DENNIS FOK

Advanced Econometric Marketing Models

Advanced Econometric Marketing Models

ERIM Ph.D. Series Research in Management, 27

Erasmus Research Institute of Management (ERIM)
Erasmus University Rotterdam
Internet: http://www.erim.eur.nl

ISBN 90–5892–049-6

This thesis was prepared with LaTeX2e, all computations were done using Ox 3.30 (Doornik, 2002).

Advanced Econometric Marketing Models

Geavanceerde econometrische marketing modellen

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus

Prof.dr.ir. J.H. van Bemmel

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 6 november 2003 om 16:00 uur

door

Dennis Fok
geboren te Dordrecht

Promotiecommissie

Promotor:	Prof.dr. P.H.B.F. Franses
Overige leden:	Prof.dr.ir. G.H. van Bruggen Prof.dr. P.K. Chintagunta Prof.dr. M.G. Dekimpe
Copromotor:	Dr. R. Paap

Preface

Next to a challenging research topic there are many other things that contribute to a pleasant Ph.D. period. This is a good time to thank the human factors. First of all, I would like to thank my promotor and co-promotor, Philip Hans Franses and Richard Paap. Not only are they pleasant people to work with, they have been very important in my development as a researcher and for the research itself. Philip Hans is a never ending source of ideas and inspiration. Many of these ideas were related to my thesis, others were on completely different topics and provided a welcome distraction. Philip Hans also aroused my interest in other research fields, ranging from modeling employment figures to forecasting the number of citations to academic publications. Richard is an expert in many different areas of econometrics. Over the last four years we had many discussions on all kinds of topics, ranging from complex Bayesian sampling schemes to the best choice of audio equipment. Concerning econometrics, I learnt the most from the times when we did not directly agree. I thank Richard for taking the time to discuss things until at last I decided to accept his arguments, or (at fewer occasions) the other way around. I am looking forward to a continued cooperation with Philip Hans and Richard in the future.

Secondly, I thank the members of the small committee, that is, Gerrit van Bruggen, Pradeep Chintagunta and Marnik Dekimpe for evaluating this thesis. I am also grateful to a number of anonymous referees who have given detailed comments on earlier versions of many of the papers that form the basis of this thesis. I also thank Marnik Dekimpe for his detailed and valuable comments on Chapter 4.

During my Ph.D. track I had the opportunity to travel around the world to visit conferences, present my work, and to meet fellow researchers. During the third year, I spent three months at the Graduate School of Business at the University of Chicago. I would like to thank the marketing group of the Graduate School of Business for their hospitality, especially Pradeep Chintagunta and Peter Rossi. The financial support for these trips of ERIM, the Econometric Institute, and the Trustfonds are gratefully acknowledged.

I thank Erjen van Nierop who was willing to share an office with me in my first year, and Björn Vroomen who even had to put up with me for three years. I hope both of you had a good time as well. Next, I would like to thank several generations of fellow Ph.D.-students for sometimes serious and at other times hilarious coffee/tea sessions,

lunches and other social activities. To name just a few I mention, Björn, David, Erjen, Ivo, Jedid-Jah, Joost, Kevin, Klaas, Patrick, Pim, Richard, Richard, Rutger, Ward, and Wilco.

Last, but certainly not least, I want to thank my parents who have always been a big support and who motivate me to get the most out of the things I do.

Dennis Fok
Rotterdam, August 2003

Contents

Chapter 1

Introduction

1.1 General motivation

The present availability of large databases in marketing, concerning for example store level sales or individual purchases, has led to an increased demand for appropriate econometric models to deal with these data. The typical database contains information on revealed preferences, measured by for example sales, market shares or brand choices. Sometimes stated preferences such as opinions, attitudes and purchase intentions are also available to the researcher through surveys.

In general, data are available at two different levels of aggregation. The data either concern marketing measures at the aggregate level, that is, the total sales of a specific product in a single store, or they concern a panel of households. In these panels all purchases of a large number of households are tracked over a period of time. Additionally, data concerning the marketing efforts of all brands in a store are available as well. One usually has information on the actual prices of all products in all stores at each day and information on promotional activities, such as features and displays. One of the main goals of the application of econometric models in marketing is to gain insight into the effects of marketing instruments on the various marketing measures.

