt}^{i} that describes the share of purchases that have attribute level l of attribute k at shopping occasion t. S_{lkt}^{i} is simply defined as the number of SKUs bought that have attribute level l of attribute k divided by the number of total category purchases on the shopping occasion t. We define

$$S_{lkt}^{i} = \frac{\sum_{j \in J} I_{lkj} PURCH_{jt}^{i}}{\sum_{j \in J} PURCH_{jt}^{i}}, \tag{9}$$

where

I_{lkj} = binary variable (see Equation (6)), which has the value 1 if the alternative j has the l'th level of attribute k, 0 otherwise, and

$PURCH_{jt}^{i}$ = purchase indicator.

The purchase indicator $PURCH_{jt}^{i}$ is defined as

$$PURCH_{jt}^{i} = \begin{cases} 1 & \text{if consumer } i \text{ bought alternative } j \text{ at shopping occasion } t, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

To start up the loyalty variable, the initial values for LOY_{lkt}^{i} variable must be defined. Understandably, the loyalty variable cannot be defined without the purchase history.

Therefore the loyalty variable is not defined at the first shopping occasion $t = 1$. We initialize the loyalty variable at the second shopping occasion $t = 2$ so that

$$LOY_{lk}^{i}(2) = \begin{cases} \gamma_k S_{lk}^{i}(1) & \text{if alternative with attribute level } l \text{ of attribute } k \\ & \text{was bought by customer } i \text{ at } t = 1, \\ \dfrac{1-\gamma_k}{N_k - M_k(1)} & \text{otherwise,} \end{cases} \tag{11}$$

where

N_k = total number of attribute levels l associated with attribute k, and

M_{kt} = total number different attribute levels l associated with attribute k bought at shopping occasion t.

This formulation for the initial values of loyalty attributes ensures that the sum across loyalty variables sums to 1 and the initial values quickly approach the long-term averages of loyalty variables.

SKU-Specific Component of Utility

The SKU-specific component of utility contains previous purchase and marketing mix variables. It can be expressed as

$$\begin{aligned} W_{jt}^{i} = \beta_{0j} + \beta_{1j} PPP_{jt}^{i} + \beta_{2j} PNPP_{jt}^{i} + \beta_{3j} PROM_{jt}^{i} + \beta_{4j} PRICE_{jt}^{i} \\ + \beta_{5j} DISC_{jt}^{i}, \end{aligned} \tag{12}$$

where

β_{0j} = SKU-specific intercept term (GL model only),

β_{1j} = SKU-specific coefficient of PPP_{jt}^{i} variable,

PPP_{jt}^{i} = SKU-specific previous promotional purchase variable, which has the value 1 if the consumer's previous purchase at shopping occasion $t-1$ was a promotional purchase of alternative j, 0 otherwise,

β_{2j} = SKU-specific coefficient of $PNPP_{jt}^{i}$ variable,

$PNPP_{jt}^{i}$ = SKU-specific previous non-promotional purchase variable, which has the value 1 if the consumer's previous purchase at shopping occasion $t-1$ was a non-promotional purchase of alternative j, 0 otherwise,

β_{3j} = SKU-specific coefficient of $PROM_{jt}^{i}$,

$PROM_{jt}^{i}$ = SKU-specific promotion variable, which has the value 1 if the alternative j is promoted on the shopping occasion t, 0 otherwise,

β_{4j} = SKU-specific coefficient of $PRICE_{jt}^{i}$,

$PRICE_{jt}^{i}$ = SKU-specific regular price variable representing undiscounted gross price of the alternative j at shopping occasion t,

β_{5j} = SKU-specific coefficient of $DISC_{jt}^{i}$, and

$DISC_{jt}^{i}$ = a non-negative SKU-specific discount variable, which shows the discount of alternative j at the shopping occasion t as a percentage of normal price, 0 otherwise.

All parameters β are estimated from the data.

Previous Promotional and Previous Non-Promotional Purchase

We model the recent choice behaviour of consumers with previous purchase variables PPP_{jt}^{i} and $PNPP_{jt}^{i}$. Previous promotional purchase and previous non-promotional purchase are treated separately because previous research shows that the promotional purchase of a brand decreases the likelihood of a subsequent purchase of the same brand compared with a previous non-promotional purchase (e.g., Shoemaker and Shoaf 1977, Dodson et al. 1978, Jones and Zufryden 1981).

The previous promotional purchase variable PPP_{jt}^{i} describes whether or not the alternative j was on promotion and was bought by consumer i at the previous shopping occasion. We define

$$PPP_j^i(t+1) = \begin{cases} 1 & \text{if one of customer } i\text{'s purchases at shopping occasion } t \text{ is a promotional purchase of alternative } j, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

We use a similar approach to modelling the previous non-promotional purchases. We define previous non-promotional variable $PNPP_{jt}^i$ with

$$PNPP_j^i(t+1) = \begin{cases} 1 & \text{if one of consumer } i\text{'s purchases at shopping occasion } t \text{ is a non-promotional purchase of alternative } j, \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

Promotion

The effect of promotion is determined with $PROM_{jt}$ variable. We define

$$PROM_{jt}^i = \begin{cases} 1 & \text{if alternative } j \text{ was on promotion at the consumer } i\text{'s } t\text{th shopping occasion,} \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

Regular Price

Variable $PRICE_{jt}^i$ captures the effect of regular price to consumer choice. We define

$$PRICE_{jt}^i = \frac{\text{Gross Price}_j}{\text{Wholesale Price}_j} = 1 + \text{Markup-\%}_{jt}^{\;i} \tag{16}$$

Discount

The effect of discount is captured in $DISC_{jt}$ variable. We define

$$DISC_{jt}^i = \text{Discount (as \% of Regular Price)} \tag{17}$$

Model 1: The Guadagni-Little Model (GL)

In the first model, we use the standard SKU-specific approach of Guadagni and Little (1983). It is the most commonly used approach to modelling the consumer's choice of SKU in the marketing literature. Inserting Equations (6), (7) and (12) into the general presentation for the deterministic component of utility in the Equation (5) results in

$$
\begin{aligned}
U_{jt}^{i} = V_{jt}^{i} + W_{jt}^{i} \\
&= \sum_{k\in K}\sum_{l\in L} I_{lkj}\alpha_{1lk}LOY_{lkt}^{i} + \beta_{0j} + \beta_{1j}PPP_{jt}^{i} \\
&+ \beta_{2j}PNPP_{jt}^{i} + \beta_{3j}PROM_{jt}^{i} + \beta_{4j}PRICE_{jt}^{i} \\
&+ \beta_{5j}DISC_{jt}^{i}
\end{aligned}
\tag{18}
$$

This is the final presentation for the deterministic component of utility in the GL model.

Model 2: The Fader-Hardie Model (FH)

The second model uses the attribute-specific approach of Fader and Hardie (1996). The final presentation for the deterministic component of utility in the FH model becomes

$$
\begin{aligned}
U_{jt}^{i} = V_{jt}^{i} + W_{jt}^{i} \\
&= \sum_{k\in K}\sum_{l\in L} I_{lkj}\left(\alpha_{0lk} + \alpha_{1lk}LOY_{lkt}^{i}\right) + \beta_{1j}PPP_{jt}^{i} \\
&+ \beta_{2j}PNPP_{jt}^{i} + \beta_{3j}PROM_{jt}^{i} + \beta_{4j}PRICE_{jt}^{i} \\
&+ \beta_{5j}DISC_{jt}^{i}.
\end{aligned}
\tag{19}
$$

The FH model is similar to the GL model except for the modelling of the intercept terms. Instead of using the SKU-specific intercept terms β_{0j} of the GL model to describe the consumer's preference toward stock-keeping units, we use the attribute level-specific intercept terms α_{0lk} to define the consumer's preference toward attribute levels.

The one-segment FH model is merely a constrained version of the GL model. Therefore, it will not perform better than the GL model in terms of log-likelihood in the training set in the calibration period. However, the FH model represents a more parsimonious approach to choice modelling at the SKU-level with insightful parameter estimates describing consumers' preferences toward attributes.

Purchase Probability

The consumer chooses the alternative with the highest utility from her choice set. The probability that the consumer i chooses alternative j from the choice set with J alternatives at shopping occasion t is

$$p_{jt}^{i} = e^{U_{jt}^{i}} \Big/ \sum_{z \in J} e^{U_{zt}^{i}} \,. \tag{20}$$

The analytic form of choice probabilities has greatly contributed to the popularity of the multinomial logit model.

4.3 Maximum Likelihood Estimation

The model parameters are obtained using the maximum likelihood estimation method. The log-likelihood function is maximized by changing the parameter values to obtain optimal fit of model parameters to the data.

The optimal parameters are obtained by maximizing the likelihood function. The likelihood function is given as

$$\mathcal{L} = \prod_{i \in I} \prod_{j \in J} \prod_{t \in T} p_{jt}^{i} {}^\wedge PURCH_{jt}^{i}. \tag{21}$$

Often it is more convenient to take the logarithm of the likelihood function because the derivative of sum terms is easier to compute than the derivative of product terms. The resulting function is called the log-likelihood function. It attains its maximum value with the same parameter values as the original likelihood function, but has the advantage of being more convenient and efficient to calculate. The log-likelihood function for the multinomial logit model is expressed as

$$LL = \ln \mathcal{L} = \sum_{i \in I} \sum_{j \in J} \sum_{t \in T} PURCH_{jt}^{i} \ \ln\left(p_{jt}^{i}\right), \tag{22}$$

where

p_{jt}^{i} = probability of consumer i choosing alternative j on shopping occasion t, and

$PURCH_{jt}^{i}$ = purchase indicator, which has the value 1 if the consumer i chooses alternative j on shopping occasion t, 0 otherwise.

McFadden (1973) shows that the likelihood function is globally concave under relatively weak conditions. Therefore, if a solution exists, it is unique.

The maximum likelihood estimator is asymptotically normal, asymptotically efficient, and consistent under very general conditions (Ben-Akiva and Lerman 1985). Asymptotic normality property allows us to approximate the distributions of the parameter estimates for large samples with a normal distribution. Asymptotic efficiency implies that, for large samples, the estimator is unbiased and attains the minimum variance for all parameters. Consistency states that as the sample size increases, the parameter estimates converge in probability to the true values of the parameters being estimated.

4.4 Evaluating the Goodness of Fit

4.4.1 Asymptotic Standard Errors and t-Values of the Parameter Estimates

The maximum likelihood estimator (MLE) is asymptotically normal. It means that with large sample sizes the distribution of the maximum likelihood estimator can be approximated with a normal distribution or a t-distribution. Because of the approximate nature of the MLE, usually the more conservative t-distribution is used rather than the normal distribution in assessing the standard errors and the t-values of parameter estimates.

The Hessian is the second derivative of the objective function. If the objective function is the negative log-likelihood function, then the Hessian is the observed Fisher information. In other words, the Fisher information is the inverse of the Hessian of the negative log-likelihood evaluated at the MLE. For k parameters, the Fisher information matrix is defined as

$$F = \begin{bmatrix} -\frac{\partial^2 LL}{\partial \theta_1^2} & -\frac{\partial^2 LL}{\partial \theta_1 \partial \theta_2} & \cdots & -\frac{\partial^2 LL}{\partial \theta_1 \partial \theta_k} \\ -\frac{\partial^2 LL}{\partial \theta_1 \partial \theta_2} & -\frac{\partial^2 LL}{\partial \theta_2^2} & \cdots & -\frac{\partial^2 LL}{\partial \theta_2 \partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial^2 LL}{\partial \theta_1 \partial \theta_k} & -\frac{\partial^2 LL}{\partial \theta_2 \partial \theta_k} & \cdots & -\frac{\partial^2 LL}{\partial \theta_k^2} \end{bmatrix}. \tag{23}$$

The Cramér-Rao bound states that the inverse of the Fisher information matrix gives a lower bound on the variance of any unbiased estimator (Ben-Akiva and Lerman 1985). An unbiased estimator which achieves this bound is said to be efficient. Because the MLE is both unbiased and asymptotically efficient, the variances of parameter estimates for large samples are approximated by the inverse of the Fisher information matrix. The variance-covariance matrix is obtained by inverting the Fisher matrix

$$\begin{bmatrix} Var(\hat{\theta}_1) & Cov(\hat{\theta}_1, \hat{\theta}_2) & \cdots & Cov(\hat{\theta}_1, \hat{\theta}_k) \\ Cov(\hat{\theta}_1, \hat{\theta}_2) & Var(\hat{\theta}_2) & \cdots & Cov(\hat{\theta}_2, \hat{\theta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\theta}_1, \hat{\theta}_k) & Cov(\hat{\theta}_2, \hat{\theta}_k) & \cdots & Var(\hat{\theta}_k) \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 LL}{\partial \theta_1^2} & -\frac{\partial^2 LL}{\partial \theta_1 \partial \theta_2} & \cdots & -\frac{\partial^2 LL}{\partial \theta_1 \partial \theta_k} \\ -\frac{\partial^2 LL}{\partial \theta_1 \partial \theta_2} & -\frac{\partial^2 LL}{\partial \theta_2^2} & \cdots & -\frac{\partial^2 LL}{\partial \theta_2 \partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial^2 LL}{\partial \theta_1 \partial \theta_k} & -\frac{\partial^2 LL}{\partial \theta_2 \partial \theta_k} & \cdots & -\frac{\partial^2 LL}{\partial \theta_k^2} \end{bmatrix}^{-1}. \tag{24}$$

The covariance between the ith and the jth parameters of the MLE is given by the ith row and the jth column of the variance-covariance matrix. The variance of any parameter can be obtained from the diagonal elements of the variance-covariance matrix.

The standard error for a parameter estimate is obtained by taking the positive square root of the variance

$$SE(\hat{\theta}) = \sqrt{Var(\hat{\theta})}. \tag{25}$$

The confidence intervals are obtained using the t-distribution. The confidence interval of the parameter estimate with two-sided significance level α is given as

$$CI = \hat{\theta} \pm t_{\frac{\alpha}{2}} SE(\hat{\theta}), \tag{26}$$

where $t_{\frac{\alpha}{2}}$ follows Student's t-distribution with $n - k - 1$ degrees of freedom. Here n denotes the number of observations in the calibration set and k is the number of parameters. As an example, the two-sided 95% confidence interval corresponds to ± 1.96 times the standard error for a large sample.

The t-statistic for the parameter estimate is obtained with

$$t_{\hat{\theta}} = \frac{\hat{\theta} - \theta_0}{SE(\hat{\theta})}, \tag{27}$$

where θ_0 is a non-random, known constant. In this thesis, we are interested in whether a parameter has an impact on the log-likelihood. Therefore, we test a hypothesis that θ_0 is zero. For a large sample, we can reject this hypothesis with a 5% significance level if the t-statistic is between (-1.96, 1.96). High t-statistic means that the variable is highly significant.

The Hessian matrix can be computed via finite difference approximation. Notice that the Fisher information matrix is symmetric. Therefore we only have to calculate the cells in the upper triangular half of the matrix, increasing the computational efficiency. The approximation for the first order partial derivative is given as

$$f_x(x,y) \approx \frac{f(x+h,y) - f(x-h,y)}{2h}. \tag{28}$$

The second order partial derivatives are approximated with

$$f_{xx}(x,y) \approx \frac{f(x+h,y) - 2f(x,y) + f(x-h,y)}{h^2} \tag{29}$$

and

$$f_{xy}(x,y) \approx \frac{f(x+h,y+h) - f(x+h,y-h) - f(x-h,y+h) + f(x-h,y-h)}{4h^2}. \tag{30}$$

4.4.2 Likelihood Ratio Index

The likelihood ratio index is often used with discrete choice models to measure how well the models fit the data (Train 2003). The statistic measures how well the model performs, with its parameters estimated, compared with a model in which all the parameters are set to zero. The comparison is made for the log-likelihood function. The likelihood ratio index is defined as

$$\rho = 1 - \frac{LL(\hat{\beta})}{LL(0)}, \tag{31}$$

where $LL(\hat{\beta})$ is the value of the log-likelihood function at the estimated parameters and $LL(0)$ is the value of the log-likelihood function with all parameters set to zero. The log-likelihood ratio index has the value zero if the estimated model does no better than the null model and value one if the estimated model is perfectly able to predict each decision maker’s choices. The likelihood ratio index is a good indicator of model performance as it maintains the comparability of the GL model and the FH model. This is because both models have the same specification and log-likelihood for the null model.

4.4.3 Cross-Validation using the Hold-Out Method

Cross validation is used to analyse how well results obtained from one data set generalize to a complementary data set. One of the simplest techniques in achieving this is the hold-out method (e.g., Halkidi and Vazirgiannis 2005, Hamel 2009). In the hold-out method, the data set is usually randomly partitioned into two complementary subsets, a training set and a test set. The model is then calibrated using only the data in the training set. The test set is held out and not looked at during training. The generalization performance of the model calibrated with the training data is then evaluated with the test data.

The downside is that the hold-out method does not use all the available data in the model calibration and the results are highly dependent on the choice of the training set/test set split (Refaeilzadeh et al. 2009). These problems can be addressed by using k-fold cross-validation where the data set is first partitioned into k mutually exclusive and equally sized subsets. The model is then calibrated k times so that within each iteration a different fold of data is held out for testing while the remaining folds are used for training. However, the k-fold cross validation increases the computational burden because the model needs to be calibrated k times.

We use the hold-out method as a means of cross-validation, because the computational burden of using k-fold cross validation would become unmanageable. This entails holding out 1/3 of data for testing while calibrating the model with the remaining 2/3 of data in the training set.

We can test the generalization performance of the calibrated models with the hold-out method by comparing increases in log-likelihood with different model specifications. The test set is reserved only for validation purposes and is not used for model calibration. We calculate the log-likelihood for the test set with the parameters calibrated using the training set. We can then compare the increases in log-likelihood in the training set and the test set to see whether the introduction of new variables increases the model performance in both data sets.

4.4.4 Comparison of Actual and Predicted Share of Purchases

The model is evaluated by tracking the predicted share of purchases against the actual share of purchases in a one-month period. The comparisons are then plotted over time, offering valuable visual representations of the quality of fit. We evaluate the model by investigating how well the model predicts the share of purchases

1) in the *training set* during the *calibration period*,

2) in the *training set* during the *forecasting period*,

3) in the *test set* during the *calibration period*, and

4) in the *test set* during the *forecasting period*.

The realized share of purchases is readily available from the data. Given the null hypothesis that the calibrated model is correct, the actual purchase is binomially distributed (Guadagni and Little 1983). The predicted share of purchases is given as

$$s = \sum p_{jt}^{i} / n \,,$$

and the standard error of predicted share of purchases is expressed as

$$SE(s) = \left[\sum p_i(1 - p_i)\right]^{1/2} /n.$$

We can use the standard error to calculate confidence intervals for the predicted share of purchases. We calculate approximate 95% confidence intervals by assuming that the mean of the prediction is normally distributed. Therefore, 95% confidence intervals are given by ± 1.96 times the standard error of predicted share of purchases, $SE(s)$.

Our model has certain limitations when used for forecasting purposes. We do not aim to forecast the timing of purchases or the identity of the consumer. Thus, we assume perfect information on the timing of purchase and the identity of the consumer. In a realistic forecasting situation, the number and timing of purchases in the future and the identity of the consumer are generally unknown to the retailer. Also, we do not try to forecast the forthcoming marketing actions but rather continue to employ the actual marketing mix actions taken by the retailer. This is not an unrealistic assumption, because the retailer can control its own marketing campaigns.

These assumptions limit our forecasting only on the choice of an SKU. As it is, the model cannot be used in forecasting the sales in the future periods. Rather, the model can be used to gain important knowledge on the consumers' buying behaviour and marketing response. The purpose of Section 6.3 is to assess the model's performance in a future period while it is calibrated with the historical data.

5 Empirical Data

5.1 Structure of the Data

The most important piece of data used in this thesis is the consumer-level point-of-sale (POS) data from the retail sales of health products to healthcare personnel in the 56-month period, January 1, 2007 to July 31, 2011. This order line-level data was automatically collected in the retailer’s enterprise resource planning (ERP) database as the purchases were made. It includes 1.8 million order lines in 140,000 orders for 7,200 unique customers. The POS data set includes the information on customers, orders, order lines, items and suppliers.

The information on past promotional activities, product categorizations and attributes were combined with the POS data set. The promotional marketing brochures sent to the customers were made available for this thesis, which made it possible to investigate the effect of promotion to consumers’ choice behaviour. Information on the campaign periods and promoted products were retrieved from these files. The final categorizations for products were created together with product managers who were responsible for the marketing and sales of these products.