In this thesis we will discuss marketing models at both levels of aggregation, that is, at the household level and at the store level. The first part of this thesis deals with models on the aggregate level, in particular we consider various extensions of the market share attraction model. In the second part, we present models based on household panel data, that is, models for brand choice and models for the purchase timing of individual households. We however abstain from comparing the relative performance of the aggregate and the individual level analysis. Although there still is a debate in the marketing literature as to which level of analysis is to be preferred, we feel that both have their relative advantages. One of the main advantages of an aggregate level analysis are the

modest data requirements. For a study at the individual household level one needs to track a large number of households over a period of time. Collecting such data is rather expensive. However, based on these data one can obtain a more detailed view on the effects of promotions than with sales or market share data. With aggregate data it is for example difficult to see whether a sales increase is caused by more customers buying the product or by larger purchased quantities by the customers. Our view to this issue is that both data sources have their merits and the choice of the appropriate data depends on the research question.

Models have been developed for all kinds of marketing data, see for example Leeflang *et al.* (2000) and Franses and Paap (2001) for recent overviews of quantitative marketing models. However, there are still a large number of unresolved issues. Some of these issues deal with econometric aspects of existing marketing models, such as estimation, model selection and forecasting. In this thesis we present a thorough econometric analysis of the market share attraction model. Other unresolved issues deal with specific marketing-theoretic questions. Examples of such issues covered in this thesis are the effects of brand introduction, dynamic effects of promotions and structural heterogeneity.

Inference in these new or extended models is often not straightforward, and advanced econometric techniques, such as Bayesian inference, simulation and simulated maximum likelihood are necessary. The intention of this thesis is to provide a contribution to two, partly overlapping, fields of research, that is, the field of applied econometrics on the one hand and the field of quantitative marketing on the other hand. Below we give a specific introduction and motivation for each chapter separately.

1.2 Outline

In Part I of this thesis, we focus on aggregate market response models, that is, models for marketing measures at the level of a specific supermarket or a specific geographical region. The typical case for this type of research is the analysis of sales or market shares of various brands within a single category. Dependent upon the research goal, sales or market shares may be the preferred dependent variable. In case one is interested in the relative positioning of brands one will probably prefer the analysis of market shares. By using market shares, one also allows for seasonal fluctuations and category expansion in a natural way without actually having to specify these effects. In the three chapters of Part I we will consider market shares as the focal marketing measure.

In Chapter 2 we review the market share attraction model. This model has been a popular tool to analyze market shares for quite some time, see for example Cooper and Nakanishi (1988). Many papers in the marketing literature have used the attraction model. However, a thorough econometric analysis of this model is lacking. For example,

in most studies one specific attraction specification is chosen from the wide range of available specifications without formal testing. We show that a formal model selection strategy based on statistical tests improves the empirical performance of the attraction model, for in-sample fit as well as out-of-sample forecasting. Furthermore, many papers have studied the forecasting performance of the attraction model. However, due to the nonlinearity of the attraction model, forecasting market shares turns out not to be a trivial issue. We demonstrate that routinely made forecasts are biased. Unbiased forecasts can be obtained through simulation. Chapter 2 is an extended version of Fok *et al.* (2002a)

In Chapter 3 we extend the market share attraction model to allow for changes in the number of brands. Such a change may occur when a new brand is introduced into the market or when a brand is removed. We focus on the former event, although the analysis presented in this chapter can easily be applied to the case of the exit of a brand. As a consequence of the brand introduction several changes may take place affecting the competition among the incumbent brands. For example, some brand managers may decide to react to the introduction using marketing instruments. A large body of research has investigated the issue of optimal response to market entry in the context of a market share model. For example, Basuroy and Nguyen (1998) demonstrate that a brand introduction should lead to price decreases. Similar studies are found in Karnani (1985) and Gruca *et al.* (1992, 2001) among others. These studies are usually normative, and do not consider empirical data. Furthermore, one usually does not account for changes in household preferences or changes in the responses of households to marketing instruments.

We develop statistical tests with which we can explicitly test for changes in the competitive structure. That is, we model the period before and after the introduction in a single attraction model. The difficulty that arises is that the attraction model was originally developed for a constant number of brands. The advantage of capturing both periods in a single model is that we can easily perform statistical tests to see whether preferences or marketing responses have changed due to the entry, while correcting for the impact of the new brand's marketing mix and possible competitive reactions. Next to tests for changes in (aggregate) household behavior, we develop tests for possible changes in the use of the marketing mix by the incumbent brands. The results of these tests can then be used to validate the normative predictions available in the literature.