All of our data was compiled into an Access database. The structure of the data is presented in Figure 1. Practically all modern cash registers and terminals record point of sale (POS) data that contains similar information on orders, order lines and items sold. These POS terminals are usually integrated in a back-office POS system that may have additional functionality such as inventory control, pricing, information exchange between terminals and planning of promotional campaigns. Information on suppliers and categories can be combined with the sales transactions data quite easily, but not all POS systems are able to identify customers or record non-discount promotions. Some industries traditionally collect very specific data on customers (e.g., healthcare, hotels, banking, car retailing). The ability to collect disaggregate data automatically and efficiently has recently become available to other industries (e.g., retail trade) through customer loyalty programmes and online retailing. The impact of promotional activities to consumers' choice behaviour can be evaluated by combining POS and marketing data together, as we have done, even though POS systems do not automatically collect historical records of promotional activities.

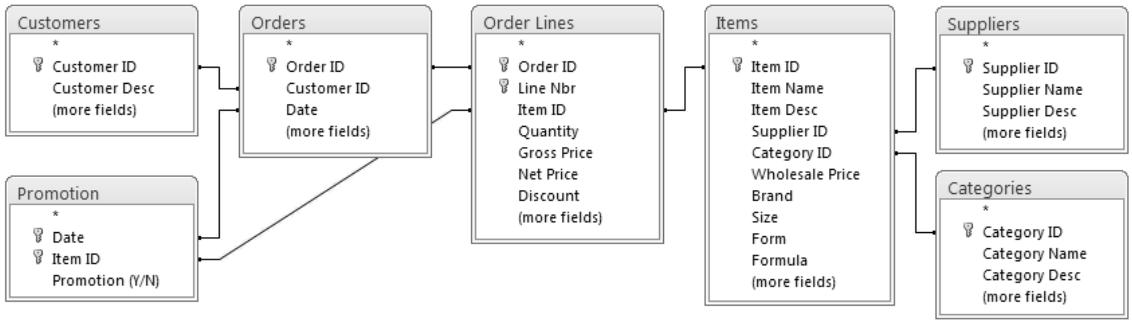

Figure 1: Data structure

5.2 Data Periods

We divided the data into three periods for the purposes of initialization, calibration and forecasting. The purchases for the first 12 months (January 2007 to December 2007) were used for the initialization of the loyalty variables. The next 24 months (January 2008 to December 2009) were used for the calibration of the model. The last 19 months (January 2010 to July 2011) were used for out-of-sample forecasting. The data periods are illustrated in Figure 2.

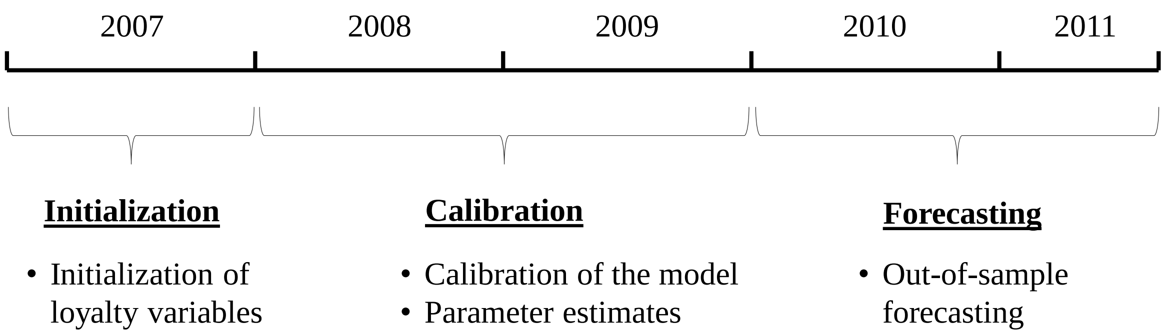

Figure 2: Data periods

5.3 Consumers

In this thesis, we analyse data from the retail sales of health product to healthcare personnel. These personnel customers work in pharmacies, hospitals, health shops, veterinarians, convenience stores and healthcare shops. Each workplace has a designated contact person who makes the order on behalf of the other employees. The order is made on an online service to which only the contact person has access. The contact person collects the money in the workplace from the other employees that have participated in the order in exchange for the purchased goods. Due to the ordering procedure, every consumer whose purchases we analyse in this thesis comprises of a group of individuals.

Loyalty variables are based on each consumer's historical record of purchases. Therefore, we require that a customer has had at least one purchase in the chosen category during the initialization period. Only consumers that have made a purchase in the category during the initialization period are included in the data set, leaving out newly acquired customers and those who did not make a purchase in the chosen category in 2007. This screening method is repeated for all categories separately.

For model evaluation, consumers were divided into two groups. Two thirds of the consumers were randomly placed in the training set and the remaining consumers were assigned to the test set. The purpose of this is to have an independent set of observations which can be used in out-of-sample evaluation of the model in the calibration period and the forecasting period (see Section 4.4.3).

5.4 Categories

The products were categorized so that all products in a category were more or less substitutes for each other. Some categories were selected for further analysis based on total category sales, number of purchase occasions, number of unique SKUs, clear category boundaries and existence of a house brand in the category. Categories with high total sales were preferred because they are more important to the retailer than the smaller categories. A larger number of purchase occasions in a category offered more observations for the model calibration. We required more than ten unique SKUs in each category to ensure that attribute preferences describe consumers' preferences truthfully. Categories with clear boundaries were preferred so that all of the consumers' realistic choice alternatives could be included and unlikely substitutes would not be placed in the same category. An existence of a house brand in the category ensured that the results of this thesis could be applied in optimizing the marketing mix decisions of the retailer's own products.

Six categories were selected for further studies using the category selection criteria. The chosen categories include hand sanitizer, magnesium, multivitamin, omega-3, probiotic and vitamin D. The number of observations for each period and category is presented in Table 9. Notice that the frequency of observations decreases in time. This is because our screening method insists that the consumers included in the data set have made a purchase from the category in 2007. Therefore, all new consumers that have joined afterwards are excluded from the data set.

Table 9: Number of observations included in the data set

	Hand Sanitizer	Magnesium	Multi-vitamin	Omega-3	Probiotic	Vitamin D
Initialization						
Order lines	9,850	5,868	21,550	17,014	7,430	3,019
Orders	7,142	4,980	12,115	11,483	5,460	2,633
Consumers	3,579	2,792	4,666	4,428	2,608	1,788
Calibration						
Order lines	9,982	6,083	25,281	22,170	10,519	4,499
Orders	7,089	4,934	13,719	13,546	7,087	3,391
Consumers	2,391	1,668	3,336	3,191	1,809	1,117
Forecasting						
Order lines	2,231	2,514	8,159	8,085	5,740	3,469
Orders	1,799	2,033	4,930	5,284	3,590	2,300
Consumers	981	892	1,690	1,671	1,133	809

We assume that each purchase decision is independent of other decisions. Thus, each order line is considered an independent purchase decision. Table 10 shows that two thirds of order lines are located into the training set and the remaining into the test set.

Table 10: Number of purchase incidences in the training set and the test set

	Hand Sanitizer	Magnesium	Multi-vitamin	Omega-3	Probiotic	Vitamin D
Initialization						
Training set	6,598	3,885	14,183	11,308	4,804	2,034
Test set	3,252	1,983	7,367	5,706	2,626	985
Calibration						
Training set	6,660	4,033	16,744	14,772	6,853	3,021
Test set	3,322	2,050	8,537	7,398	3,666	1,478
Forecasting						
Training set	1,469	1,617	5,345	5,321	3,727	2,288
Test set	762	897	2,814	2,764	2,013	1,181

Figure 3 shows the number of unique SKUs in the assortment during each promotion period. The number of SKUs in vitamin D and probiotic categories has been growing due to the increasing popularity of these categories. The number of SKUs in the omega-3 category has decreased sharply in early 2009 as a large manufacturer moved to another distributor. The other categories have remained quite stable during the whole observation period.

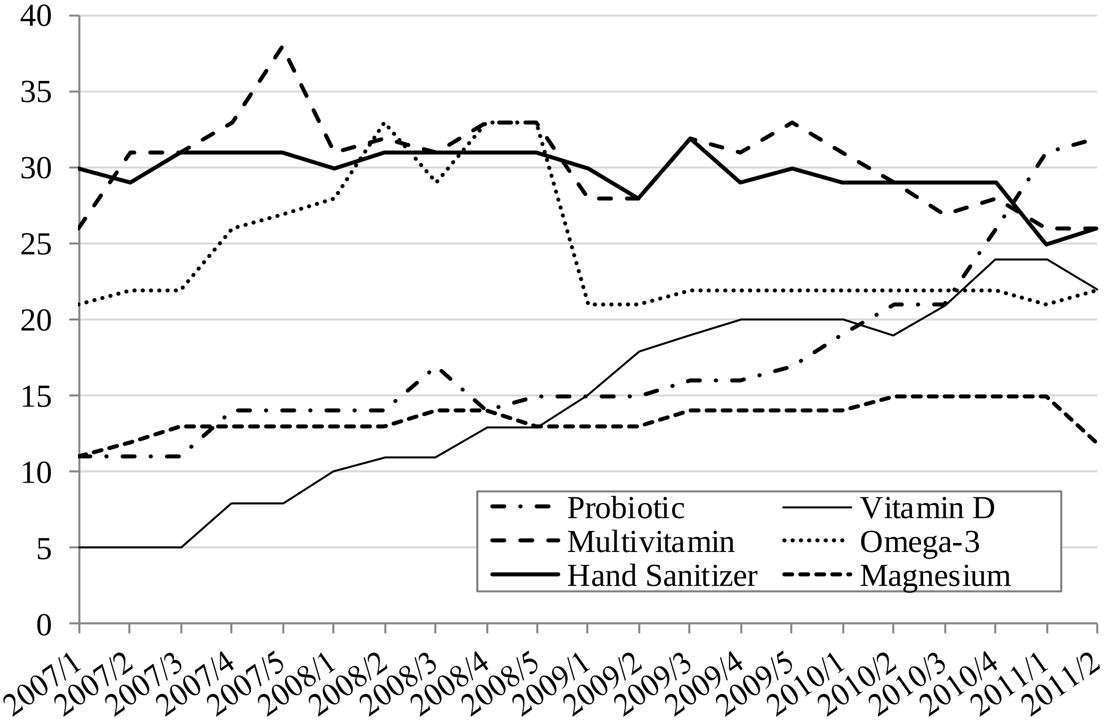

Figure 3: Number of unique SKUs in assortment

5.5 Attributes of Alternatives

Attributes for the chosen categories were selected so that they carried all relevant information on the products. Three important criteria are typically used in determining attributes of SKUs (Fader and Hardie 1996). First, the attribute must be consumer recognizable, meaning that the attribute must be easily observable by examining the package or the product information. Second, the attribute should be objective. There should be no ambiguity or individual differences in observing the precise attribute level for each SKU attribute. Third, the attribute must be collectively exhaustive, meaning that every attribute must apply to every SKU in the category. The objective is to define a minimal number of distinct attributes that capture all relevant product characteristics within the category.

The attribute levels associated with each category are presented in Table 11. Some attributes, such as brand and package size, were used across all categories whereas other attributes, such as form and formula, were meaningful only for some categories. Only two attributes (brand and size) were needed to characterize the hand sanitizer category whereas all four attributes (brand, size, form and formula) were employed to characterize multivitamin and omega-3 products. Product differentiation and large number of unique SKUs in a category require more attributes as the complexity of the category increases.

SKU-specific attribute levels for probiotics are presented in Table 12. Three different attributes are used to characterize probiotics including brand, size and form. However, one can note that the item descriptions include additional attributes such as taste and formula. These attributes could be incorporated in the model, but in this case they did not produce an improvement in the model fit.

Table 11: Attribute levels for categories

	Hand Sanitizer	Magnesium	Multivitamin	Omega-3	Probiotic	Vitamin D
Brand	Brand 1 Brand 2 Brand 3 Brand 4 Brand 5 Brand 6 Brand 7 Brand 8	Brand 1 Brand 2 Brand 3 Brand 4 Brand 5 Brand 6 Brand 7	Brand 1 Brand 2 Brand 3 Brand 4 Brand 5 Brand 6 Brand 7	Brand 1 Brand 2 Brand 3 Brand 4 Brand 5 Brand 6 Brand 7 Brand 8	Brand 1 Brand 2 Brand 3 Brand 4	Brand 1 Brand 2 Brand 3 Brand 4 Brand 5
Size	Extra Large Large Medium Small	Large Medium Small	Large Medium Small Extra Small	Large Medium Small	Extra Large Large Medium Small	Large Medium
Form	-	Chewable Effervescent Granular Liquid Powder Swallowed	Chewable Effervescent Liquid Swallowed	Capsule Liquid	Capsule Chewable Drops Powder	-
Formula	-	-	Added Omega-3 For Adults For Children For Pregnancy For Women Regular	Antioxidant Calcium Children E-EPA Flavoured (14 more)	-	5 μg 7.5 μg 10 μg 20 μg 25 μg 50 μg

Table 12: SKU-specific attribute levels for probiotics

#	Brand	Size	Form	Description
1	Brand 1	Small	Capsule	Brand 1, 20 Capsules
2	Brand 1	Small	Capsule	Brand 1, Formula A, 20 Capsules
3	Brand 1	Small	Capsule	Brand 1, Formula B, 30 Capsules
4	Brand 1	Medium	Capsule	Brand 1, 50 Capsules
5	Brand 1	Medium	Chewable	Brand 1, 30 Chewable Tables
6	Brand 1	Large	Capsule	Brand 1, 100 Capsules
7	Brand 1	Large	Drops	Brand 1, Drops, 7 ml
8	Brand 2	Small	Capsule	Brand 2, 20 Capsules
9	Brand 2	Small	Chewable	Brand 2, Raspberry, 15 Chewable Tablets
10	Brand 2	Medium	Capsule	Brand 2, 50 Capsules
11	Brand 2	Medium	Capsule	Brand 2, Formula C, 50 Capsules
12	Brand 2	Medium	Chewable	Brand 2, Raspberry, 30 Chewable Tablets
13	Brand 2	Medium	Powder	Brand 2, Orange, 7 Powder Bags
14	Brand 2	Large	Chewable	Brand 2, Raspberry, 60 Chewable Tablets
15	Brand 2	Large	Drops	Brand 2, Drops, Added Vitamin D, 7.5 ml
16	Brand 2	Large	Drops	Brand 2, Drops, 7.5 ml
17	Brand 3	Small	Capsule	Brand 3, 20 Capsules
18	Brand 3	Small	Capsule	Brand 3, 30 Capsules
19	Brand 3	Small	Chewable	Brand 3, 10 Chewable Tablets
20	Brand 3	Small	Chewable	Brand 3, 20 Chewable Tablets
21	Brand 3	Medium	Chewable	Brand 3, 30 Chewable Tablets
22	Brand 3	Medium	Chewable	Brand 3, 40 Chewable Tablets
23	Brand 3	Medium	Chewable	Brand 3, Formula D, 30 Chewable Tablets
24	Brand 3	Medium	Drops	Brand 3, Drops, 7 ml
25	Brand 3	Large	Capsule	Brand 3, 100 Capsules
26	Brand 3	Large	Capsule	Brand 3, 100 Capsules
27	Brand 3	Large	Chewable	Brand 3, 60 Chewable Tablets
28	Brand 4	Small	Chewable	Brand 4, Lemon, 10 Chewable Tablets
29	Brand 4	Medium	Chewable	Brand 4, Strawberry, 30 Chewable Tablets
30	Brand 4	Medium	Chewable	Brand 4, Lemon, 30 Chewable Tablets
31	Brand 4	Large	Chewable	Brand 4, Strawberry, 60 Chewable Tablets
32	Brand 4	Large	Chewable	Brand 4, Lemon, 60 Chewable Tablets
33	Brand 4	Large	Drops	Brand 4, Drops, 10 ml
34	Brand 4	Large	Drops	Brand 4, Drops, Added Vitamin D, 10 ml
35	Brand 4	Large	Drops	Brand 4, Drops, 10 ml
36	Brand 4	Extra Large	Chewable	Brand 4, Strawberry, 90 Chewable Tablets
37	Brand 4	Extra Large	Chewable	Brand 4, Strawberry, 90 Chewable Tablets

The number of unique attribute levels for each category is presented in Table 13. For all categories, the total number of attribute levels used in describing SKUs in a category is equal to or smaller than the number of unique SKUs. Therefore, attribute-based modelling provides a parsimonious approach to describing product categories.

Table 13: Number of attribute levels and SKUs

	Hand Sanitizer	Magnesi-um	Multi-vitamin	Omega-3	Probiotic	Vitamin D
# of Brands	8	7	8	8	4	5
# of Sizes	4	3	4	3	4	2
# of Forms	-	6	4	2	4	-
# of Formulas	-	-	6	19	-	6
# of Prices	-	2	-	-	-	-
Total Attribute Levels	12	18	22	32	12	13
# of Unique SKUs	43	18	54	41	37	25

5.6 Marketing Mix

The products are promoted in a marketing brochure which is published four to five times a year. These brochures are printed and sent to all customers who have made an order during the past year. They contain product and price information for chosen health products accompanied by discounts. Seasonal and non-health products are included to diversify the product mix and make shopping more interesting.

Table 14 shows the number of products promoted, the number of products discounted and the number of products simultaneously promoted and discounted during the promotion periods. The product is considered discounted when the average discount percentage for the product is greater than equal to 1 % and less than 100 % during the time period. Most of the products that are discounted are also promoted to maximize the sales impact. However, discounted products are not always promoted, because some of the discounted products are only sold in the outlet. Products sold in the outlet are generally small batches of unsold stock and items close to their expiration date, so there is no point in promoting these products in the marketing brochure.

Table 14: Promotion campaign data

Campaign	Promotion Period	Promoted	Discounted	Promoted & Discounted
2007/1	Jan 12, 2007 – Feb 23, 2007	90	134	44
2007/2	Feb 26, 2007 – Apr 20, 2007	319	215	50
2007/3	Apr 30, 2007 – Jul 13, 2007	401	218	116
2007/4	Aug 13, 2007 – Oct 12, 2007	332	130	65
2007/5	Oct 22, 2007 – Dec 21, 2007	373	149	96
2008/1	Jan 12, 2008 – Feb 15, 2008	125	80	47
2008/2	Feb 18, 2008 – Apr 18, 2008	325	184	59
2008/3	Apr 28, 2008 – Jul 11, 2008	438	237	149
2008/4	Aug 11, 2008 – Oct 10, 2008	330	107	73
2008/5	Oct 27, 2008 – Dec 19, 2008	343	174	132
2009/1	Jan 12, 2009 – Feb 20, 2009	130	51	32
2009/2	Mar 2, 2009 – Apr 24, 2009	199	121	79
2009/3	Apr 27, 2009 – Jul 24, 2009	290	221	117
2009/4	Aug 3, 2009 – Sep 25, 2009	290	220	93
2009/5	Oct 5, 2009 – Dec 18, 2009	250	203	98
2010/1	Jan 18, 2010 – Mar 28, 2010	216	62	45
2010/2	Apr 26, 2010 – Jul 4, 2010	323	139	83
2010/3	Aug 2, 2010 – Sep 30, 2010	313	118	76
2010/4	Nov 1, 2010 – Dec 31, 2010	317	92	88
2011/1	Feb 15, 2011 – Apr 15, 2011	190	78	72
2011/2	Apr 25, 2011 – Jul 4, 2011	259	91	65

The initial data suggests that total sales are heavily dependent on promotion campaigns. Figure 4 shows how the consumers postpone their orders over the non-promotion periods until new promotion campaign is launched. However, we note that the non-promotional periods are located in holiday periods, e.g., summer holiday and Christmas vacation. Therefore, it is too early to draw far-reaching conclusions on the impact of promotional campaigns on sales based on the data in Figure 4 alone. We continue with analysis on different marketing actions during the observation period.

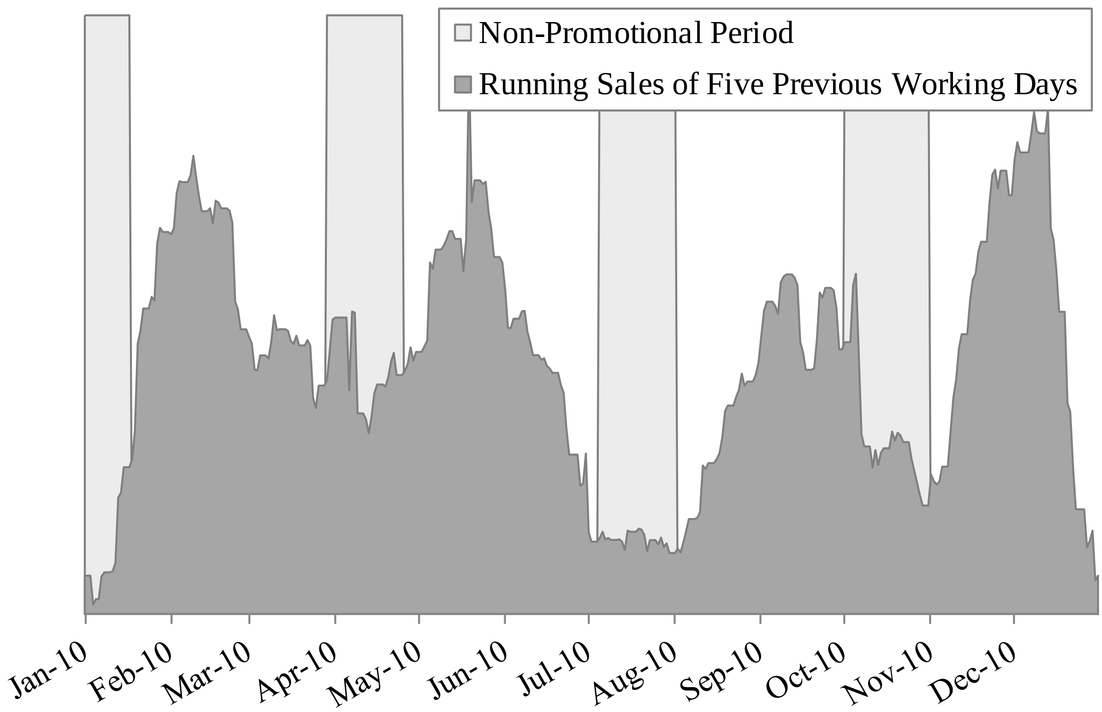

Figure 4: Running sales of five previous working days in 2010

Promotion

The promotion data is further analysed to illustrate how many unique SKUs were promoted during each promotion period. Table 15 shows the number of unique SKUs promoted by category in each promotion period. Omega-3 is the most heavily promoted category in terms of number of products promoted. Probiotic became heavily promoted only after the calibration period. Magnesium, multivitamin and vitamin D are the least promoted categories, mainly due to the lack of a strong house brand in these categories. Non-house brands are only infrequently promoted when the principal buys advertisement space from the marketing brochure. The retailer heavily promotes its own house brands because they are a major driver of profitability. Therefore, the same house brands are promoted in almost every marketing brochure. This may lead to challenges in modelling because we may be unable to determine whether the consumer is truly loyal to the brand or sensitive to promotions. High correlation between loyalty parameters and promotion parameters would imply that this effect is strong.

Table 15: Unique SKUs promoted during the promotion period

Campaign	Hand Sanitizer	Magnesi-um	Multi-vitamin	Omega-3	Probiotic	Vitamin D
2007/1	2	1	0	0	0	0
2007/2	2	2	0	5	0	2
2007/3	8	2	0	5	1	0
2007/4	2	1	1	5	4	3
2007/5	6	2	7	14	3	4
2008/1	2	0	6	2	3	0
2008/2	4	3	7	14	3	2
2008/3	6	3	0	10	3	0
2008/4	2	1	0	11	3	3
2008/5	4	0	1	7	3	0
2009/1	2	0	1	4	3	4
2009/2	2	2	0	12	5	2
2009/3	9	0	1	3	3	0
2009/4	4	0	4	4	4	2
2009/5	7	1	2	8	4	0
2010/1	8	0	0	3	12	2
2010/2	4	2	0	5	3	0
2010/3	3	2	2	6	4	2
2010/4	2	0	0	5	5	1
2011/1	2	0	2	7	4	1
2011/2	4	0	2	7	5	1
Average	4.0	1.0	1.7	6.5	3.6	1.4

Discount

The number of SKUs discounted on each promotion period is presented in Table 16. We consider that the product is discounted if the average discount percentage for the product is greater than or equal to 1 % and less than 100 % during the time period. Discount has been used similarly with promotion. Omega-3 is the most often discounted category. Other categories, most notably probiotic and vitamin D, rely only on a few observations on discounts during the calibration period. This suggests that the parameter estimates for discount variables may not be accurate for all categories.

Table 16: Unique SKUs discounted during the promotion period

Campaign	Hand Sanitizer	Magnesi-um	Multi-vitamin	Omega-3	Probiotic	Vitamin D
2007/1	3	1	0	1	0	0
2007/2	0	0	0	0	0	0
2007/3	3	0	0	1	0	0
2007/4	0	0	0	0	0	0
2007/5	2	0	0	5	0	0
2008/1	0	0	0	1	0	0
2008/2	2	1	0	10	0	0
2008/3	2	1	0	10	0	0
2008/4	2	1	1	13	0	3
2008/5	2	0	2	8	0	1
2009/1	0	0	2	2	0	0
2009/2	0	2	0	12	4	0
2009/3	7	0	0	0	0	0
2009/4	0	0	4	1	0	0
2009/5	5	0	0	1	0	0
2010/1	5	0	0	1	6	1
2010/2	2	2	0	4	2	0
2010/3	1	0	2	4	3	2
2010/4	0	0	1	5	4	1
2011/1	0	0	2	7	4	1
2011/2	0	0	2	7	5	1
Average	1.7	0.4	0.8	4.4	1.3	0.5

We also investigate the magnitude of discounts in each category. Discount is given as

$$\text{Discount} = 1 - \frac{\text{Net Price}}{\text{Gross Price}}. \tag{32}$$

We calculate average discount for each category by taking the average of this metric across all category purchases. The average discounts for product categories are presented in Figure 5. The results confirm that omega-3 is the most heavily discounted product category. Based on this preliminary analysis, we expect omega-3 category to provide the most significant estimates on the impact of promotion and discount on choice behaviour.

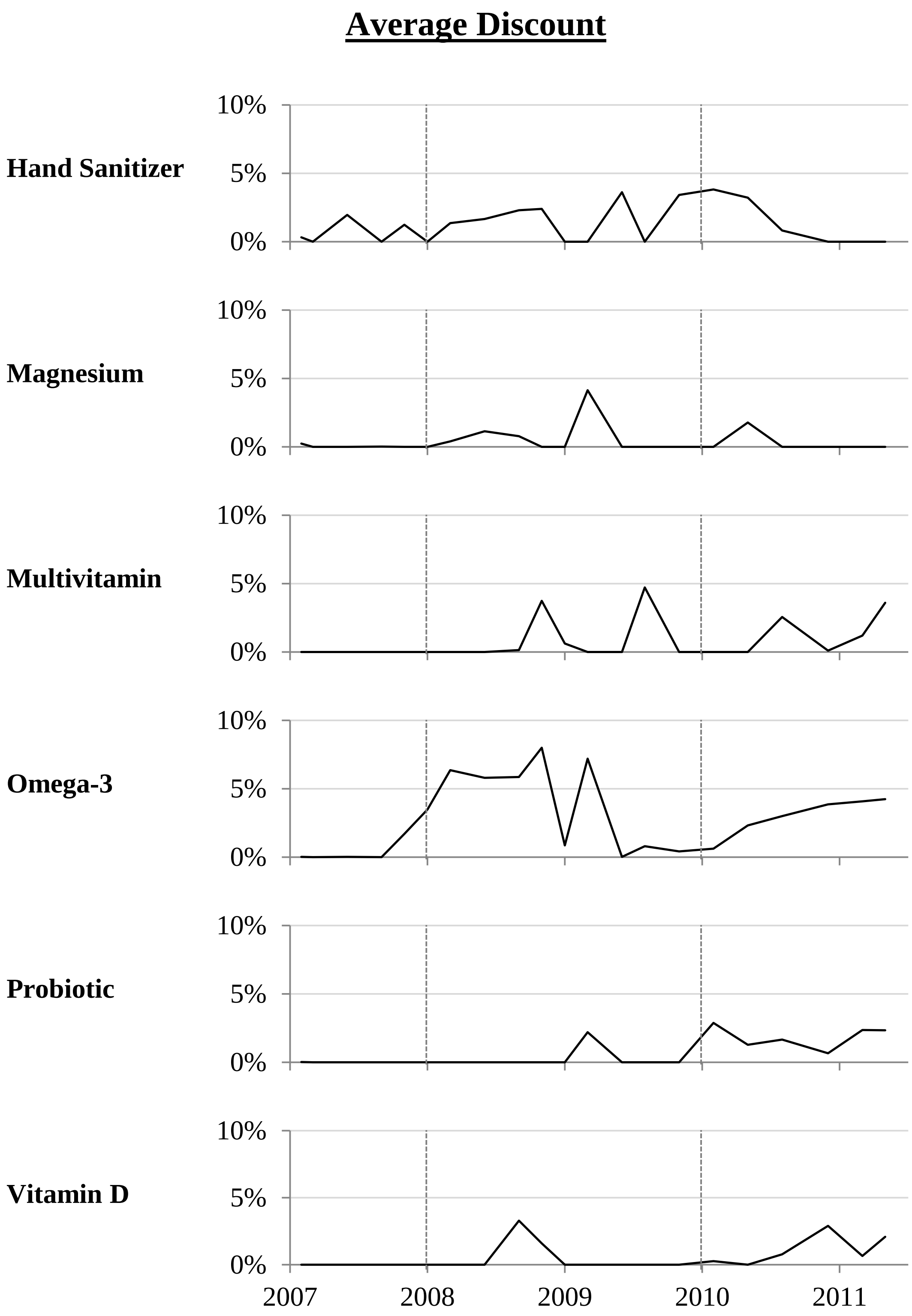

Figure 5: Average discounts for product categories

Regular Price

We examine the impact of price level and regular prices on choice behaviour by studying the mark-ups for different SKUs. Mark-up is the difference between the gross price and the wholesale price of an SKU. The mark-up percentages have been very stable during the whole observation period. This is because the retailer generally makes a separate contract with each manufacturer on selling its products to the healthcare personnel. The mark-up percentage used in selling the manufacturer's products is agreed on and written in these contracts. This leads to very inflexible pricing because the retailer cannot make changes to prices without making a new contract with the manufacturer. As an exception, the retailer can fully determine the prices for its own house brands (products that it sells and markets itself), representing some 40% of personnel sales. However, the prices of house brands are also very rigid because of the difficulty of managing the prices in the retailer's ERP system.

The wholesale prices generally change each year to adjust to inflation, but these changes are not captured in our model. This is because the main reason for the customers to buy from the retailer is the lower prices compared to a pharmacy or a health store. When the wholesale price of a product is increased, the pharmacies and the health stores also increase their retail prices aiming to keep the gross margin percentage constant. The consumer's utility of buying the desired product from the retailer's personnel sales channel compared to a pharmacy or a health store should therefore remain the same. However, the increase in the retail price of a product should lower the utility of buying the product relative to the utilities of its substitutes. This effect is not captured in our model. Anyhow, the impact should be small considering the magnitude of the increases in wholesale prices (only 2-4% a year).

6 Results and Analysis

6.1 Parameter Estimates

In Section 6.1.1, we examine attribute preference parameters obtained with the FH model to understand the attractiveness of attributes. However, the one-segment FH model is only a constrained version of the GL model. Therefore, the GL model arrives at more accurate parameter estimates on attribute loyalties, previous purchases and marketing mix variables. Sections 6.1.2–6.1.4 present parameter estimates on attribute loyalties, previous purchases and marketing mix variables obtained with the GL model.

6.1.1 Attribute Preference

The FH model allows us to characterize the SKUs in a category through their attributes (e.g., brand, size, form, formula). The parameter estimates can be interpreted as the consumers' preference toward the attribute levels of products. Results are analysed to gain deeper understanding of the consumers' preferences in each product category.

Hand Sanitizer

Hand sanitizers are used for hand disinfection as an alternative to washing hands with soap and water. All products in this category are gel based preparations. The minor differences (e.g., skin-softening ingredients) are captured in the brand as all products with the same brand name share the same product characteristics. Only two attributes, brand and size, were needed to explain consumers' choice behaviour in this category.

Consumers' preferences toward brand and size are presented in Figure 6. We notice that consumers prefer small (20ml-100ml) and medium (250ml-500ml) package sizes over larger ones, which are mainly intended for professional use.

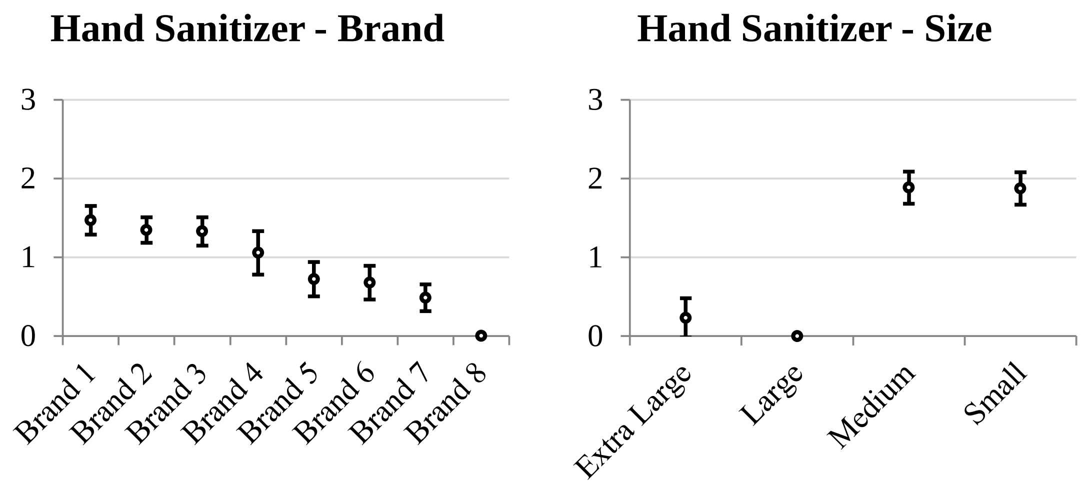

Figure 6: Attribute preferences for hand sanitizer with 95% confidence intervals

Multivitamin

Multivitamins are products intended to supplement a human diet with essential vitamins and minerals. They usually come in the form of chewable, swallowed or effervescent tablets and liquid. Two brand names dominate the Finnish market with several different formulas targeted for different consumer segments. There are many package sizes available from 20 to 250 daily doses.

Figure 7 presents the consumers' preferences toward attribute levels of multivitamins. Consumers prefer large package sizes probably because they deliver better value for money. Tablets intended to be swallowed are the least preferred form even though most of the products in the market are intended to be swallowed.

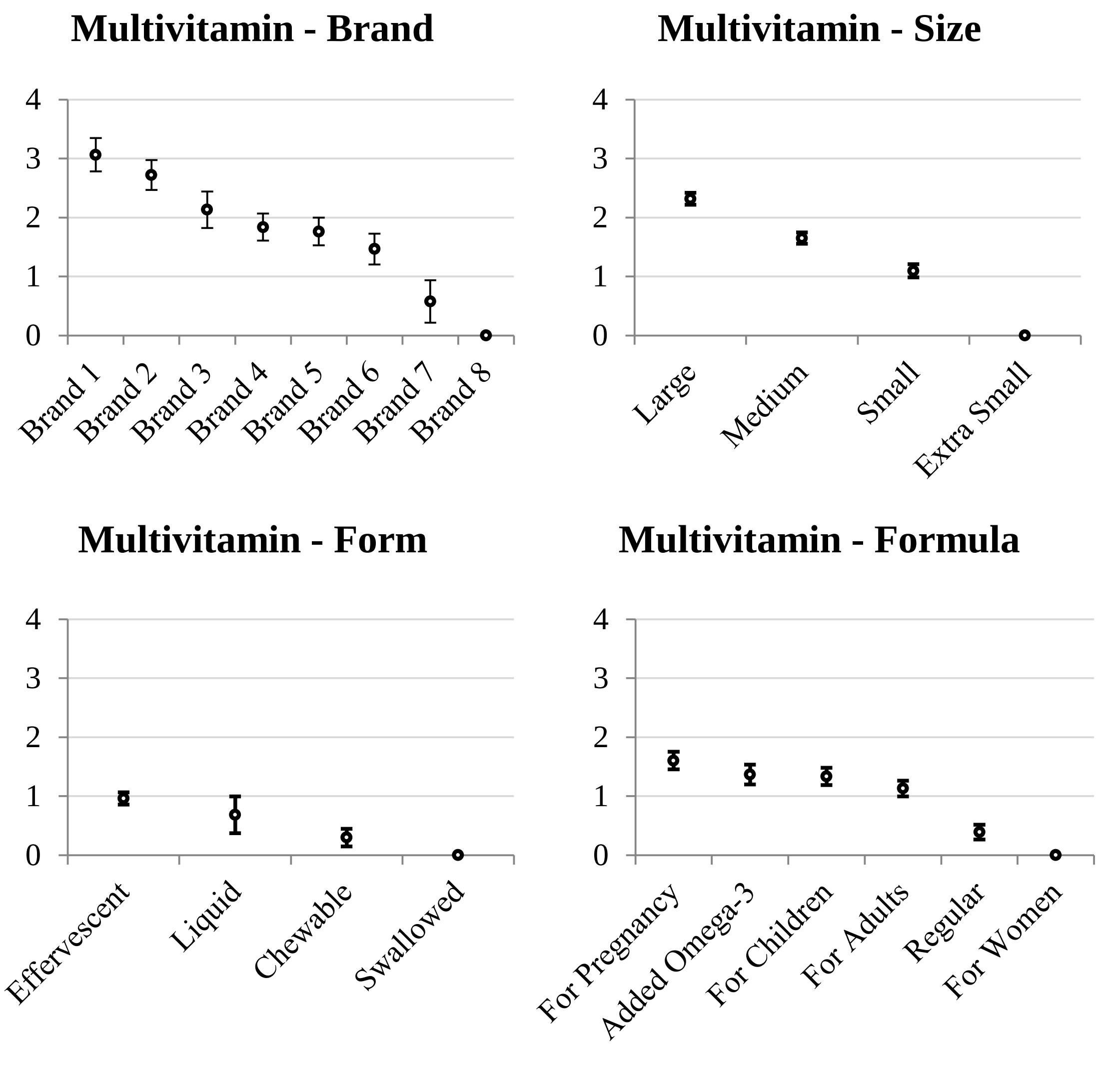

Figure 7: Attribute preferences for multivitamin with 95% confidence intervals

Omega-3

Omega-3 is an unsaturated fatty acid essential to the human body. Its health benefits include the reduced risk of heart disease and improved brain function. Omega-3 must be consumed through food because the body cannot produce it. Omega-3 can be found in fish, other seafood and nut oils.

Omega-3 supplements are found in capsules and liquid. Liquid products offer high concentrations of omega-3 with lower cost than capsules. However, most people do not like the taste of omega-3 fatty acids. Capsules are a tasteless alternative to enjoying the daily dosage of omega-3. The products in the market are heavily differentiated with many competing formulas. Figure 8 shows the calibration results for omega-3.