In our empirical application, we do not find evidence for the hypothesis developed in the game-theoretic literature that prices should decrease as a reaction to the introduction. Furthermore we find that some part of the competitive structure in the market changes, indicating that there are consumer reactions to the introduction.

In Chapter 4, the final chapter of Part I, we discuss the analysis of dynamic effects in the context of market shares. A flexible dynamic version of the market share model can easily be obtained by including lagged marketing-mix instruments and lagged mar-

ket shares in the attraction specification. However, the interpretation of the dynamical features of such a specification is not straightforward. We show which restrictions have to be imposed on this specification to yield an attraction model for which it is possible to derive interesting dynamical features. The necessary restrictions turn out to correspond well to attraction models found to be appropriate in practice.

As market shares and marketing-mix series are usually statistically stationary series (see for example Dekimpe and Hanssens (1995a) and Nijs *et al.* (2001)) a temporary promotion, for example a feature promotion or a price cut, cannot have a permanent effect on market shares. The relevant dynamic features of a model are therefore the direct, or short-run, effect and the cumulative, or long-run, effect of a temporary promotion. The attraction model can be rewritten into the so-called error-correction format, in which one can easily identify these two practically relevant features. Next to deriving an appropriate dynamic specification, we study the relation between the short and long-run effect and characteristics of the brand and the market. To this end, we develop a Hierarchical Bayes model in which the short-run and long-run effects are related to these characteristics in a second level of the model.

The resultant model is applied to a database of seven product categories in two distinct geographical areas. Our main finding is that, in absolute sense, the short-run price elasticity usually exceeds the long-run elasticity. Furthermore, we find strong correlations between the price elasticity and various brand and market characteristics. For example, the price elasticity tends to be larger, in absolute sense, for higher priced brands or brands that often issue coupons. This chapter is based on Fok *et al.* (2003).

Part II of this thesis also consists of three chapters. In these chapters we focus on marketing measures at the individual level. The typical data set on revealed preferences has the household as the unit of analysis. Therefore, we will refer to the household as the decision maker throughout Part II of this thesis. At the household level, three variables capture the purchase process, these are, purchase timing, brand choice, and purchased quantity, see for example Gupta (1988). Together, these three variables determine the market shares or sales of all brands in the market. By studying marketing measures at the disaggregated level one may obtain more insight in household behavior compared to a study based on sales or shares. In this thesis we will consider purchase timing and brand choice. We leave the modeling of purchase quantity to further research.

Chapters 5 and 6 both concern the modeling of interpurchase times. These two chapters are based on Fok and Paap (2003) and Fok *et al.* (2002b), respectively. Contrary to the previous chapters, we now study aspects of the purchase process that are related to the category level. Popular models to describe interpurchase timing are the Negative Binomial model, for the discrete case, and the hazard model in case time is measured on a continuous scale. The aim of these models is to infer the impact of household character-

istics and marketing instruments on the purchase timing of households. Variables in the first category are quite easy to include in a model. However, this is not the case for the marketing instruments of brands. As purchase timing is defined at the category level, the researcher has to somehow summarize the marketing efforts of the different brands into a category measure. This problem is studied in Chapter 5. In this chapter we elaborate on this issue and discuss the advantages and disadvantages of various strategies available in the literature. Furthermore, we suggest some new alternatives that are based on summarizing the marketing mix of individual brands using brand preference probabilities obtained through a choice model. Our findings suggest that the commonly used method of using household-specific choice shares to aggregate the brand-level marketing mix performs quite well. However, this method is not suitable when the focus of study is on out-of-sample forecasting. In that case, the newly proposed "latent preference purchase timing" model performs best.

In Chapter 6, we study the dynamic aspects of purchase timing. In the literature, consecutive interpurchase spells are assumed to be independent. However, based on findings in, for example, the areas of consumer behavior and sales and market share modeling, one expects correlation over time to be present. For example, consider a household persuaded to make a purchase in a category earlier than originally planned, for example due to a large price promotion. If the usage rate of the product is not affected, it will take the household a longer time before the extra stock of the product is depleted. Therefore, in this example, there is a negative correlation between interpurchase times. We develop a dynamic version of the popular hazard model and we show how to derive the dynamic properties of this model, thereby allowing for an intuitive and straightforward interpretation of the model parameters.

Our empirical analysis of the purchase timing in three different product categories indicates that there are significant dynamic effects in interpurchase times. Concerning the dynamic effects of marketing instruments, we find substantial differences across these categories.

Chapter 7 is a completely revised version of Fok *et al.* (2001). In this chapter we study the brand choice decision of households. After having decided to make a purchase in a specific category, the household must choose the brand to buy. However, households may differ in the decision process they use to make this choice. The decision process used may even differ over time for the same household.

We consider two different decision processes. Under the first decision process households actively compare the brands using price and take into account promotional activities. We coin this decision process "responsive to marketing efforts". Under the alternative decision process the household invests less effort. In this case the household does not take into account the promotional activities and does not pay attention to promotional pricing.

Instead, the brand choice is driven by base preferences and state dependence, where state dependence refers to a household's tendency to buy the same brand as bought on the previous shopping occasion. The actual decision process used at a specific purchase occasion cannot be observed, instead, this has to be inferred from the data. The probability of being in the "responsive state" is related to observable characteristics, such as household income, interpurchase time and the amount of money spent at the shopping trip.

Next to the structural heterogeneity it is important to capture differences in base preferences across households. For the responsiveness model we choose to model the base preferences using the normal distribution. This however does complicate the estimation procedure as for such a model there is no closed-form expression for the likelihood. To obtain parameter estimates, we maximize an approximation of the likelihood, where the approximation is obtained using simulation. This procedure is known as simulated maximum likelihood. However, to obtain an accurate approximation one needs many simulation draws. To improve the accuracy of the sampler, we propose using importance sampling. Our results show that this reduces the number of draws to a large extent.

In the empirical application we found that the responsiveness model fits rather well on observed brand choices in the detergent category. In fact the model outperforms various heterogenous variants of the commonly used multinomial logit model. The main behavioral conclusions are that most households act responsive to marketing efforts, and households buying large volumes of detergent or households buying many items on the same shopping trip tend to be less responsive.

In Chapter 8 we conclude the thesis with a brief overview of the findings. In this chapter we will also present some of the implications of our research and outline some topics for further research.

Part I

Aggregate market response models

Chapter 2

Modeling market shares

2.1 Introduction

In this chapter we will consider the econometric analysis of a popular model in marketing research, which is the market share attraction model. This model is typically considered for data on market shares, where the data have been collected at a weekly or monthly interval. Compared to models for sales, market share models are useful for categories which show instability (for example, due to rapid expansion), for which it is difficult to compare sales figures, whereas market shares provide a natural way of performing comparative analysis. Some studies consider the vector autoregression [VAR] model to describe market shares, like Takada and Bass (1998) and Srinivasan *et al.* (2000). These models do not explicitly impose the restriction that market shares sum to unity and that market shares of individual brands are between zero and one. If one wants to impose this, one usually ends up considering the attraction model.

Market share attraction models are seen as useful tools for analyzing competitive structures, see Cooper and Nakanishi (1988) and Cooper (1993), among various others. The models can be used to infer cross-effects of marketing-mix variables, but one can also learn about the effects of own efforts while conditioning on competitive reactions. The econometric analysis of the attraction model is complicated by the logical consistency feature of the model, that is that the model rightfully assumes that market shares sum to unity and that the market shares of individual brands are in between zero and one. However, an attraction model can be written as a system of equations concerning all market shares, and the parameters can then be estimated using standard methods, see for example Cooper (1993) and Bronnenberg *et al.* (2000).

Interestingly, a casual glance at the relevant marketing literature on market share attraction models indicates that there seems to have been little attention to how to specify the attraction model, how to estimate its parameters, how to analyze it virtues in the sense

that the models capture the salient data characteristics, and about how to use the models for forecasting. In sum, it seems that an (empirical) econometric view in these models is lacking. Therefore, in this chapter we aim to contribute to this view by addressing these issues concerning attraction models when they are to be used for describing and forecasting market shares. The first issue concerns the specification of the models. A literature check immediately indicates that many studies simply assume one version of an attraction model to be relevant and start from there. In this chapter we first start with a fairly general and comprehensive attraction model, and we show how various often applied models fit into this general framework. We also indicate how one can arrive from the general model at the more specific models, thereby immediately suggesting a general-to-simple testing strategy. Second, we discuss the estimation of the model parameters. We show that a commonly advocated method is unnecessarily complicated and that a much simpler method yields equivalent estimates. Finally, we address the issue of generating forecasts for market shares. As the market share attraction model ultimately gets analyzed as a system of equations for (natural) log transformed shares, generating unbiased forecasts is far from trivial. We discuss a simulation-based method which yields unbiased forecasts.