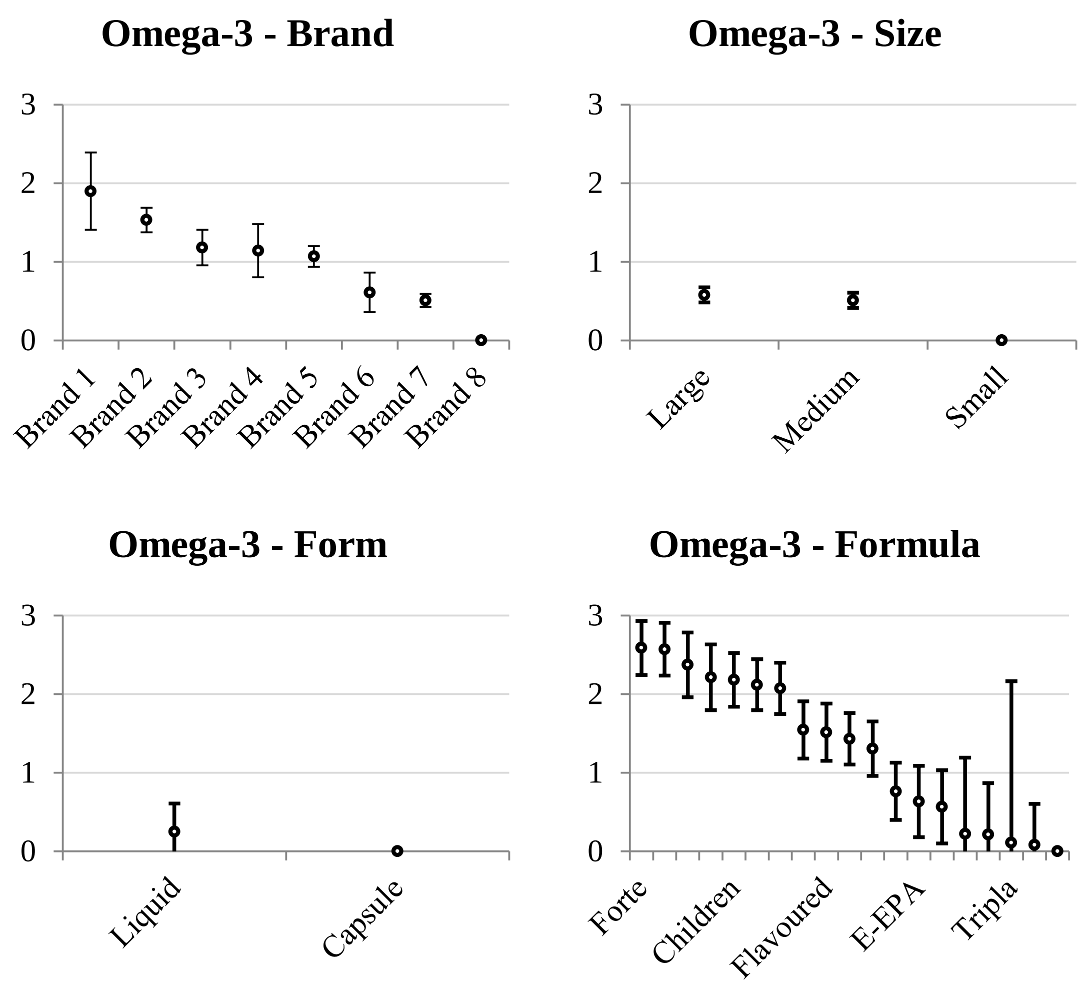

Figure 8: Attribute preferences for omega-3 with 95% confidence intervals

Probiotic

Probiotics are beneficial bacteria that help balance the intestinal flora and fight disease. As a supplement, they are used to treat diarrhea and boost the immune system. Probiotics have gained popularity in the past few years, mainly because of increased focus on the need to keep the digestive system healthy.

Products are available in capsule, chewable tablet, powder and drop. There are four major brands in the market. All brands offer different forms to suit the preferences of all consumers. The consumers' preferences toward attribute levels are illustrated in Figure 9. Large package sizes are favoured over smaller ones. Capsules are the most preferred form even though chewable tablets and drops are increasing their popularity very fast.

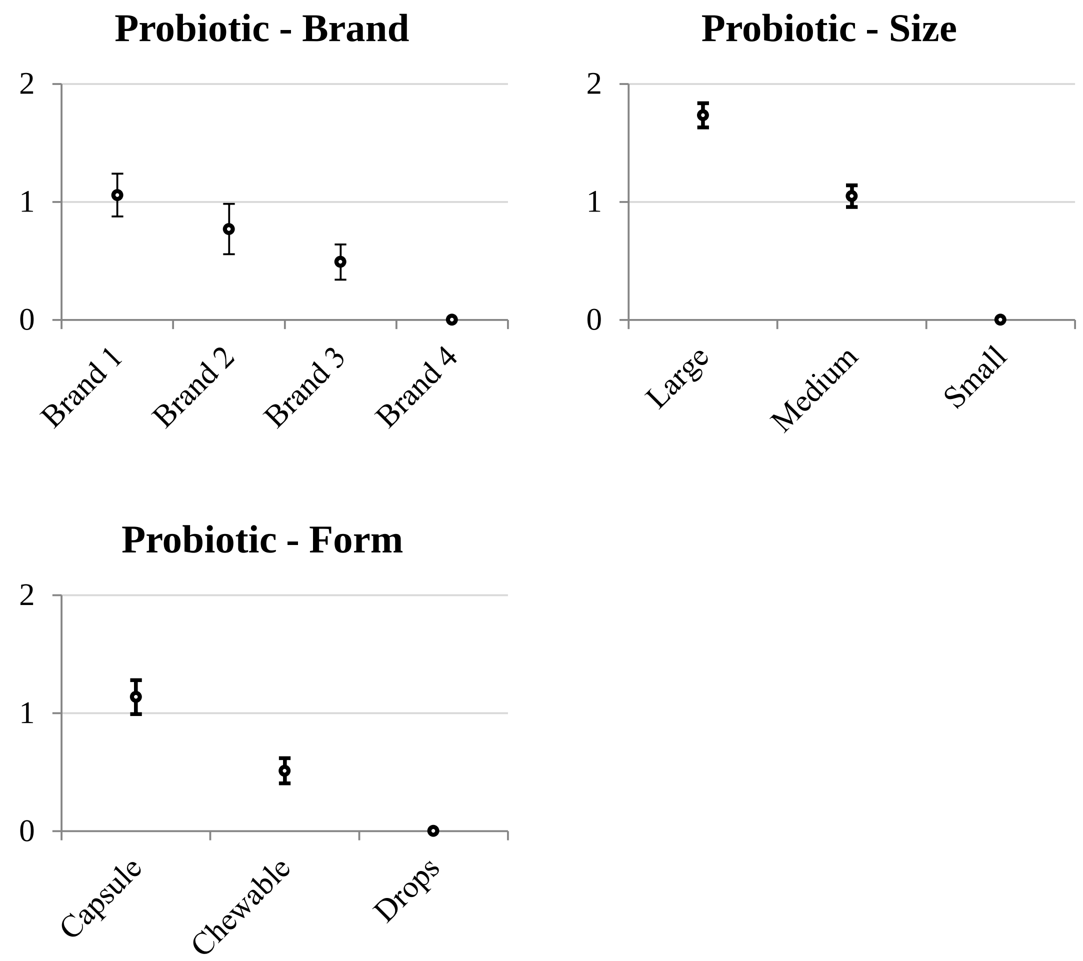

Figure 9: Attribute preferences for probiotic with 95% confidence intervals

Magnesium

In the human body, magnesium is found in bones and muscle tissue. Nuts, vegetables and whole grains are good sources of magnesium. It is critical to cellular functioning in energy production, cell reproduction and protein formation. Inadequate magnesium intake can cause fasciculation or even muscle cramps. Magnesium supplements are mostly used by athletes and those suffering from magnesium deficiency. The two most common forms of magnesium supplements are magnesium citrate and magnesium oxide. Magnesium citrate is absorbed by the body more effectively but is much more expensive than magnesium oxide. The preference toward the form of magnesium used (magnesium citrate or magnesium oxide) is captured in the brand, because all manufacturers use the same form of magnesium across their product lines.

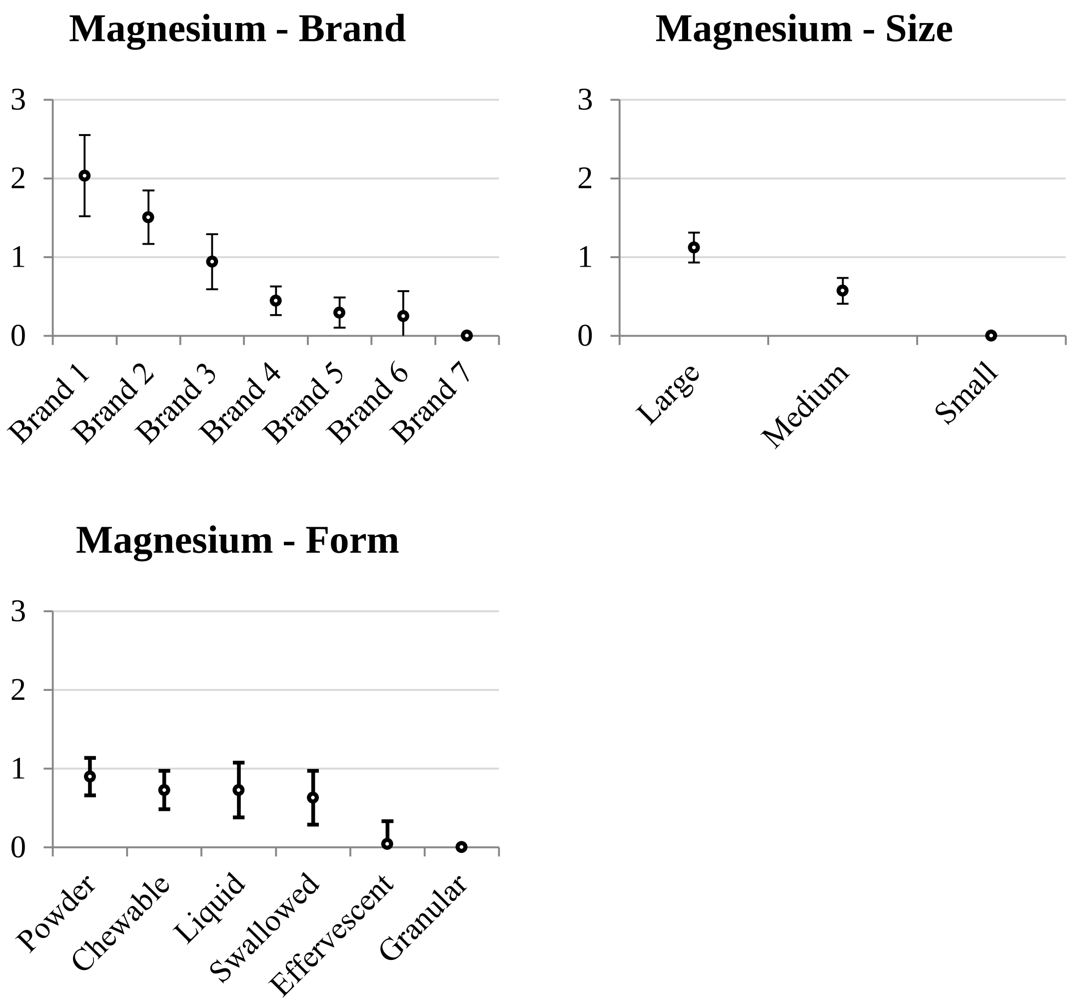

Figure 10: Attribute preferences for magnesium with 95% confidence intervals

Figure 10 presents consumers' preferences toward attribute levels of magnesium products. Products are sold in chewable, swallowed or effervescent tablet, powder, liquid and granular forms. Large package sizes are favoured over smaller ones. Generally, only the products that consist of magnesium oxide are sold in larger sizes to keep the price of an SKU manageable. Effervescent and granular forms are disliked whereas powder is the most preferred form.

Vitamin D

Vitamin D is essential for the effective utilization of calcium in the human body and the functioning of the immune system. Vitamin D is obtained naturally from food and sun exposure. Supplements are be used to replace the natural sources of vitamin D, especially during winter months when the amount of sunlight is insufficient.

Products are available in the form of chewable tablets, drops and powder. However, form is not used in the model because products available in drops and powder came to the market only after the calibration period. The content of vitamin D in each tablet is captured in the formula attribute. Products are sold in medium and large package sizes containing 100 to 300 tablets each.

The consumers' preferences toward attributes are presented in Figure 11. We find that consumers prefer large package sizes and high concentrations of vitamin D. However, the confidence intervals are wide and the results are statistically insignificant. Wide confidence intervals on parameter estimates are a symptom of multicollinearity. It occurs when independent variables are highly correlated with each other. If independent variables are perfectly correlated, then one independent variable may be presented as a linear combination of the other independent variables. Multicollinearity does not affect the goodness of fit or the predictive power of the model. However, multicollinearity leads to lack of statistical significance of independent variables even though the overall model may be significant.

Table 17 shows the variance-covariance matrix for the attribute preference parameters. We notice that covariance between brand parameters and formula parameters are very high, meaning that they are highly correlated among themselves. This implies that there exists a unique solution for the parameter estimates, but the estimates are unstable and their standard errors are inflated. This reduces our ability to draw conclusions on the consumers' preferences toward attributes of vitamin D products.

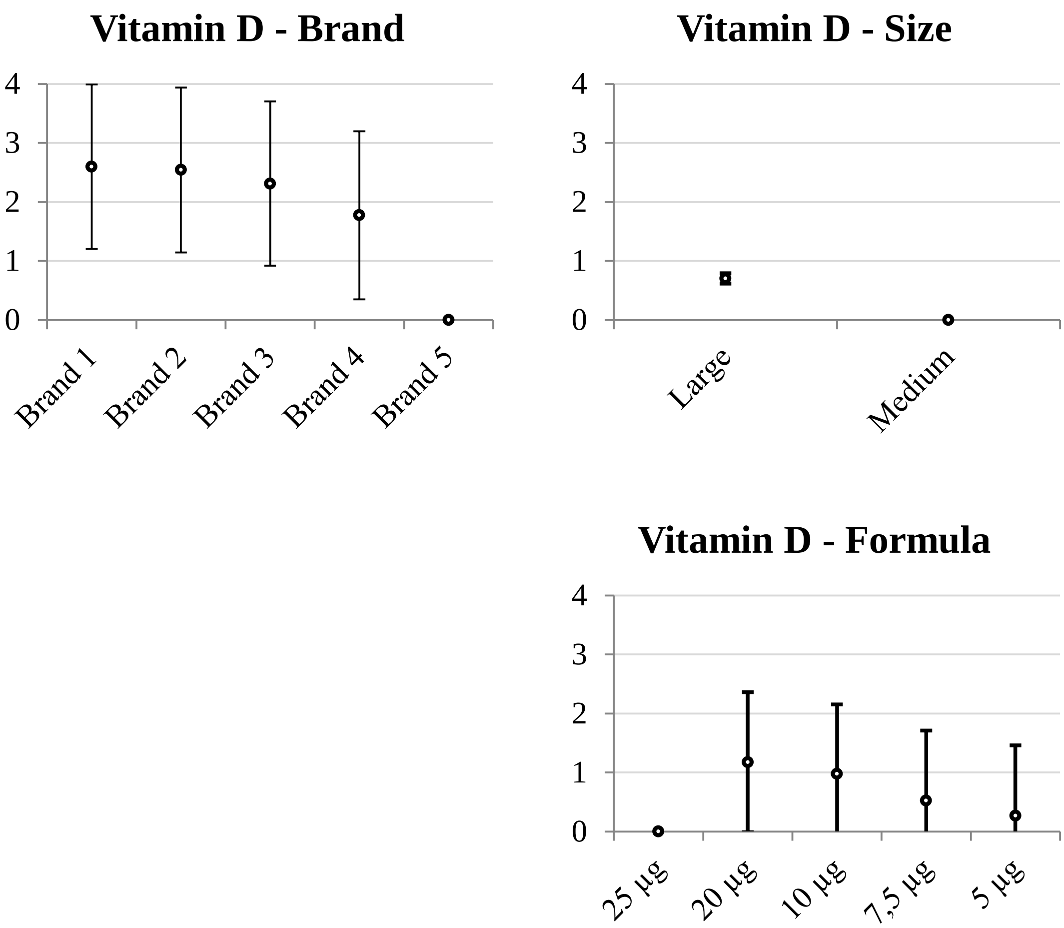

Figure 11: Attribute preferences for vitamin D with 95% confidence intervals

Table 17: Variance-covariance matrix for the brand and formula parameters for vitamin D. Covariance between brand parameters and formula parameters shaded with light grey

	Brand 1	Brand 2	Brand 3	Brand 4	$5\mu g$	$7.5\mu g$	$10\mu g$	$20\mu g$
Brand 1	0.504	0.500	0.504	0.503	0.000	-0.001	0.001	0.000
Brand 2	0.500	0.526	0.500	0.501	0.026	0.026	0.026	0.025
Brand 3	0.504	0.500	0.509	0.504	0.002	0.001	0.002	0.001
Brand 4	0.503	0.501	0.504	0.505	-0.001	0.000	0.001	-0.001
$5\mu g$	0.000	0.026	0.002	-0.001	0.369	0.364	0.361	0.361
$7.5\mu g$	-0.001	0.026	0.001	0.000	0.364	0.365	0.361	0.361
$10\mu g$	0.001	0.026	0.002	0.001	0.361	0.361	0.360	0.360
$20\mu g$	0.000	0.025	0.001	-0.001	0.361	0.361	0.360	0.367

Problems with multicollinearity arise from three frequent situations. First, a product has two or more unique attribute levels. This is true for the product that is both the only small product and the only product with 50μg formula. These unique attribute levels are perfectly correlated. As a result, we cannot determine whether the consumer prefers the small size or the 50μg formula of the product. Second, there is a subset of attribute levels among two or more attributes that are characterized only by certain products. For example, only products for children are in chewable or liquid form. Third, too many attribute levels are employed compared to the number of unique SKUs in the market. Most problems with multicollinearity can be easily avoided with better design of attributes and attribute levels.

6.1.2 Attribute Loyalty

Loyalty Parameters

Loyalty parameters are estimated for all attributes and product categories. The value of loyalty parameter indicates the parameter's importance on consumer's choice. Higher value implies higher impact. Calibration results for loyalty parameters with 95% confidence intervals are presented in Figure 12. Statistical significance of parameter estimates is measured with t-statistic in Table 18.

We find that brand loyalty is the most important factor affecting consumers' choice. Brand loyalty has the largest coefficients and t-values when compared to attribute loyalties. Probiotic and omega-3 are highly promoted categories with strong brands which explains the high values of loyalty variables for these categories. Multivitamin has the lowest brand loyalty, which is partially explained by the consumers' formula loyalty in multivitamins.

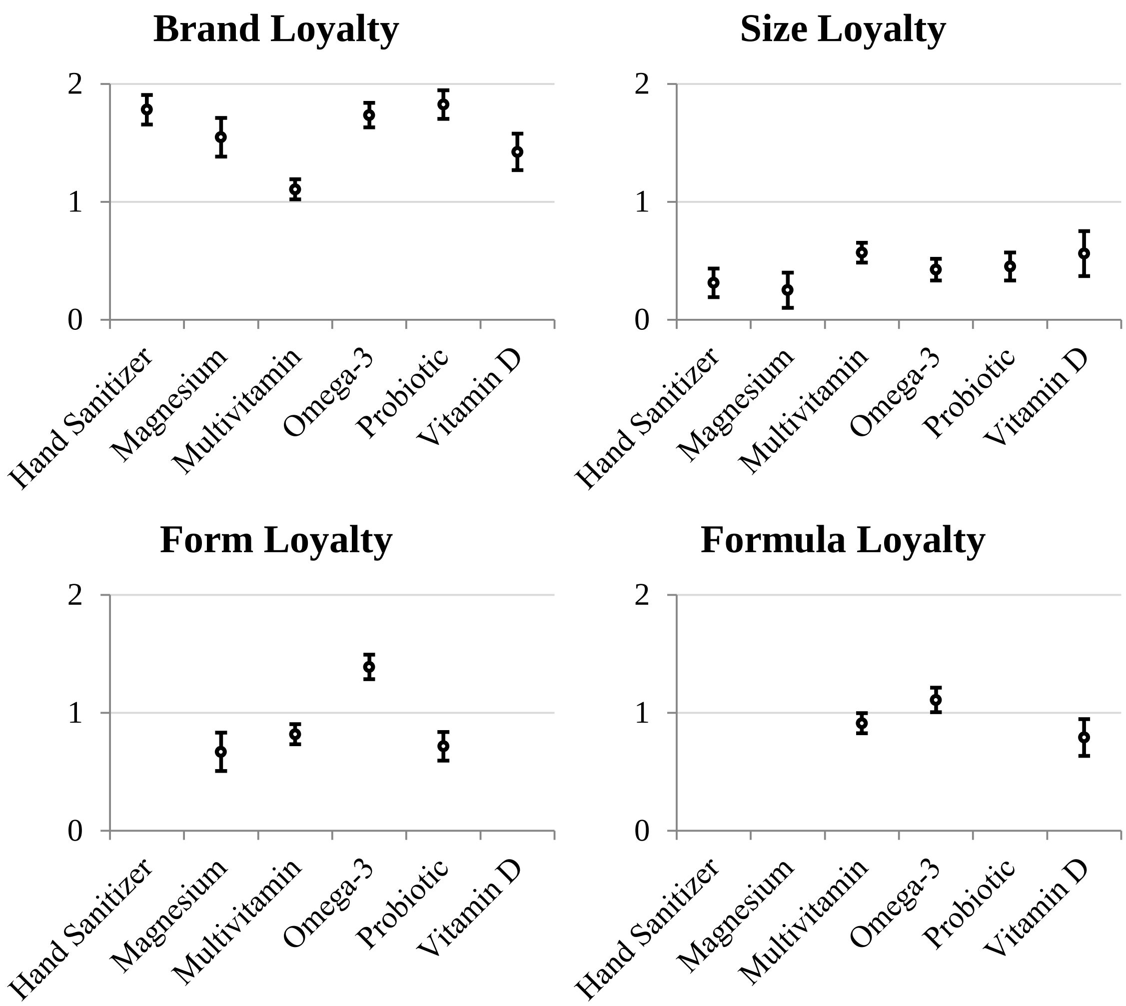

Figure 12: Loyalty parameters with 95% confidence intervals

Table 18: t-statistics for loyalty parameter estimates

	Hand Sanitizer	Magnesi-um	Multi-vitamin	Omega-3	Probiotic	Vitamin D
Brand loyalty	27.93	18.54	25.40	32.62	29.47	17.94
Size loyalty	5.09	3.29	13.34	9.08	7.47	5.74
Form loyalty	-	7.29	14.26	13.99	10.71	-
Formula loyalty	-	-	18.86	16.68	-	8.03

Formula loyalty is the second most important loyalty parameter. It seems that consumers appreciate the formula of the product almost as highly as the brand. On the other hand, repeat purchasing of the same SKU can be equally attributed to the brand as to the formula. Making conclusions on the importance of formula attribute would need further analysis on the consumers' substitution behaviour.

Form loyalty is the third most important determinant of consumer choice. It is placed in greater emphasis when the preferences of form are sharply divided. For example, some people consume omega-3 in capsules because they dislike the taste of fish oil whereas others have difficulties in swallowing the capsules, thus preferring the more inexpensive liquid form. This contributes to the high form loyalty of omega-3 products.

Size loyalty is the least important loyalty parameter. Consumers are more willing to substitute another package size than to substitute another brand, form or formula. However, all size loyalty coefficients are positive and significant.

Carry-Over Parameters

Carry-over parameters are very closely linked to loyalty parameters. Large values of carry-over indicate that loyalty is carried over at a higher rate, or it decays at a slower rate. Conversely, small values of carry-over put greater emphasis on the recent purchases than the long-term purchase history. Carry-over parameters are presented in Figure 13. Their average values and standard deviations are presented in Table 19.

The carry-over parameter estimates range between 0.57 and 0.85. In previous research, the values of carry-over parameters are usually between 0.7 and 0.9 (Fader et al. 1993). Our parameter estimates are well in line with the values used in the marketing literature.

Brand carry-over has smaller parameter estimates than size or form. A possible explanation could be that the consumers' loyalty toward brand is shorter term in nature than the loyalty toward size or form. Brand and formula carry-over have the smallest standard errors, implying that they are the most significant parameters. Size carry-over has wide confidence intervals, suggesting that the consumers' loyalty toward size is not a significant determinant of the purchase decision.

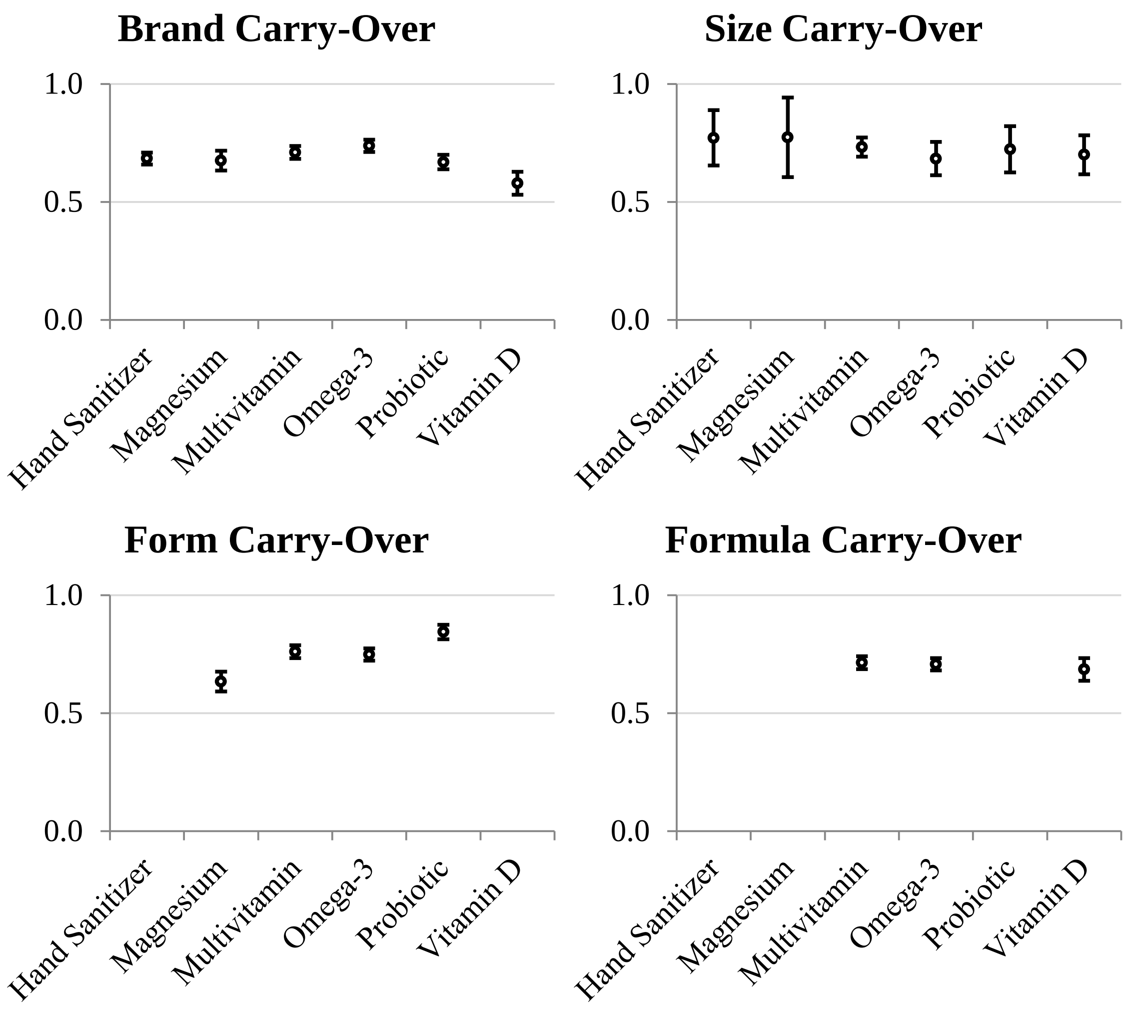

Figure 13: Carry-over parameters with 95% confidence intervals

Table 19: Average values and standard errors for carry-over parameter estimates

	Average value of parameter	Average standard error of parameter
Brand carry-over	0.675	0.017
Size carry-over	0.731	0.049
Form carry-over	0.747	0.031
Formula carry-over	0.702	0.022

6.1.3 Previous Purchase

The impact of consumers' previous purchase on the choice probability is captured in the previous purchase parameters. To investigate the delayed impact of promotion on choice behaviour, we distinguish between the previous promotional purchase and the previous non-promotional purchase. The calibration results for previous purchase parameters are presented in Figure 14.

As expected, the values of previous non-promotional purchase parameters are generally higher than the values of previous promotional purchase parameters. This suggests that promotional activity induces consumers to switch from their favourite brand to the promoted brand. When the promotion campaign is over, consumers are likely to switch back to their favourite brand.

All parameters are positive in sign, meaning that the previously purchased SKU is likely to be repurchased in the following choice occasion. Negative parameter estimates would imply that rather than buying the same SKU again, the consumer would be more willing to change to similar SKUs, e.g., different formula or package size of the same brand. In this situation, historical purchases of the consumer would be captured in the attribute-specific loyalty parameters rather than the SKU-specific previous purchase parameters. Large, positive parameter estimates indicate that repeat purchasing of the same SKU is very common.

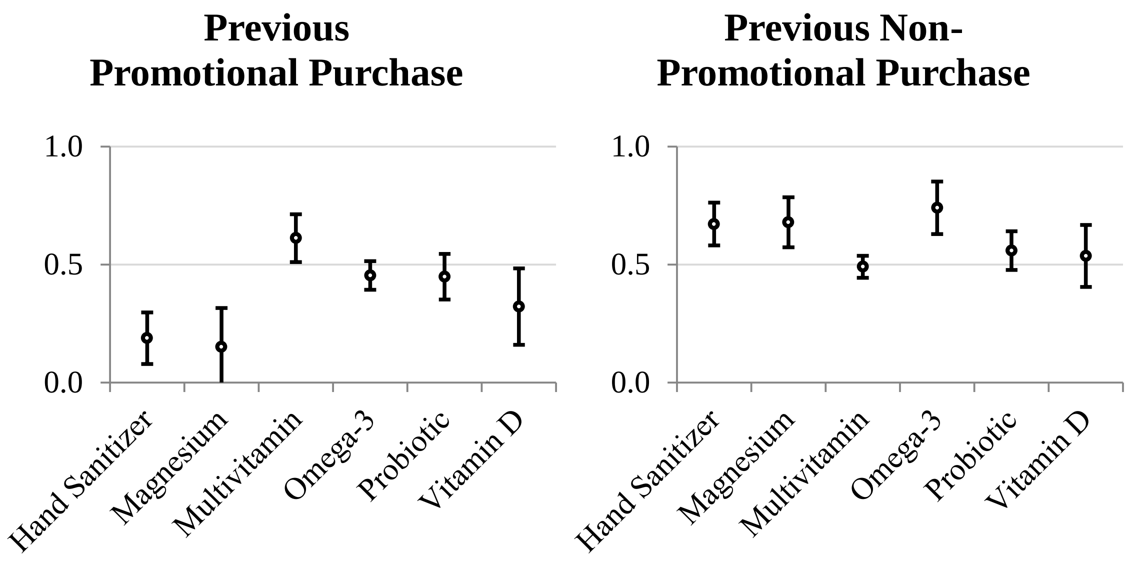

Figure 14: Previous purchase parameters with 95% confidence intervals

6.1.4 Marketing Mix

Marketing mix represents the determinant of consumer choice that is most easily controlled by the retailer. Marketing mix components include promotion, regular price and discount. The parameter estimates are presented in Figure 15 and the t-statistics for the parameter estimates in Table 20.

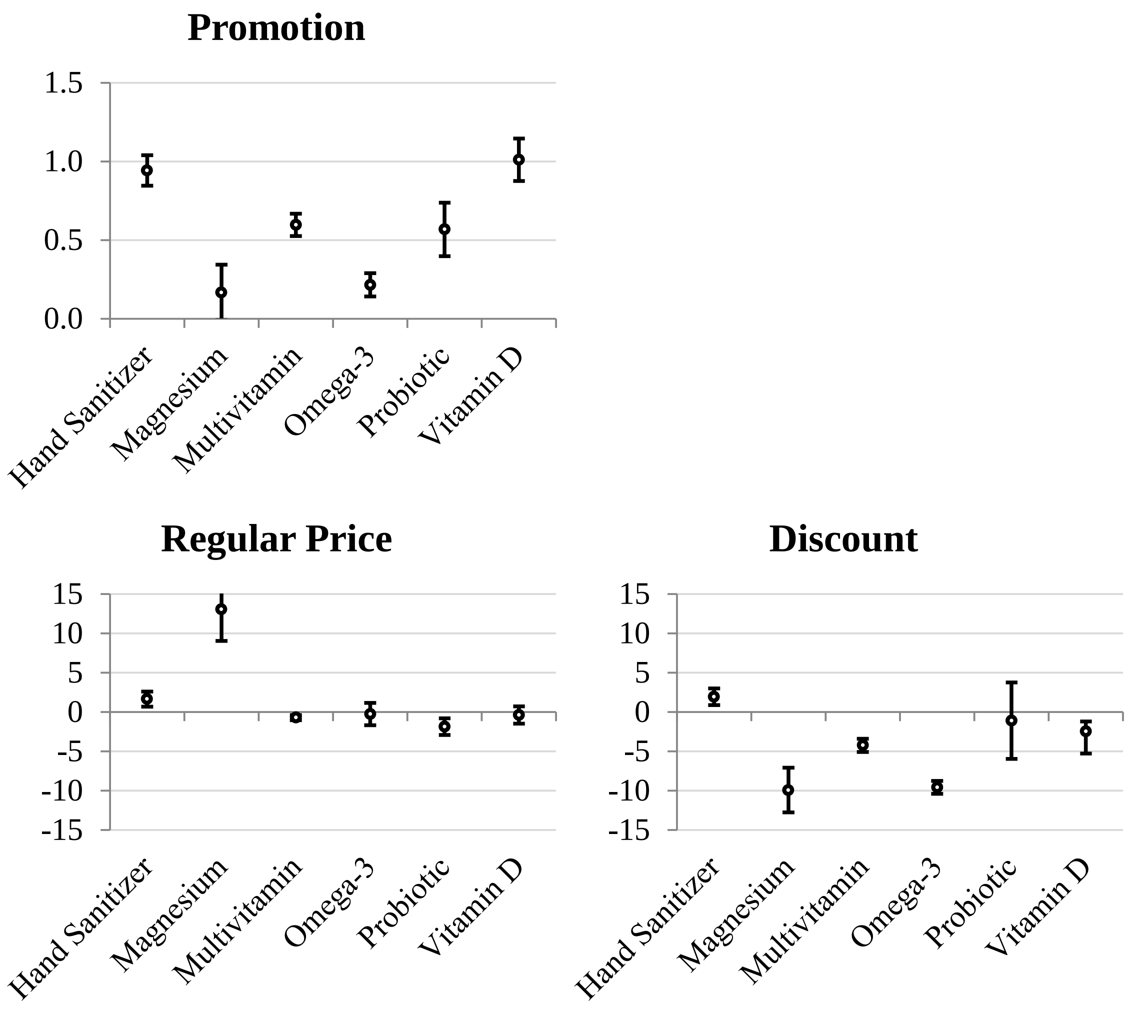

Figure 15: Marketing mix parameters with 95% confidence intervals

Table 20: t-statistics for marketing mix parameters

	Hand Sanitizer	Magnesi-um	Multi-vitamin	Omega-3	Probiotic	Vitamin D
Promotion	19.01	1.86	16.46	5.74	6.55	14.66
Regular price	3.32	6.36	-4.29	-0.38	-3.41	-0.70
Discount	3.55	-6.85	-9.95	-22.89	-0.45	-1.71

Promotion

Promotion parameter captures the impact of including the SKU in the marketing brochure. All parameter estimates have positive signs as could be expected considering that promotion attracts consumer choice. The average rise in utility due to promotion is 0.6. The parameter estimates are significant for all categories except magnesium with 5% confidence level.

Regular Price

Regular price variables have larger values when the retail prices increase. We expect lower regular price to increase the SKU's attractiveness, resulting in a negative sign for the regular price parameter. This applies only to four categories out of six. Confidence intervals are wide and t-values are small for regular price parameters across all categories. This suggests that regular price has not correctly captured the effect of price level to choice behaviour. The main reason is that the mark-ups for products changed very rarely during the observation period.

Discount

Discount is defined as the percentage decrease in price from the regular price of the SKU. Discount parameters are expected to have negative signs. All except hand sanitizer have negative values for discount parameter estimates. The average value for a discount parameter across categories is -4. Thus, a discount of -15% would result in a rise of 0.6 in the utility of an SKU. Heavily discounted category omega-3 has smaller and more significant discount parameter than other categories. The results suggests that discount is an important factor in determining consumers' choice behaviour.

6.2 Specifications with Increasing Number of Variables

We now assess the contributions of variables in explaining the consumers' purchase behaviour. This is done by creating alternative model specifications with increasing number of variables and calibrating the models on the data. We can follow the improvements in log-likelihood by adding groups of variables in a hierarchical fashion to see which parameters cause the largest increases in log-likelihood. This also allows us to investigate the stability of the coefficients against changes in model specifications.

The model specifications are explained in Table 21. Specification S0 is the null model with all coefficients set to zero. Specification S1 is the baseline model that contains only the SKU-specific coefficients of the GL model or the attribute level-specific coefficients of the FH model. In specification S2, loyalty variables are introduced in addition to the variables already included in the model S1. Each new specification brings in more variables in addition to the ones introduced in earlier specifications. The final specification S6 includes all variables presented in this thesis.

Table 21: Model specifications with increasing number of variables

Specification	Description
S0	Null model (all coefficients set to zero)
S1	SKU- or attribute level-specific coefficients only
S2	S1 + loyalty variables
S3	S2 + previous purchase variables
S4	S3 + promotion variable
S5	S4 + regular price variable
S6	S5 + discount variable

Table 22 and Table 23 present the parameter estimates for the GL model and the FH model, respectively. These parameter estimates are calculated with different model specifications for the magnesium category. We see that the GL model requires more parameters but results in higher likelihood ratio and lower log-likelihood. Both models have very similar parameter estimates for loyalty, carry-over, previous purchase and marketing mix variables. Parameter estimates are also very stable as more variables are introduced.

Table 22: Parameter estimates for the GL model for magnesium with different model specifications

	S0	S1	S2	S3	S4	S5	S6
SKU #1		1.03	1.29	1.24	1.21	3.35	3.35
SKU #2		1.95	1.75	1.61	1.61	2.93	2.92
SKU #3		2.32	1.77	1.68	1.68	2.48	2.46
SKU #4		2.44	1.64	1.64	1.47	2.14	2.16
SKU #5		0.77	0.66	0.64	0.64	1.96	1.95
SKU #6		1.44	1.28	1.31	0.80	1.53	1.85
SKU #7		1.40	1.54	1.43	1.25	1.91	1.75
SKU #8		1.97	1.19	1.22	1.05	1.72	1.75
SKU #9		1.54	1.63	1.61	1.38	2.03	1.74
SKU #10		1.42	0.92	0.92	0.92	1.72	1.71
SKU #11		0.48	0.69	0.60	0.65	1.31	1.27
SKU #12		1.38	1.30	1.22	1.22	1.21	1.21
SKU #13		0.67	0.57	0.56	0.56	1.23	1.21
SKU #14		1.53	1.16	1.12	1.12	1.12	1.12
SKU #15		0.29	0.18	0.21	0.17	1.00	1.00
SKU #16		0.34	0.34	0.35	0.35	0.35	0.35
SKU #17		0.00	0.00	0.00	0.00	0.00	0.00
Brand loyalty			1.87	1.54	1.54	1.54	1.55
Size loyalty			0.55	0.25	0.25	0.25	0.25
Form loyalty			0.82	0.66	0.66	0.66	0.67
Brand carry-over			0.64	0.68	0.68	0.68	0.67
Size carry-over			0.63	0.77	0.77	0.77	0.77
Form carry-over			0.60	0.63	0.63	0.63	0.63
Previous prom. purchase				0.12	0.10	0.10	0.15
Previous non-prom. purchase				0.68	0.69	0.69	0.68
Promotion					0.54	0.53	0.17
Regular price						13.19	13.04
Discount							-9.94
Log-likelihood	-10,418	-9,514	-8,596	-8,517	-8,488	-8,468	-8,444
Likelihood ratio	0.000	0.087	0.175	0.183	0.185	0.187	0.190
Parameters	0	16	22	24	25	26	27

Table 23: Parameter estimates for the FH model for magnesium with different model specifications

	S0	S1	S2	S3	S4	S5	S6
Brand 1		0.59	0.62	0.63	0.59	2.03	2.03
Brand 2		1.39	0.96	0.88	0.87	1.51	1.51
Brand 3		1.58	1.16	1.19	0.97	0.97	0.94
Brand 4		0.70	0.31	0.32	0.30	0.44	0.44
Brand 5		1.02	0.36	0.45	0.24	0.25	0.29
Brand 6		1.34	0.99	0.89	0.90	0.25	0.25
Brand 7		0.00	0.00	0.00	0.00	0.00	0.00
Large		1.40	1.28	1.12	1.13	1.13	1.12
Medium		0.75	0.66	0.57	0.58	0.58	0.57
Small		0.00	0.00	0.00	0.00	0.00	0.00
Powder		1.06	0.97	0.89	0.93	0.92	0.90
Chewable		1.13	0.75	0.72	0.77	0.76	0.73
Liquid		0.93	1.11	0.95	0.95	0.94	0.73
Swallowed		0.71	0.68	0.62	0.65	0.66	0.63
Effervescent		0.02	0.02	0.04	0.07	0.06	0.04
Granular		0.00	0.00	0.00	0.00	0.00	0.00
Brand loyalty			1.87	1.53	1.53	1.53	1.54
Size loyalty			0.58	0.26	0.26	0.27	0.26
Form loyalty			0.80	0.64	0.64	0.64	0.64
Brand carry-over			0.64	0.68	0.68	0.68	0.68
Size carry-over			0.64	0.77	0.78	0.78	0.78
Form carry-over			0.59	0.63	0.63	0.63	0.63
Previous prom. purchase				0.09	0.08	0.08	0.12
Previous non-prom. purchase				0.72	0.73	0.73	0.72
Promotion					0.59	0.59	0.28
Regular price						12.98	12.75
Discount							-8.61
Log-likelihood	-10,418	-9,587	-8,665	-8,574	-8,538	-8,518	-8,500
Likelihood ratio	0.000	0.080	0.168	0.177	0.180	0.182	0.184
Parameters	0	13	19	21	22	23	24

Log-likelihoods for the FH model and GL model in the training set and the test set are presented in Table 24–Table 27. The inclusion of SKU-specific and attribute level-specific coefficients in specification S1 results in the highest improvement. The addition of loyalty variables in S2 produces a large increase in log-likelihood. Introduction of the previous purchase variables in S3 produces a further increase in log-likelihood, but the impact is significantly lower. Improvement of approximately the same magnitude is achieved when the promotion variable is introduced in S4. The introduction of regular price and discount variables in specifications S4 and S5, respectively, do not show a significant improvement in log-likelihood except for the most heavily promoted category omega-3.