The outline of this chapter is as follows. In Section 2.2, we first discuss the basics of the attraction model by reviewing various specifications of the model. We discuss the interpretation of the model in Section 2.3, and we discuss parameter estimation of the model in Section 2.4. The topic of model selection is discussed in Section 2.5. Forecasting market shares with the attraction model is presented in Section 2.6. Some of the practical implications of our model selection strategy and the forecasting method are discussed in Section 2.7. In Section 2.8, we illustrate some of the techniques using scanner data. We conclude in Section 2.9.

2.2 Representation

In this section we start off with discussing a general market share attraction model and we deal with various of its nested versions which currently appear in the academic marketing literature. We first start with the so-called fully extended attraction model in Section 2.2.1. This model has a flexible structure as it includes many variables. Naturally this increases the empirical uncertainty about the relevant parameters. Therefore, in practice one may want to consider restricted versions of this general model. In Section 2.2.2, we discuss some of the restricted versions, where we particularly focus on those models which are often applied in practice.

2.2.1 A general market share attraction model

Let A_{it} be the attraction of brand i at time t, $t = 1, \ldots, T$, given by

$$A_{it} = \exp(\mu_i + \varepsilon_{it}) \prod_{j=1}^{I} \prod_{k=1}^{K} x_{kjt}^{\beta_{kji}} \quad \text{for } i = 1, \ldots, I, \tag{2.1}$$

where x_{kjt} denotes the k-th explanatory variable (such as price level, distribution, advertising spending) for brand j at time t and where β_{kji} is the corresponding coefficient for brand i. The parameter μ_i is a brand-specific constant. Let the error term $(\varepsilon_{1t}, \ldots, \varepsilon_{It})'$ be normally distributed with zero mean and Σ as a possibly non-diagonal covariance matrix, see Cooper and Nakanishi (1988). As we want the attraction to be non-negative, x_{kjt} has to be non-negative, and hence rates of changes are usually not allowed. The variable x_{kjt} may be a 0/1 dummy variable to indicate promotional activities for brand j at time t. Note that for this dummy variable, one should transform x_{kjt} to $\exp(x_{kjt})$ to avoid that A_{it} becomes zero in case of no promotional activity.

The attraction specification in (2.1) is known as the Multiplicative Competitive Interaction [MCI] specification. A more general version of the attraction model uses a transformation of the explanatory variables, that is, it uses $f(x_{kjt})$ instead of x_{kjt}. When $f(\cdot)$ is taken to be the exponential function one obtains a specification known as the Multinomial Logit [MNL] specification. The difference between the MCI and the MNL specification is the assumed pattern of the elasticity of marketing instruments. The MCI specification assumes that the elasticity declines with increasing values of the explanatory variable, while the MNL specification assumes that the elasticity increases up to a specific level and then decreases. The ultimate choice of a specification therefore depends on the marketing instruments used. The MNL specification seems to be appropriate for advertising spending, while the MCI specification would better fit pricing, see Cooper (1993) or Cooper and Nakanishi (1988) for elaborate discussions on the choice of $f(\cdot)$. In order not to complicate matters, we only consider the MCI specification, but note that all results can be extended to the MNL specification.

The market shares for the I brands follow from the, what is called, Market Share Theorem, see Bell *et al.* (1975). This theorem states that the market share of brand i is equal to its attraction relative to the sum of all attractions, that is,

$$M_{it} = \frac{A_{it}}{\sum_{j=1}^{I} A_{jt}} \quad \text{for } i = 1, \ldots, I. \tag{2.2}$$

The model in (2.1) with (2.2) is usually called the market share attraction model. Notice that the definition of the market share of brand i at time t given in (2.2) implies that the attraction of the product category is the sum of the attractions of all brands and that $A_{it} = A_{lt}$ results in $M_{it} = M_{lt}$.

An interesting aspect of the attraction model is that the A_{it} in (2.1) is unobserved. As we will see below, this implies that not all the brand intercepts (μ_i) and not all of the marketing-mix parameters (β_{kji}) are identified. As we will indicate, there are many possible model specifications for the attraction of a brand. For example, to describe potential dependencies in market shares over time, which represents purchase reinforcement effects, one may include lagged attractions A_{it} in (2.1). For example, one may consider

$$A_{it} = \exp(\mu_i + \varepsilon_{it}) A_{i,t-1}^{\gamma_i} \prod_{j=1}^{I} \prod_{k=1}^{K} x_{kjt}^{\beta_{kji}}. \tag{2.3}$$