Likelihood ratio allows us to compare the performance of different models in different data sets. Likelihood ratios are presented in Table 28–Table 31. When we compare the GL model's likelihood ratios for the same model specifications in the training set and the test set, we notice that the values for likelihood ratios are of equal size. This suggests that there has not been significant deterioration of model performance when we use our model to forecast consumers' choice behaviour in the independent test set. The results are very encouraging as they allow us to generalize our model with confidence to a larger population of the retailer's customers. However, we note that there is some deterioration of model performance in the vitamin D category. This may be a result of a smaller set of observations obtained for the vitamin D category, highlighting the importance of large data sets in quantitative studies.

Table 24: Training set log-likelihood for the GL model in the calibration period

	Hand Sanitizer	Magnesi-um	Multi-vitamin	Omega-3	Probiotic	Vitamin D
S0	-22,644	-10,418	-58,155	-48,293	-18,402	-8,109
S1	-19,057	-9,514	-48,646	-36,467	-15,761	-7,496
S2	-18,208	-8,596	-46,377	-33,573	-13,824	-7,048
S3	-18,107	-8,517	-46,113	-33,288	-13,713	-7,015
S4	-17,847	-8,488	-45,813	-32,965	-13,689	-6,889
S5	-17,845	-8,468	-45,804	-32,965	-13,683	-6,888
S6	-17,838	-8,444	-45,737	-32,685	-13,683	-6,887

Table 25: Test set log-likelihood for the GL model in the calibration period

	Hand Sanitizer	Magnesi-um	Multi-vitamin	Omega-3	Probiotic	Vitamin D
S0	-11,286	-5,298	-29,657	-24,185	-9,836	-3,990
S1	-9,585	-4,913	-24,756	-18,228	-8,465	-3,747
S2	-9,126	-4,369	-23,464	-16,734	-7,421	-3,607
S3	-9,064	-4,323	-23,338	-16,555	-7,367	-3,595
S4	-8,931	-4,304	-23,208	-16,423	-7,362	-3,544
S5	-8,929	-4,296	-23,197	-16,423	-7,360	-3,543
S6	-8,928	-4,285	-23,176	-16,306	-7,360	-3,542

Table 26: Training set log-likelihood for the FH model in the calibration period

	Hand Sanitizer	Magnesi-um	Multi-vitamin	Omega-3	Probiotic	Vitamin D
S0	-22,644	-10,418	-58,155	-48,293	-18,402	-8,109
S1	-20,821	-9,587	-51,500	-37,079	-15,925	-7,583
S2	-19,070	-8,665	-49,049	-34,102	-13,964	-7,126
S3	-18,808	-8,574	-48,442	-33,859	-13,822	-7,088
S4	-18,569	-8,538	-48,125	-33,467	-13,796	-6,935
S5	-18,569	-8,518	-47,966	-33,402	-13,790	-6,931
S6	-18,562	-8,500	-47,903	-33,085	-13,790	-6,927

Table 27: Test set log-likelihood for the FH model in the calibration period

	Hand Sanitizer	Magnesi-um	Multi-vitamin	Omega-3	Probiotic	Vitamin D
S0	-11,286	-5,298	-29,657	-24,185	-9,836	-3,990
S1	-10,308	-4,951	-26,095	-18,517	-8,518	-3,767
S2	-9,490	-4,403	-24,736	-16,987	-7,473	-3,629
S3	-9,338	-4,349	-24,427	-16,833	-7,409	-3,616
S4	-9,220	-4,327	-24,265	-16,672	-7,403	-3,561
S5	-9,219	-4,319	-24,182	-16,633	-7,402	-3,559
S6	-9,218	-4,309	-24,158	-16,507	-7,402	-3,558

Table 28: Training set likelihood ratio for the GL model in the calibration period

	Hand Sanitizer	Magnesi-um	Multi-vitamin	Omega-3	Probiotic	Vitamin D
S0	0.000	0.000	0.000	0.000	0.000	0.000
S1	0.158	0.087	0.164	0.245	0.144	0.076
S2	0.196	0.175	0.203	0.305	0.249	0.131
S3	0.200	0.183	0.207	0.311	0.255	0.135
S4	0.212	0.185	0.212	0.317	0.256	0.150
S5	0.212	0.187	0.212	0.317	0.256	0.151
S6	0.212	0.190	0.214	0.323	0.256	0.151

Table 29: Test set likelihood ratio for the GL model in the calibration period

	Hand Sanitizer	Magnesi-um	Multi-vitamin	Omega-3	Probiotic	Vitamin D
S0	0.000	0.000	0.000	0.000	0.000	0.000
S1	0.151	0.073	0.165	0.246	0.139	0.061
S2	0.191	0.175	0.209	0.308	0.246	0.096
S3	0.197	0.184	0.213	0.315	0.251	0.099
S4	0.209	0.188	0.217	0.321	0.252	0.112
S5	0.209	0.189	0.218	0.321	0.252	0.112
S6	0.209	0.191	0.219	0.326	0.252	0.112

Table 30: Training set likelihood ratio for the FH model in the calibration period

	Hand Sanitizer	Magnesi-um	Multi-vitamin	Omega-3	Probiotic	Vitamin D
S0	0.000	0.000	0.000	0.000	0.000	0.000
S1	0.080	0.080	0.114	0.232	0.135	0.065
S2	0.158	0.168	0.157	0.294	0.241	0.121
S3	0.169	0.177	0.167	0.299	0.249	0.126
S4	0.180	0.180	0.172	0.307	0.250	0.145
S5	0.180	0.182	0.175	0.308	0.251	0.145
S6	0.180	0.184	0.176	0.315	0.251	0.146

Table 31: Test set likelihood ratio for the FH model in the calibration period

	Hand Sanitizer	Magnesi-um	Multi-vitamin	Omega-3	Probiotic	Vitamin D
S0	0.000	0.000	0.000	0.000	0.000	0.000
S1	0.087	0.066	0.120	0.234	0.134	0.056
S2	0.159	0.169	0.166	0.298	0.240	0.090
S3	0.173	0.179	0.176	0.304	0.247	0.094
S4	0.183	0.183	0.182	0.311	0.247	0.107
S5	0.183	0.185	0.185	0.312	0.247	0.108
S6	0.183	0.187	0.185	0.317	0.247	0.108

6.3 Forecasting Performance

We can evaluate the forecasting performance of our model by comparing the predicted shares of purchases to the actual share of purchases by a one-month period. This is done separately to the training set and the test set. We also examine the forecasting performance of the model outside the calibration period.

We use the GL model in predicting the share of purchases for existing products. When applicable, the GL model is more accurate than the one-segment FH model. However, the GL model cannot be used to forecast sales of new products that enter the market during the forecasting period because it lacks information on the consumers' preferences toward these products. The FH model is able to determine the consumers' preference toward product line extensions that employ the attribute levels already found in the market. Therefore, the FH model is used to forecast the share of new products.

6.3.1 Existing Products

Figure 16 shows actual and predicted share of purchases of the best-selling probiotic for both training and test sets. The dashed line in the middle is the prediction of the GL model, the grey shading indicates approximate 95% confidence intervals for the model prediction and the solid line represents the actual share of purchases. The vertical dashed line separates the calibration period (2008-2009) from the forecasting period (2010-2011). The share of purchases is calculated as the number of purchases of the SKU divided by the total number of purchases of existing products during the one-month period. Note that we leave out all new products that were introduced during the forecasting period because the GL model is not able to forecast their share of purchases.

We find that the model is accurate in forecasting the share of the best-selling SKU in the probiotic category. The model correctly explains the rise in share in 2008, the promotional peaks in 2010 and the downward trend in 2011. The confidence intervals are wider for the test set as it contains only some 1/3 of observations while the remaining 2/3 are included in the training set.

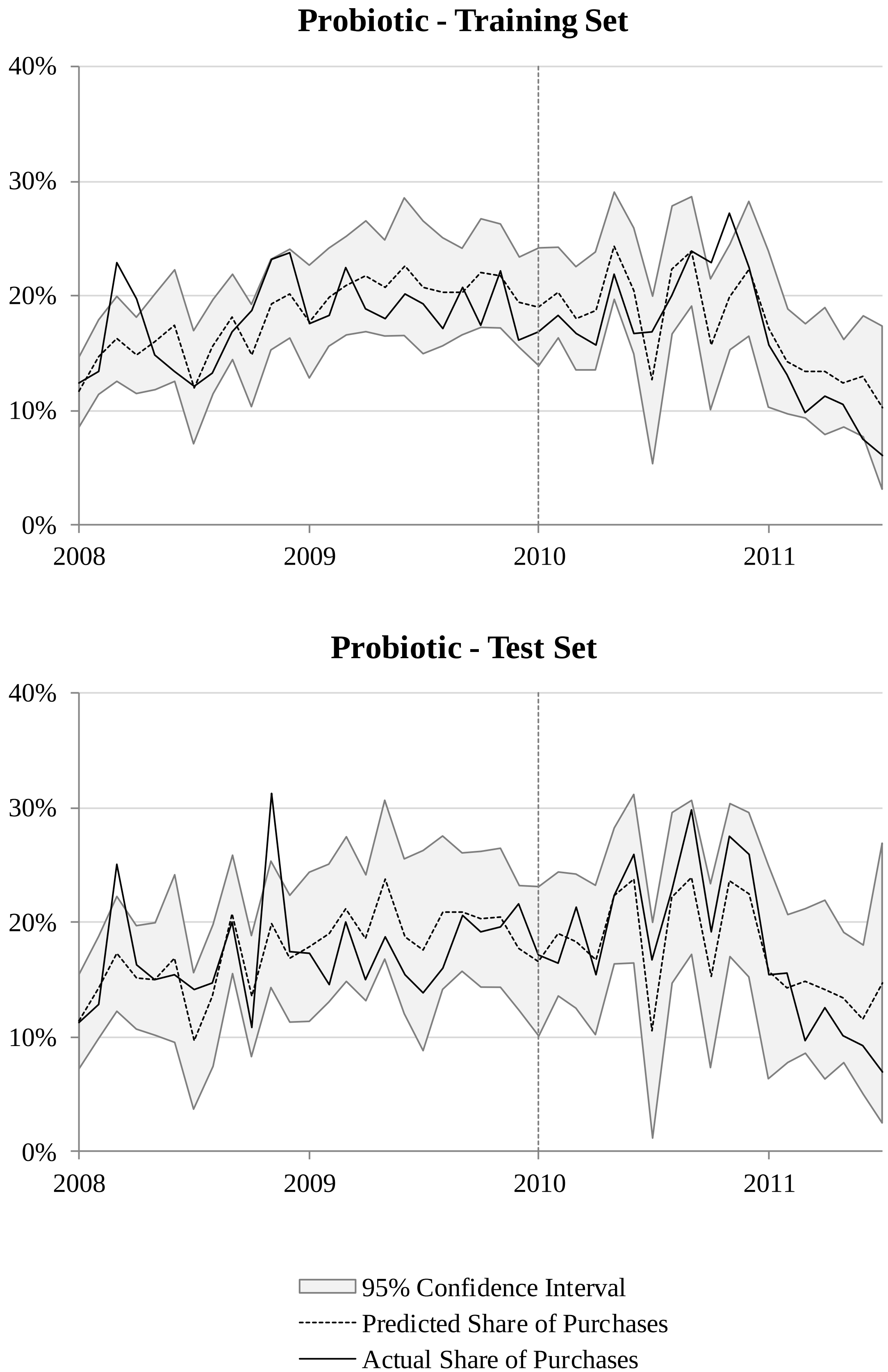

Figure 16: Actual and predicted share of purchases of the best-selling probiotic in the training set and the test set

As we see for the best-selling probiotic SKU in the training set, 5 of the 43 observations (11.6%) lie outside the 95% confidence intervals. In the test set, 2 of the 43 observations (4.7%) fall outside the 95% confidence intervals. This suggests that the 95% confidence intervals are too optimistic, or that the calculation procedure underestimates the sizes of the confidence intervals.

We test the hypothesis that the confidence intervals are too optimistic. Table 32 presents the share of observations located outside the 95% confidence intervals for existing SKUs in each category. Results are presented separately for the training set and the test set, and the calibration period and the forecasting period. We notice that significantly more than 5% of observations lie outside the theoretical 95% confidence intervals. On average, 12.4% of observations lie outside theoretical 95% confidence intervals in the training set during the calibration period. This verifies that the confidence intervals are underestimated and the assumption that the purchase decision is an independent binomial draw does not hold.

A probable explanation is that our model has not captured all variables affecting consumer choice and the model could still be improved. The purchase occasions also may not be independent of each other. We notice that there is generally no deterioration of tracking performance when we move from the training set to the test set. However, there is a notable deterioration of model performance when we move from the calibration period to the forecasting period. A partial explanation is that our model cannot capture trends in the attractiveness of SKUs and their attributes. During the long observation periods, some products have gained in popularity and some have lost share, which leads to systematic errors in forecasting performance.

The results also indicate that there are significant differences in explanatory power of the model across categories. Our model performs very well for hand sanitizer, magnesium and probiotic categories, which have clear category boundaries and more homogeneous assortment. In contrast, heavy product differentiation and large number of unique SKUs leads to weaker performance for omega-3 and multivitamin categories.

Table 32: Share of observations outside the 95% confidence intervals for the GL model

	Hand Sanitizer	Magnesi-um	Multi-vitamin	Omega-3	Probiotic	Vitamin D	Average
Training Set							
Calibration	11.5%	8.1%	13.4%	22.2%	8.4%	11.0%	12.4%
Forecasting	10.8%	13.2%	25.3%	20.1%	12.2%	16.3%	16.3%
Test Set							
Calibration	10.9%	9.0%	11.7%	27.5%	10.3%	9.8%	13.2%
Forecasting	7.6%	9.0%	16.8%	26.1%	8.8%	12.1%	13.4%

Figure 17 presents actual and predicted share of purchases of the best-selling SKUs of each category in the training set. The model does well in forecasting the promotional peaks and the overall trends. It seems that the forecasts for omega-3 are constantly ahead of the actual results. Note that the SKUs in hand sanitizer and magnesium categories experienced stock-outs during the calibration period.

Figure 18 presents the share of purchases for the same SKUs in the test set. We notice that the confidence intervals are wider and actual share has higher variance due to the smaller sample size. However, the promotional peaks and trends are still explained accurately.

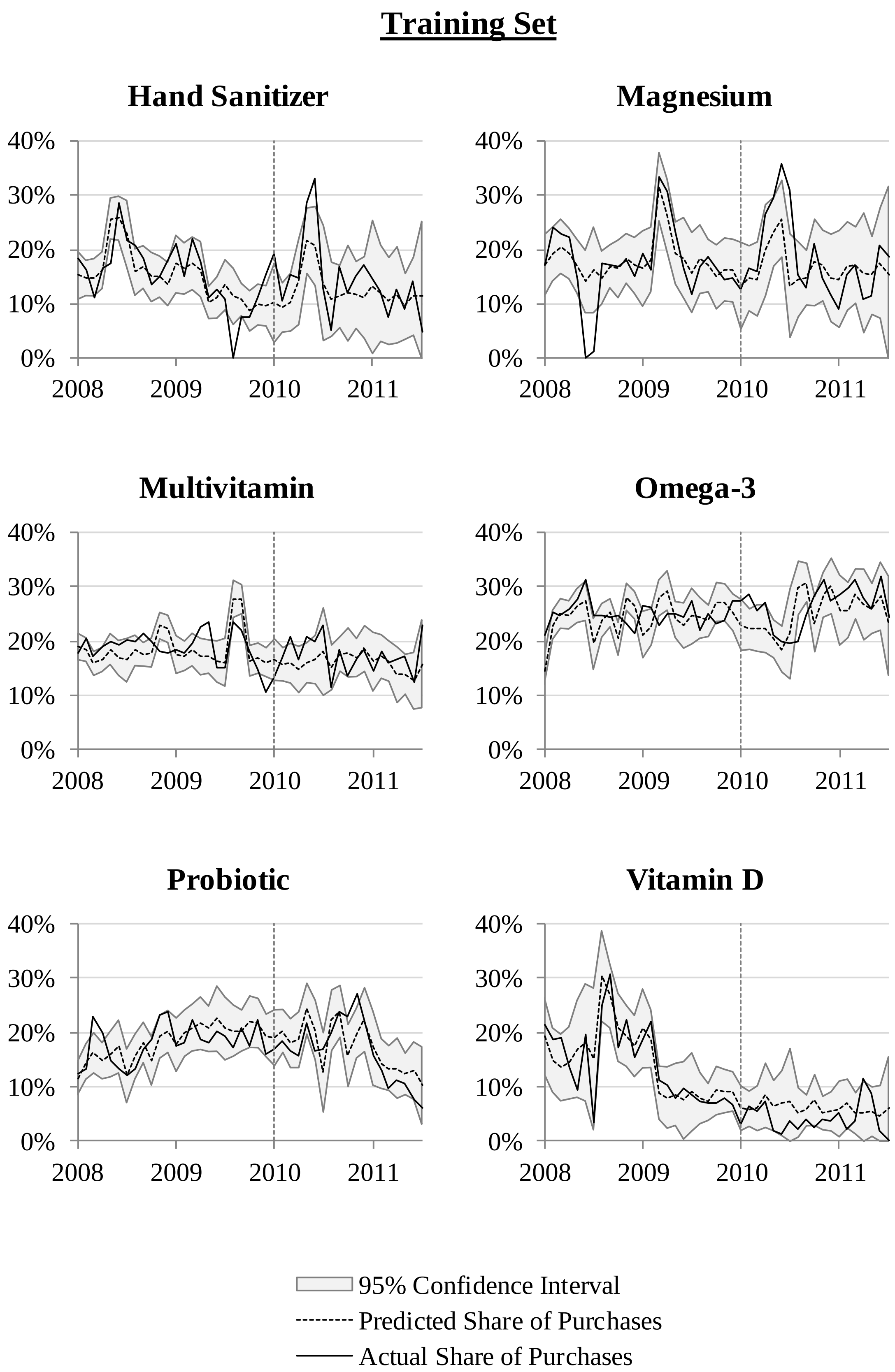

Figure 17: Actual and predicted share of purchases of the best-selling SKUs of each category in the training set

Test Set

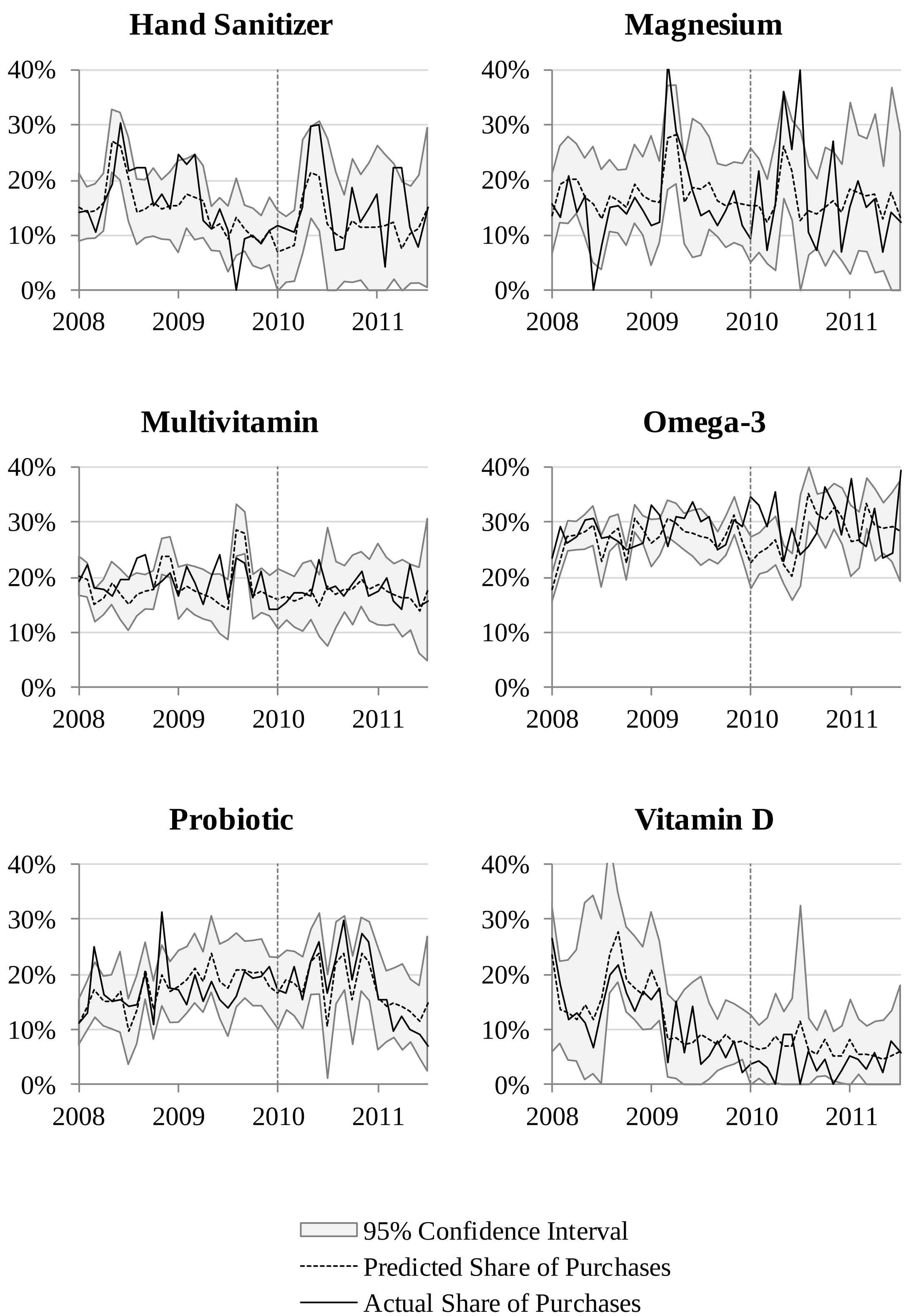

Figure 18: Actual and predicted share of purchases of the best-selling SKUs of each category in the test set

6.3.2 New Products

The FH model is used to determine the share of purchases for the new SKUs that are introduced after the calibration period. Most of these new SKUs are line extensions that feature attribute levels that are already found in the market. The FH model can be readily used to forecast their share of purchases. These SKUs and their attributes are presented in Table 33.