However, due to the fact that we do not observe A_{it}, it turns out only possible to estimate the parameters in this model if the lag parameter γ_i is assumed to be the same across brands, see Chen *et al.* (1994). As this may be viewed as too restrictive, an alternative strategy to account for dynamics is to include lagged values of the observed variables M_{jt} and x_{kjt} in (2.1). The most general autoregressive structure follows from the inclusion of lagged market shares and lagged explanatory variables of all brands. In that case, the attraction specification with a P-th order autoregressive structure becomes

$$A_{it} = \exp(\mu_i + \varepsilon_{it}) \prod_{j=1}^{I} \left(\prod_{k=1}^{K} x_{kjt}^{\beta_{kji}} \prod_{p=1}^{P} \left(M_{j,t-p}^{\alpha_{pji}} \prod_{k=1}^{K} x_{kj,t-p}^{\beta_{pkji}} \right) \right), \tag{2.4}$$

where the α_{pji} parameters represent the effect of lagged market shares on attraction and where the β_{pkji} parameters represent the effect of lagged explanatory variables. To illustrate, this model allows that the market share for brand 1 at $t-1$ has an effect on that of brand 2 at t, and also that there is a relationship between brand 2's market share and the price of brand 1 at $t-1$. The lagged endogenous variables capture dynamics in purchase behavior that cannot be attributed to specific marketing instruments. For example, consider state dependence in behavior. If brand i is purchased at time t by consumers who act state dependent, there will be a higher probability that they will purchase brand i again at time $t+1$. Whether the brand was chosen at time t because it was promoted or just by chance does not influence the dynamics in the behavior. On the other hand, part of the dynamics in the behavior can be attributed to specific marketing instruments. As an example, consider price promotions. A well-known feature of promotions is the post-promotional dip, see van Heerde *et al.* (2000). In the period after a promotion it is often observed that sales or market shares decrease temporarily, as due to the promotion there has been stock piling by the consumers. To capture such dynamic patterns we include lagged exogenous variables in our attraction specification.

The flexibility of this general specification is reflected by the potentially large number of parameters. For example with $I = 4$ brands, $K = 3$ explanatory variables and $P = 2$ lags, there are over 150 parameters to estimate (although they are not all identified,

see below). It is however not necessary that the order P for the lagged market shares and lagged explanatory variables is the same. To obtain a different lag order for the explanatory variables, one can restrict the corresponding β_{pkji} parameters to be zero.

The model that consists of equations (2.4) and (2.2) is sometimes called the fully extended multiplicative competitive interaction [FE-MCI] model, see Cooper (1993). To enable parameter estimation, one can linearize this model in two steps. First, one can take one brand as the benchmark brand. Choosing brand I as the base brand yields

$$\frac{M_{it}}{M_{It}} = \frac{\exp(\mu_i + \varepsilon_{it}) \prod_{j=1}^{I} \left(\prod_{k=1}^{K} x_{kjt}^{\beta_{kji}} \prod_{p=1}^{P} \left(M_{j,t-p}^{\alpha_{pji}} \prod_{k=1}^{K} x_{kj,t-p}^{\beta_{pkji}} \right) \right)}{\exp(\mu_I + \varepsilon_{It}) \prod_{j=1}^{I} \left(\prod_{k=1}^{K} x_{kjt}^{\beta_{kjI}} \prod_{p=1}^{P} \left(M_{j,t-p}^{\alpha_{pjI}} \prod_{k=1}^{K} x_{kj,t-p}^{\beta_{pkjI}} \right) \right)}. \tag{2.5}$$

In Section 2.4.2, we will discuss another approach to linearizing the model, but we will show that both transformations lead to the same parameter estimates, while the estimation procedure based on (2.5) is much simpler. Next, one can take the natural logarithm (denoted by ln) of both sides of (2.5). Together, this results in the $(I-1)$-dimensional set of equations given by

$$\begin{aligned} \ln M_{it} - \ln M_{It} = (\mu_i - \mu_I) + \sum_{j=1}^{I} \sum_{k=1}^{K} (\beta_{kji} - \beta_{kjI}) \ln x_{kjt} + \\ \sum_{j=1}^{I} \sum_{p=1}^{P} \left((\alpha_{pji} - \alpha_{pjI}) \ln M_{j,t-p} + \sum_{k=1}^{K} (\beta_{pkji} - \beta_{pkjI}) \ln x_{kj,t-p} \right) + \eta_{it}. \end{aligned} \tag{2.6}$$

for $i = 1, \ldots, I-1$. Note that not all μ_i parameters ($i = 1, \ldots, I$) are identified. Also for each k and p, one of the β_{kji} and β_{pkji} parameters is not identified. In fact, only the parameters $\tilde{\mu}_i = \mu_i - \mu_I$, $\tilde{\beta}_{kji} = \beta_{kji} - \beta_{kjI}$ and $\tilde{\beta}_{pkji} = \beta_{pkji} - \beta_{pkjI}$ are identified. This is however sufficient to completely identify elasticities, see Section 2.3 below and Cooper and Nakanishi (1988, p. 145). Finally, one can only estimate the parameters $\tilde{\alpha}_{pji} = \alpha_{pji} - \alpha_{pjI}$.

The error variables in (2.6) are $\eta_{it} = \varepsilon_{it} - \varepsilon_{It}$, $i = 1, \ldots, I-1$. Hence, given the earlier assumptions on ε_{it}, $(\eta_{1t}, \ldots, \eta_{I-1,t})'$ is normally distributed with mean zero and the $((I-1) \times (I-1))$-dimensional covariance matrix $\tilde{\Sigma} = L \Sigma L'$, where $L = (\mathbf{I}_{I-1} \vdots -\mathbf{i}_{I-1})$ with $\mathbf{I}_{I-1}$ an $(I-1)$-dimensional identity matrix and where $\mathbf{i}_{I-1}$ is an $(I-1)$-dimensional unity vector. Note that therefore only $\frac{1}{2} I(I-1)$ parameters of the covariance matrix Σ can be identified.

In sum, the general attraction model can be written as a $(I-1)$-dimensional P-th order vector autoregression with exogenous variables [sometimes abbreviated as VARX(P)],

given by

$$\ln M_{it} - \ln M_{It} = \tilde{\mu}_i + \sum_{j=1}^{I}\sum_{k=1}^{K} \tilde{\beta}_{kji} \ln x_{kjt} + \sum_{j=1}^{I}\sum_{p=1}^{P} \left(\tilde{\alpha}_{pji} \ln M_{j,t-p} + \sum_{k=1}^{K} \tilde{\beta}_{pkji} \ln x_{kj,t-p} \right) + \eta_{it}, \quad (2.7)$$

$i = 1, \ldots, I-1$, where the covariance matrix of the error variables $(\eta_{1t}, \ldots, \eta_{I-1,t})'$ is $\tilde{\Sigma}$. Note that the model is only valid for the observations starting at time $t = P+1$. For inference, it is common practice to condition on the first P initial values of the log market shares and the explanatory variables as is also done in vector autoregressions, see Lütkepohl (1993). For further reference, we will consider (2.7) as the general attraction specification. We will take it as a starting point in our within-sample model selection strategy, which follows the general-to-specific principle, see Section 2.5 below.

It is not possible to write the log market shares in (2.7) as a function of current and lagged explanatory variables and disturbances only. It is even not possible to solve (2.7) for $\ln M_{it} - \ln M_{It}$. This is mainly due to the complex dynamic structure. This means that it is difficult to derive restrictions for stationarity of the log market shares themselves. In practice, this may not be a serious problem. Indeed, Srinivasan and Bass (2000) and Franses *et al.* (2001) consider testing for unit roots in market shares in a different model and their results suggest that generally market shares appear to be stationary. In Chapter 4 we will discuss modeling and interpreting dynamics in market share models in detail.

2.2.2 Various restricted models

As can be understood from (2.7), the general attraction model contains many parameters and in practice this will absorb many degrees of freedom. Therefore, one usually assumes a simplified version of this general model. Obviously, the general model can be simplified in various directions, and, interestingly, the academic marketing literature indicates that in many cases one simply assumes some form without much further discussion. Selecting an appropriate model may be a non-trivial exercise, as there are many possible simpler models. One can for example impose restrictions on the β coefficients, on the covariance structure Σ, and on the autoregressive parameters α. In this section we will discuss a few of these potentially empirically relevant restrictions on the attraction specification in (2.4).

Restricted Covariance Matrix [RCM]

If the covariance matrix of the error variables ε_{it} in (2.4) is a diagonal matrix, where each ε_{it} has its own variance σ_i^2, that is, $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_I^2)$, then the covariance matrix for the $(I-1)$-dimensional vector η_{it} in (2.7) becomes

$$\text{diag}(\sigma_1^2, \ldots, \sigma_{I-1}^2) + \sigma_I^2 \mathbf{i}_{I-1} \mathbf{i}_{I-1}', \tag{2.8}$$

where $\mathbf{i}_{I-1}$ denotes a $(I-1)$-dimensional unity vector. In Section 2.5 we discuss how one can examine the validity of (2.8). If this restriction holds, the errors in the attraction specifications are independent, implying that the unexplained components of the attraction equations are uncorrelated.