Table 33: Line extensions and their attribute levels

SKU	Category	Brand	Size	Form	Formula
SKU #1	Hand Sanitizer	Brand 5	Small	-	-
SKU #2	Hand Sanitizer	Brand 5	Small	-	-
SKU #3	Magnesium	Brand 4	Large	Swallowed	-
SKU #4	Multivitamin	Brand 5	Medium	Swallowed	For Women
SKU #5	Multivitamin	Brand 1	Medium	Swallowed	For Women
SKU #6	Multivitamin	Brand 5	Small	Swallowed	For Women
SKU #7	Multivitamin	Brand 5	Medium	Swallowed	Added Omega-3
SKU #8	Probiotic	Brand 2	Large	Drops	-
SKU #9	Probiotic	Brand 1	Large	Chewable	-
SKU #10	Probiotic	Brand 1	Medium	Chewable	-
SKU #11	Probiotic	Brand 4	Small	Capsule	-
SKU #12	Probiotic	Brand 4	Medium	Chewable	-
SKU #13	Probiotic	Brand 2	Large	Chewable	-
SKU #14	Probiotic	Brand 4	Small	Capsule	-
SKU #15	Probiotic	Brand 2	Medium	Chewable	-
SKU #16	Probiotic	Brand 2	Small	Chewable	-
SKU #17	Probiotic	Brand 3	Medium	Chewable	-
SKU #18	Probiotic	Brand 4	Large	Drops	-
SKU #19	Probiotic	Brand 1	Large	Drops	-
SKU #20	Probiotic	Brand 1	Large	Drops	-
SKU #21	Vitamin D	Brand 3	Large	-	20 µg
SKU #22	Vitamin D	Brand 3	Large	-	25 µg
SKU #23	Vitamin D	Brand 3	Medium	-	25 µg
SKU #24	Vitamin D	Brand 2	Large	-	10 µg

Although most of the new SKUs employ attribute levels already found in the market, every now and then products with unique attribute levels are introduced. They can have a unique brand, size, form or formula that has not been previously found in the market. The FH model cannot forecast the popularity of these products unless the consumers' preference toward the new attribute level can be assessed. The products with new attribute levels are introduced in Table 34. Note that we leave out all new products with unique attribute levels because the FH model cannot give reliable forecasts on their share of purchases.

Table 34: New products with the new attribute level underlined

SKU	Category	Brand	Size	Form	Formula
SKU #25	Probiotics	Brand 2	Medium	<u>Powder</u>	-
SKU #26	Probiotics	Brand 1	<u>Extra Large</u>	Chewable	-
SKU #27	Probiotics	Brand 1	<u>Extra Large</u>	Chewable	-
SKU #28	Vitamin D	Brand 2	Medium	-	<u>50 μg</u>

We are interested in the model's capability of forecasting the share of purchases for the new SKUs. The actual share is compared against the share predicted by the FH model in Table 35. SKUs are sorted in decreasing order of their actual share of purchases. The share of purchases is calculated as the number of purchases of the new SKU divided by the total number of purchases during the period that the SKU has been on the market. The results for the test set are similar with the results for the training set, and are omitted for brevity. We notice that the FH model gives rather accurate first-cut forecasts for the share of new SKUs given that no additional data is used on the consumers' preferences toward the specific SKU.

Table 35: Actual and predicted share of purchases in the training set

SKU	Actual Share	Predicted Share	Absolute Error	Percentage Error
SKU #13	7.9%	8.3%	0.4%	4.7%
SKU #2	6.8%	2.1%	-4.7%	-69.2%
SKU #9	5.6%	8.0%	2.3%	41.0%
SKU #8	4.6%	4.3%	-0.3%	-7.5%
SKU #15	4.6%	3.9%	-0.6%	-13.8%
SKU #21	3.9%	6.5%	2.6%	65.7%
SKU #1	3.5%	2.8%	-0.7%	-20.8%
SKU #3	2.9%	7.1%	4.2%	142.2%
SKU #5	2.4%	2.9%	0.5%	20.5%
SKU #22	1.5%	1.6%	0.0%	1.6%
SKU #10	1.5%	3.2%	1.7%	116.1%
SKU #20	1.2%	2.4%	1.2%	94.8%
SKU #7	1.1%	5.4%	4.3%	409.9%
SKU #24	1.0%	5.3%	4.3%	450.6%
SKU #19	0.9%	2.9%	2.0%	225.8%
SKU #4	0.8%	1.4%	0.6%	77.3%
SKU #17	0.5%	1.3%	0.8%	155.7%
SKU #12	0.4%	0.4%	0.1%	20.8%
SKU #6	0.3%	0.9%	0.5%	173.0%
SKU #23	0.3%	0.6%	0.3%	98.8%
SKU #11	0.2%	0.4%	0.1%	67.4%
SKU #18	0.2%	0.5%	0.3%	129.9%
SKU #14	0.2%	0.3%	0.1%	76.3%
SKU #16	0.2%	0.7%	0.5%	273.5%

For a visual comparison, Figure 19 presents the actual share of purchases plotted against the predicted share of purchases. We notice that the prediction of share is biased upward, meaning that the model prediction of share is generally higher than the actual outcome. This effect is more pronounced with lowest-selling SKUs. A probable explanation is that it takes some time for the consumers to find the new SKU. Usually the introduction of new products is accompanied with a marketing campaign which shortens the time it takes for a consumer to acknowledge the new SKU in the market. If the SKU is not promoted by the retailer, consumers do not know that it is included in the assortment and thus cannot buy it. Most of the lowest-selling products are not promoted by the retailer, which explains the low actual shares of the low-selling SKUs.

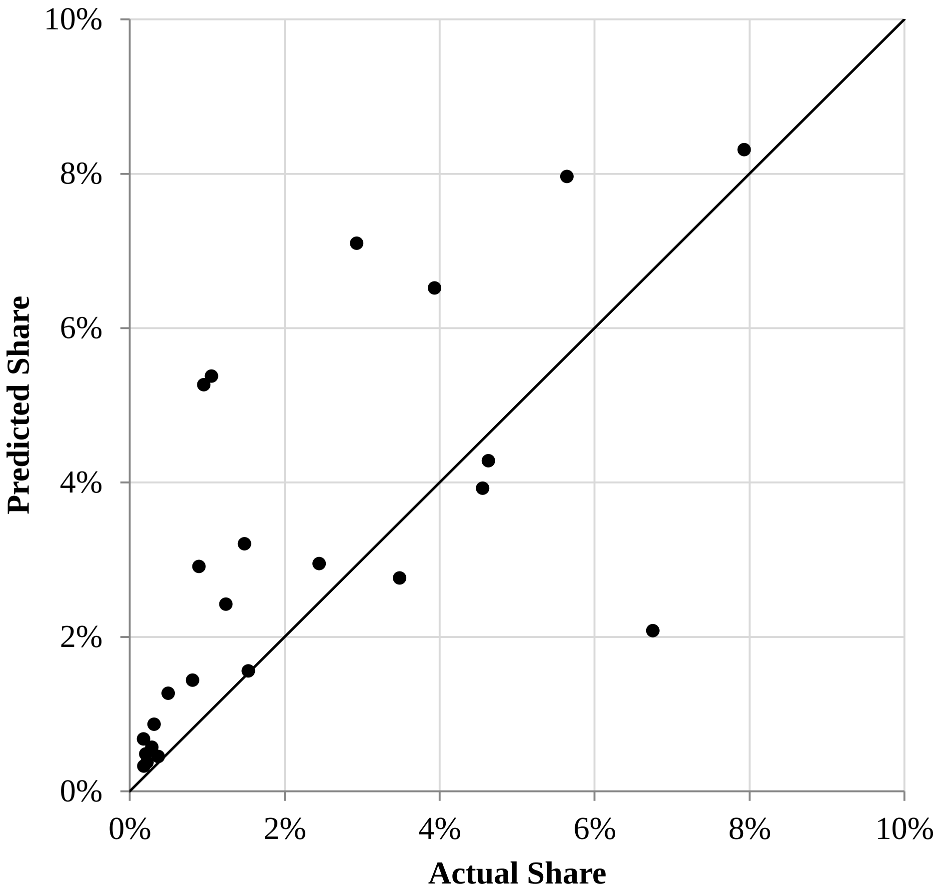

Figure 19: Actual share plotted against predicted share

Forecasting results for twelve new SKUs with the largest actual share of purchases are presented in Figure 20 and Figure 21. We see that the model predictions are remarkably well in line with the actual shares of purchases. The promotional peaks are correctly forecast for SKU #8 and SKU #21. For SKUs #1 and #2, the impact of promotion has been higher than expected. The results show that the long-term share of purchases is achieved in two to three months after the introduction.

Training Set

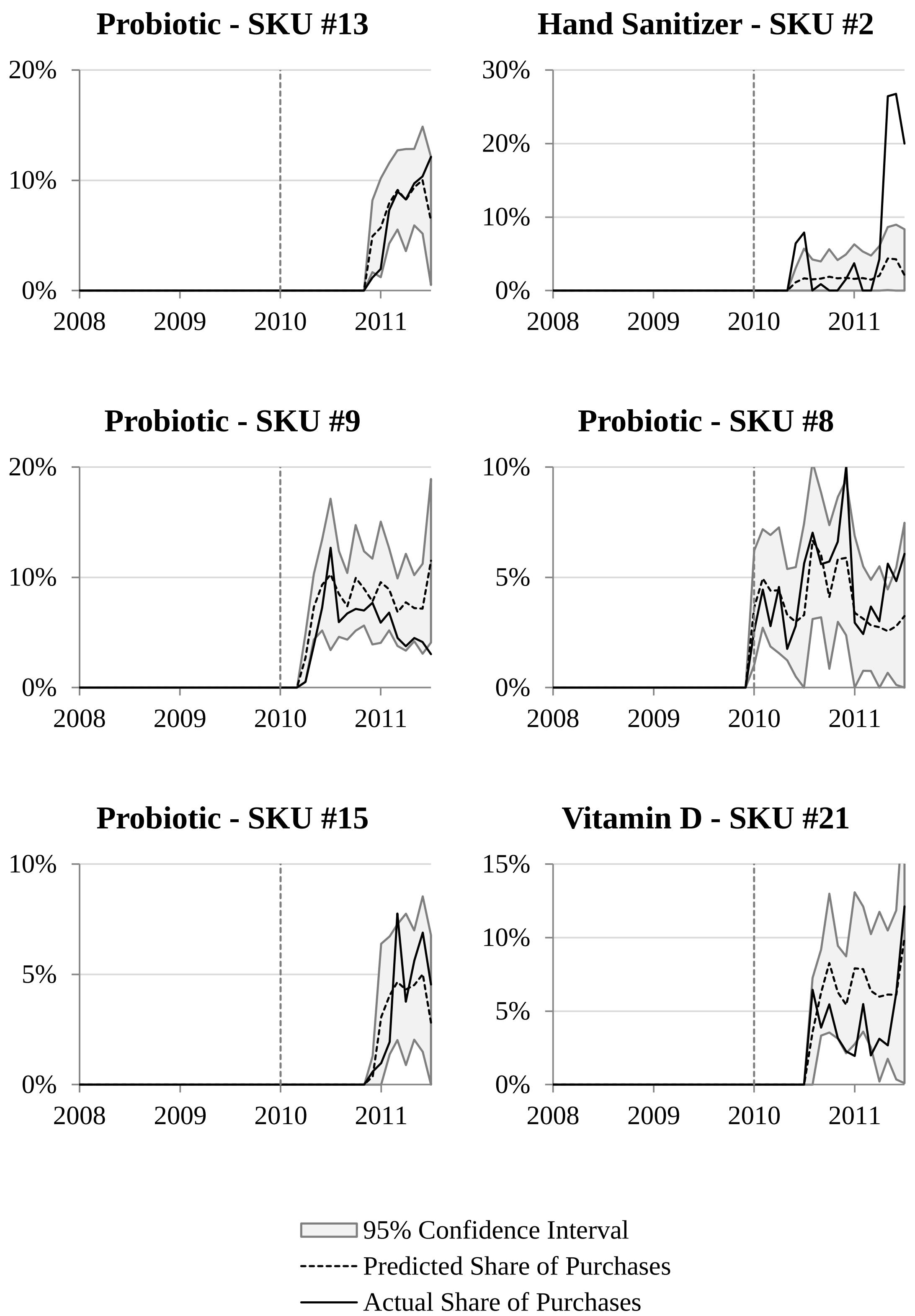

Figure 20: Actual and predicted share of purchases for new SKUs (1/2)

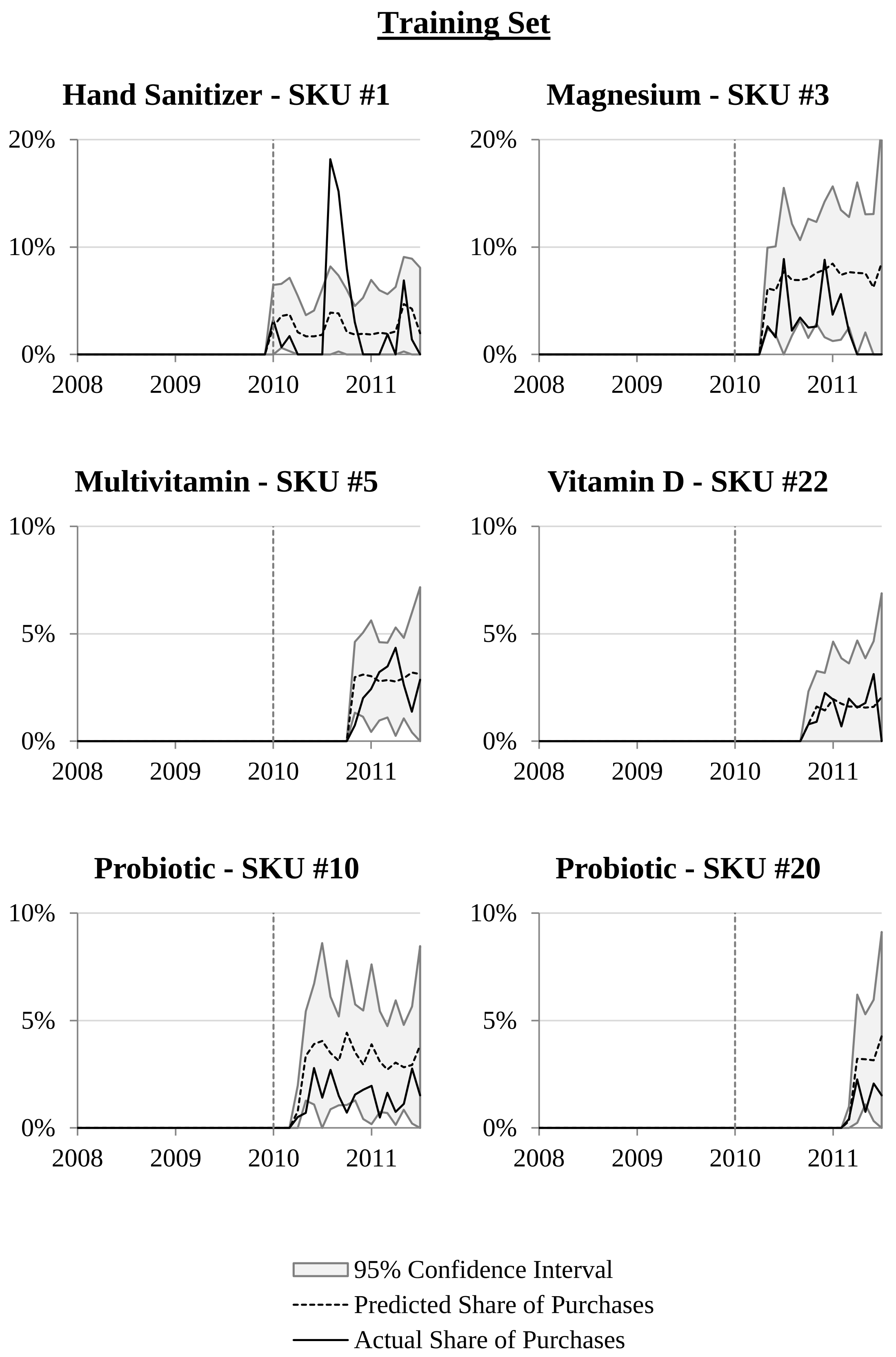

Figure 21: Actual and predicted share of purchases for new SKUs (2/2)

7 Discussion

7.1 Main Findings

The main objective of this thesis was to develop a multinomial logit model that would successfully explain consumers' purchase behaviour using only standard desktop applications and consumer-level point-of-sale (POS) data. We implemented both the SKU-specific GL model and the attribute level-specific FH model in a multinomial logit framework and compared their strengths in explaining different aspects of consumers' choice behaviour. We showed that most parameter estimates are statistically significant. The models were able to predict the purchase behaviour of consumers accurately both in the training set, which was used to calibrate the model, and the independent test set, which was used for evaluation purposes. Furthermore, we showed that the FH model is able to predict accurately the share of purchases for new products that enter the market in a future period.

We used consumer-level POS data to research consumers' purchase behaviour in retailing of health products. This data is automatically recorded by most POS terminals today and readily available for most retailers. We showed that a multinomial logit model with tens of thousands of observations can be run with a standard desktop computer using spreadsheet software, Microsoft Excel. The free Solver add-in was used to run the optimization algorithm. This shows that the applications of the multinomial logit model are now well within the reach of businesses without any additional costs associated with acquiring expensive external data or purchasing a specialized multinomial logit software.

We found that the FH model is a good alternative to the classical GL model. It provides deeper understanding of products and their attributes than the GL model. It is able to estimate the demand for new products before they are introduced to the market. The FH model also offers a more parsimonious estimation method with fewer coefficients, which is particularly important if the latent class approach is employed. However, using only one latent segment, the GL model gives more accurate estimates for the coefficients. This is because the one-segment FH model is merely a constrained version of the GL model. We showed that both models have their strengths and weaknesses. Which model is more appropriate depends on the context and aims of the modeller.

Parameter estimates of the model provide valuable insight into the consumers' purchase behaviour and responses to marketing stimuli. We showed that loyalties toward attributes, including brand, size, form and formula, were the most important factors determining consumers' choice of SKUs. Consistent with earlier research, brand loyalty had the greatest impact on choice. Form loyalty and formula loyalty were the second strongest explanatory variables, followed by size loyalty. SKU-specific previous purchase variables showed that repeat purchasing of the same SKUs was very common. Furthermore, we pointed out that a consumer who bought a product that was not promoted was more likely to repurchase the same product on the following shopping occasion than a consumer who bought a product that was promoted.

Marketing mix variables such as promotion, price and discount are important for the retailer because they are fully controllable. We showed that promotion had the largest impact on choice among marketing mix variables. Discount was a less frequently used marketing action, but highly significant for categories with enough observations on discounts. Regular price was the least significant marketing mix variable due to low frequency of price changes. Based on the results, the retailer can analyse the cost-effectiveness of marketing actions by comparing the impact of marketing stimuli with their costs.

We showed that the FH model is able to give accurate first-cut demand forecasts for new products that have attributes and attribute levels already found in the market. Our approach constitutes a very cost-effective option for new product planning because only readily available POS data is required. However, our model cannot forecast the demand of fundamentally different products that employ new attributes or attribute levels that are not previously found in the market. This limitation does not restrict the applicability of our model too much, as we showed that most new products that were introduced during the observation period were product line extensions.

7.2 Limitations and Suggestions

There are several possible ways to improve our model and further increase our understanding of consumer choice. Currently, our model is limited to explaining the consumers' choice of an SKU (what to buy). Incorporating the consumers' purchase incidence decision (whether to buy) and purchase quantity decision (how much to buy) would significantly expand our model's usability. These extensions would allow us to forecast the sales instead of only share. We could also determine the extent of category expansion due to introduction of new products and marketing activities.

In our model, consumers were treated as a homogeneous group. This means that the preferences toward attribute levels or responses to marketing activities did not vary across consumers. Heterogeneity in purchase probabilities is captured by the loyalty variables. We could improve our model by applying the latent class approach (Kamakura and Russell 1989). The latent class model identifies a finite number of homogeneous latent classes, or segments, that differ substantially in purchase behaviour with each other. Each of these segments has its own coefficients for preferences and marketing stimuli. Consumers are assigned to these segments with a probability of membership that depends on observed choice history, demographics and other covariates. The choice probability is given by a sum of the segment-specific choice probabilities weighted by the probabilities of the consumer belonging to the segment. Latent class model is shown to substantially improve the accuracy of prediction and is able to identify consumer segments for targeted marketing.

We assumed that the consumers' preferences and responses to marketing stimuli do not change over time. The only source of change in consumers' preferences came from the time-variant loyalty variables. However, tastes and habits of consumers change rapidly, particularly in the fast-growing product categories. During our observation period from January 2007 to June 2011, we have witnessed the increasing popularity of chewable tablets in probiotics and stronger formulas in vitamin D. Allowing for variation of parameters over time could reveal important information on trends in consumers' preferences (e.g., Lattin 1987, Mela et al. 1997).

In this thesis, we have studied six product categories independently of each other. However, consumers' purchase decisions across product categories may not be independent of each other. First, the choice to purchase a product may increase the probability that a product from another category is also purchased. For example, many e-commerce companies have introduced free shipping thresholds to ensure economic order sizes. These schemes encourage the consumer to buy more products from the other categories so that the order value reaches the free shipping threshold, making the shipment free of charge. Second, the choice probabilities of complementary or similarly branded products in different categories may be highly correlated. Examples of complementary health products include razor handles and blades, and electronic toothbrushes and replacement brush heads. Similarly branded products and brands spanning many categories are very common in health products. See Seetharaman et al. (2005) for a review on models of multiple-category choice.

Finally, retailers would benefit the most from a holistic approach to assortment planning and marketing. Joint optimization of assortment, pricing, promotion and advertising could produce interesting results on properties of optimal assortment and marketing policy. Assortment optimization models call for the correct characterization of demand for each product and accurate determination of substitution patterns across products. However, models that permit flexible substitution patterns are computationally very expensive. The complexity of these models can be reduced by using the more parsimonious attribute-based approaches. Combining an advanced discrete choice model with an attribute-based approach could show promise in assortment planning applications.

References

Ailawadi, K.L. and Neslin, S.A. 1998, "The effect of promotion on consumption: Buying more and consuming it faster". *Journal of Marketing Research,* vol. 35, no. 3, pp. 390-398.

Allenby, G.M. and Lenk, P.J. 1994, "Modeling household purchase behavior with logistic normal regression". *Journal of the American Statistical Association,* vol. 89, no. 428, pp. 1218-1231.

Allenby, G.M., Shively, T.S., Yang, S. and Garratt, M.J. 2004, "A choice model for packaged goods: Dealing with discrete quantities and quantity discounts". *Marketing Science,* vol. 23, no. 1, pp. 95-108.

Anderson, S.P., De Palma, A. and Thisse, J.F. 1992, *Discrete Choice Theory of Product Differentiation.* Cambridge, MA: MIT Press.

Arora, N., Allenby, G.M. and Ginter, J.L. 1998, "A hierarchical Bayes model of primary and secondary demand". *Marketing Science,* vol. 17, no. 1, pp. 29-44.

Arunotayanun, K. 2009, "Modelling freight supplier behaviour and response". Unpublished PhD thesis, Imperial College, London.

Bell, D.R., Chiang, J. and Padmanabhan, V. 1999, "The decomposition of promotional response: An empirical generalization". *Marketing Science,* vol. 18, no. 4, pp. 504-526.

Bell, D.R. and Lattin, J.M. 1998, "Shopping behavior and consumer preference for store price format: Why "large basket" shoppers prefer EDLP". *Marketing Science,* vol. 17, no. 1, pp. 66-88.

Bell, D.R., Bonfrer, A. and Chintagunta, P.K. 2005, "Recovering stockkeeping-unit-level preferences and response sensitivities from market share models estimated on item aggregates". *Journal of Marketing Research,* vol. 42, no. 2, pp. 169-182.

Ben-Akiva, M.E. 1973, "Structure of passenger travel demand models". Unpublished PhD thesis, Department of Civil Engineering, MIT, Cambridge, MA.

Ben-Akiva, M.E. and Lerman, S.R. 1985, *Discrete Choice Analysis: Theory and Application to Travel Demand.* Cambridge, MA: MIT Press.

Bhat, C.R. 1997, "Recent methodological advances relevant to activity and travel behavior analysis". *Resource paper presented at the IATBR Conference, Austin, Texas*.

Bhat, C.R., Eluru, N. and Copperman, R.B. 2000, "Flexible model structures for discrete choice analysis" in *Handbook of Transport Modelling,* vol. 1, pp. 71-90.

Boyd, J.H. and Mellman, R.E. 1980, "The effect of fuel economy standards on the US automotive market: An hedonic demand analysis". *Transportation Research Part A: General,* vol. 14, no. 5-6, pp. 367-378.

Bucklin, R.E. and Gupta, S. 1992, "Brand choice, purchase incidence, and segmentation: An integrated modeling approach", *Journal of Marketing Research,* vol. 29, no. 2, pp. 201-215.

Bucklin, R.E., Gupta, S. and Siddarth, S. 1998, "Determining segmentation in sales response across consumer purchase behaviors". *Journal of Marketing Research,* vol. 35, no. 2, pp. 189-197.

Bucklin, R.E. and Lattin, J.M. 1991, "A two-state model of purchase incidence and brand choice", *Marketing Science,* vol. 10, no. 1, pp. 24-39.

Cardell, N.S. and Dunbar, F.C. 1980, "Measuring the societal impacts of automobile downsizing". *Transportation Research Part A: General,* vol. 14, no. 5-6, pp. 423-434.

Chamberlain, G. 1980, "Analysis of covariance with qualitative data". *Review of Economic Studies*, vol. 47, no. 1, pp. 225-238.

Chiang, J. 1991, "A simultaneous approach to the whether, what and how much to buy questions". *Marketing Science,* vol. 10, no. 4, pp. 297-315.

Chib, S., Seetharaman, P. and Strijnev, A. 2004, "Model of brand choice with a no-purchase option calibrated to scanner-panel data". *Journal of Marketing Research,* vol. 41, no. 2, pp. 184-196.

Chintagunta, P.K. 1993, "Investigating purchase incidence, brand choice and purchase quantity decisions of households". *Marketing Science,* vol. 12, no. 2, pp. 184-208.

Chintagunta, P.K. and Dubé, J.P. 2005, "Estimating a stockkeeping-unit-level brand choice model that combines household panel data and store data". *Journal of Marketing Research*, vol. 42, no. 3, pp. 368-379.

Chintagunta, P.K., Jain, D.C. and Vilcassim, N.J. 1991, "Investigating heterogeneity in brand preferences in logit models for panel data". *Journal of Marketing Research,* vol. 28, no. 4, pp. 417-428.

Chong, J.K. and Ho, T.H. 2009, "Determining optimal assortment: Forecasting demand of new products and estimating substitution pattern". *Operations Research (Online Forum Commentary),* pp. 1.

Chong, J.K., Ho, T.H. and Tang, C.S. 2001, "A modeling framework for category assortment planning". *Manufacturing & Service Operations Management,* vol. 3, no. 3, pp. 191-210.

Daganzo, C. 1979, *Multinomial Probit: The Theory and Its Application to Demand Forecasting*. New York: Academic Press.

Daly, A. and Zachary, S. 1978, "Improved multiple choice models". *Determinants of Travel Choice,* vol. 335, pp. 357.

Debreu, G. 1960, "Review of RD Luce, Individual choice behavior: A theoretical analysis". *American Economic Review,* vol. 50, no. 1, pp. 186-188.

Decker, R. and Scholz, S.W. 2010, "Determining the attractiveness of product attributes in consumer goods markets using POS scanner data". *The Marketing Review,* vol. 10, no. 3, pp. 225-237.

DeSarbo, W.S., Ramaswamy, V. and Cohen, S.H. 1995, "Market segmentation with choice-based conjoint analysis". *Marketing Letters,* vol. 6, no. 2, pp. 137-147.

Dillon, W.R. and Gupta, S. 1996, "A segment-level model of category volume and brand choice". *Marketing Science,* vol. 15, no. 1, pp. 38-59.

Dodson, J.A., Tybout, A.M. and Sternthal, B. 1978, "Impact of deals and deal retraction on brand switching". *Journal of Marketing Research,* vol. 15, no. 1, pp. 72-81.

Erdem, T. 1996, "A dynamic analysis of market structure based on panel data". *Marketing Science,* vol. 15, no. 4, pp. 359-378.

Fader, P.S. and Hardie, B.G.S. 1996, "Modeling consumer choice among SKUs". *Journal of Marketing Research,* vol. 33, no. 4, pp. 442-452.

Fader, P.S. and Lattin, J.M. 1993, "Accounting for heterogeneity and nonstationarity in a cross-sectional model of consumer purchase behavior". *Marketing Science,* vol. 12, no. 3, pp. 304-317.

Fisher, M.L. and Vaidyanathan, R. 2009, "An algorithm and demand estimation procedure for retail assortment optimization", Working Paper, The Wharton School, PA.

Greene, W.H. and Zhang, C. 2003, *Econometric Analysis.* New Jersey: Prentice Hall.

Guadagni, P.M. and Little, J.D.C. 1983, "A logit model of brand choice calibrated on scanner data". *Marketing Science,* vol. 2, no. 3, pp. 203-238.

Guadagni, P.M. and Little, J.D.C. 1998, "When and what to buy: A nested logit model of coffee purchase". *Journal of Forecasting,* vol. 17, no. 3-4, pp. 303-326.

Guadagni, P.M. and Little, J.D.C. 2008, "Commentary—A logit model of brand choice calibrated on scanner data: A 25th anniversary perspective". *Marketing Science,* vol. 27, no. 1, pp. 26-28.

Gupta, S. 1988, "Impact of sales promotions on when, what, and how much to buy". *Journal of Marketing Research,* vol. 25, no. 4, pp. 342-355.

Halkidi, M. and Vazirgiannis, M. 2005, "Quality assessment approaches in data mining" in *The Data Mining and Knowledge Discovery Handbook*, eds. O. Maimon & L. Rokach. Boston, MA: Springer Verlag. pp. 661-696.

Hamel, L. 2009, *Knowledge Discovery with Support Vector Machines,* Wiley Online Library.

Hausman, J.A. and Wise, D.A. 1978, "A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences". *Econometrica: Journal of the Econometric Society,* vol. 46, pp. 403-426.

Hensher, D.A. and Greene, W.H. 2003, "The mixed logit model: The state of practice". *Transportation,* vol. 30, no. 2, pp. 133-176.

Hess, S. 2005, "Advanced discrete choice models with applications to transport demand" Unpublished PhD thesis, Imperial College, London.

Ho, T.H. and Chong, J.K. 2003, "A parsimonious model of stockkeeping-unit choice". *Journal of Marketing Research,* vol. 40, no. 3, pp. 351-365.

Hübner, A. 2011, *Retail category management: decision support systems for assortment, shelf space, inventory and price planning.* Berlin: Springer Verlag.

Jain, D.C. and Vilcassim, N.J. 1991, "Investigating household purchase timing decisions: A conditional hazard function approach". *Marketing Science,* vol. 10, no. 1, pp. 1-23.

Jedidi, K., Mela, C.F. and Gupta, S. 1999, "Managing advertising and promotion for long-run profitability". *Marketing Science,* vol. 18, no. 1, pp. 1-22.

Jones, J.M. and Landwehr, J.T. 1988, "Removing heterogeneity bias from logit model estimation", *Marketing Science,* vol. 7, no. 1, pp. 41-59.

Jones, J.M. and Zufryden, F.S. 1981, "Relating deal purchases and consumer characteristics to repeat purchase probability". *Journal of the Market Research Society,* vol. 23, no. 2, pp. 84-99.

Kalwani, M.U., Yim, C.K., Rinne, H.J. and Sugita, Y. 1990, "A price expectations model of customer brand choice". *Journal of Marketing Research,* vol. 27, no. 3, pp. 251-262.

Kamakura, W.A. and Russell, G.J. 1989, "A probabilistic choice model for market segmentation and elasticity structure". *Journal of Marketing Research,* vol. 26, no. 4, pp. 379-390.

Keane, M.P. 1997, "Modeling heterogeneity and state dependence in consumer choice behavior". *Journal of Business & Economic Statistics*, vol. 15, no. 3, pp. 310-327.

Kök, A.G., Fisher, M.L. and Vaidyanathan, R. 2009, "Assortment planning: Review of literature and industry practice" in *Retail Supply Chain Management*, eds. N.M. Agrawal & S.A. Smith. Boston, MA: Springer. pp. 1-55.

Krishnamurthi, L., Mazumdar, T. and Raj, S. 1992, "Asymmetric response to price in consumer brand choice and purchase quantity decisions". *Journal of Consumer Research*, vol. 19, no. 3, pp. 387-400.

Krishnamurthi, L. and Raj, S. 1988, "A model of brand choice and purchase quantity price sensitivities". *Marketing Science*, vol. 7, no. 1, pp. 1-20.

Lattin, J.M. 1987, "A model of balanced choice behavior". *Marketing Science*, vol. 6, no. 1, pp. 48-65.

Lattin, J.M. and Bucklin, R.E. 1989, "Reference effects of price and promotion on brand choice behavior". *Journal of Marketing Research*, vol. 26, no. 3, pp. 299-310.

Manski, C.F. 1977, "The structure of random utility models". *Theory and Decision*, vol. 8, no. 3, pp. 229-254.

McFadden, D. 1978, "Modelling the choice of residential location" in *Spatial Interaction Theory and Planning Models*, ed. A. Karlquist. Amsterdam: Elsevier Science. pp. 75-96.

McFadden, D. 1973, "Conditional logit analysis of qualitative choice behavior" in *Frontiers in Econometrics*, ed. P. Zarembka. New York: Barnes & Noble. pp. 105-142.

McFadden, D. and Train, K. 2000, "Mixed MNL models for discrete response". *Journal of Applied Econometrics*, vol. 15, no. 5, pp. 447-470.

Mehta, N. 2007, "Investigating consumers' purchase incidence and brand choice decisions across multiple product categories: A theoretical and empirical analysis". *Marketing Science*, vol. 26, no. 2, pp. 196-217.

Mela, C.F., Gupta, S. and Lehmann, D.R. 1997, "The long-term impact of promotion and advertising on consumer brand choice". *Journal of Marketing Research*, vol. 34, no. 2, pp. 248-261.

Mela, C.F., Jedidi, K. and Bowman, D. 1998, "The long-term impact of promotions on consumer stockpiling behavior". *Journal of Marketing Research*, vol. 35, no. 2, pp. 250-262.

Pauwels, K., Hanssens, D.M. and Siddarth, S. 2002, "The long-term effects of price promotions on category incidence, brand choice, and purchase quantity". *Journal of Marketing Research*, vol. 39, no. 4, pp. 421-439.

Refaeilzadeh, P., Tang, L. and Liu, H. 2009, "Cross-validation" in *Encyclopedia of Database Systems*, eds. M.T. Özsu & L. Liu. New York: Springer. pp. 532-538.

Revelt, D. and Train, K. 2000, "Customer-specific taste parameters and mixed logit: Households' choice of electricity supplier", Working Paper, University of California at Berkeley, Berkeley, CA.

Revelt, D. and Train, K. 1998, "Mixed logit with repeated choices: Households' choices of appliance efficiency level" *Review of Economics and Statistics*, vol. 80, no. 4, pp. 647-657.

Roy, R., Chintagunta, P.K. and Haldar, S. 1996, "A framework for investigating habits, "the hand of the past," and heterogeneity in dynamic brand choice". *Marketing Science,* vol. 15, no. 3, pp. 280-299.

Sándor, Z. and Wedel, M. 2002, "Profile construction in experimental choice designs for mixed logit models". *Marketing Science,* vol. 21, no. 4, pp. 455-475.

Seetharaman, P. 2004, "Modeling multiple sources of state dependence in random utility models: A distributed lag approach". *Marketing Science,* vol. 23, no. 2, pp. 263-271.

Seetharaman, P., Ainslie, A. and Chintagunta, P.K. 1999, "Investigating household state dependence effects across categories". *Journal of Marketing Research,* vol. 36, no. 4, pp. 488-500.

Seetharaman, P., Chib, S., Ainslie, A., Boatwright, P., Chan, T., Gupta, S., Mehta, N., Rao, V. and Strijnev, A. 2005, "Models of multi-category choice behavior". *Marketing Letters,* vol. 16, no. 3, pp. 239-254.

Shoemaker, R.W. and Shoaf, F.R. 1977, "Repeat rates of deal purchases". *Journal of Advertising Research,* vol. 17, no. 2, pp. 47-53.

Song, I. and Chintagunta, P.K. 2007, "A discrete-continuous model for multicategory purchase behavior of households". *Journal of Marketing Research,* vol. 44, no. 4, pp. 595-612.

Sun, B., Neslin, S.A. and Srinivasan, K. 2003, "Measuring the impact of promotions on brand switching when consumers are forward looking". *Journal of Marketing Research,* vol. 40, no. 4, pp. 389-405.

Swait, J. and Adamowicz, W. 1996, "The effect of choice environment and task demands on consumer behavior: Discriminating between contribution and confusion". Working Paper, Department of Rural Economy, University of Alberta, Edmonton, Canada.

Tellis, G.J. 1988, "Advertising exposure, loyalty, and brand purchase: A two-stage model of choice". *Journal of Marketing Research,* vol. 25, no. 2, pp. 134-144.

Train, K. 2003, *Discrete Choice Methods with Simulation.* Cambridge: Cambridge University Press.

Trinh, G., Dawes, J. and Lockshin, L. 2009, "Do product variants appeal to different segments of buyers within a category?". *Journal of Product & Brand Management*, vol. 18, no. 2, pp. 95-105.

Vovsha, P. 1997, "Application of cross-nested logit model to mode choice in Tel Aviv, Israel, metropolitan area". *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1607, pp. 6-15.

Vovsha, P. and Bekhor, S. 1998, "Link-nested logit model of route choice: Overcoming route overlapping problem". *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1645, pp. 133-142.

Williams, H.C.W.L. 1977, "On the formation of travel demand models and economic evaluation measures of user benefit". *Environment and Planning A*, vol. 9, no. 3, pp. 285-344.

Zhang, J. and Krishnamurthi, L. 2004, "Customizing promotions in online stores". *Marketing Science*, vol. 23, no. 4, pp. 561-578.