Restricted Competition [RC]

One can also assume that the attraction of brand i only depends on its own explanatory variables. This amounts to the assumption that marketing effects of competitive brands do not have an attraction effect, see for example Kumar (1994) among others. For (2.4), this corresponds to the restriction $\beta_{kji} = 0$ (and $\beta_{pkji} = 0$) for $j \neq i$. More precisely, this RC restriction implies that (2.4) reduces to

$$A_{it} = \exp(\mu_i + \varepsilon_{it}) \prod_{k=1}^{K} x_{kit}^{\beta_{ki}} \prod_{j=1}^{I} \prod_{p=1}^{P} \left(M_{j,t-p}^{\alpha_{pji}} \prod_{k=1}^{K} x_{ki,t-p}^{\beta_{pki}} \right) \quad \text{for } i = 1, \ldots, I, \tag{2.9}$$

where we write β_{ki} for β_{kii} and β_{pki} for β_{pkii}. Consequently, the linearized multiple equation model in (2.7) becomes

$$\begin{aligned} \ln M_{it} - \ln M_{It} = \tilde{\mu}_i + \sum_{k=1}^{K} \beta_{ki} \ln x_{kit} - \sum_{k=1}^{K} \beta_{kI} \ln x_{kIt} + \\ \sum_{j=1}^{I} \sum_{p=1}^{P} \left(\tilde{\alpha}_{pji} \ln M_{j,t-p} + \sum_{k=1}^{K} \beta_{pki} \ln x_{ki,t-p} - \sum_{k=1}^{K} \beta_{pkI} \ln x_{kI,t-p} \right) + \eta_{it} \end{aligned} \tag{2.10}$$

for $i = 1, \ldots, I-1$. Notice that this means that the coefficients β_{kI} are equal across the $(I-1)$ equations and that these restrictions should be taken into account when estimating the parameters. The RC assumption in (2.9) imposes $K(P+1)I(I-2)$ restrictions on the parameters in the general model in (2.7), which amounts to a substantial increase in the degrees of freedom. In Section 2.5 we will discuss how this restriction can be tested.

Restricted Effects [RE]

An even further simplified model arises if we assume, additional to RC, that the β parameters are the same for each brand, that is, $\beta_{ki} = \beta_k$ (and $\beta_{pki} = \beta_{pk}$), see Danaher

(1994) for an implementation of this combined restrictive model. This model assumes that marketing efforts for brand i only have an effect on the market share of brand i, and also that these effects are the same across brands. In other words, price effects, for example, are the same for all brands. It should be noted here that these similarities do not hold for *elasticities*, as will become apparent in Section 2.3. One may coin this model as an attraction model with restricted effects. Based on (2.4), the attraction for brand i at time t then further simplifies to

$$A_{it} = \exp(\mu_i + \varepsilon_{it}) \prod_{k=1}^{K} x_{kit}^{\beta_k} \prod_{j=1}^{I} \prod_{p=1}^{P} \left(M_{j,t-p}^{\alpha_{pji}} \prod_{k=1}^{K} x_{ki,t-p}^{\beta_{pk}} \right) \qquad \text{for } i = 1, \ldots, I, \tag{2.11}$$

and the linearized multiple equation model (2.7) simplifies to

$$\begin{aligned} \ln M_{it} - \ln M_{It} = \tilde{\mu}_i + \sum_{k=1}^{K} \beta_k (\ln x_{kit} - \ln x_{kIt}) + \\ \sum_{j=1}^{I} \left(\sum_{p=1}^{P} \tilde{\alpha}_{pji} \ln M_{j,t-p} + \sum_{k=1}^{K} \beta_{pk} (\ln x_{ki,t-p} - \ln x_{kI,t-p}) \right) + \eta_{it} \end{aligned} \tag{2.12}$$

for $i = 1, \ldots, I-1$. This RE assumption imposes an additional $K(P+1)(I-1)$ parameter restrictions on the β coefficients of (2.7). Of course, it may occur that the restrictions only hold for a few and not for all β_{kji} parameters, that is, for only a few marketing variables. In that case, less parameter restrictions should be imposed.

Restricted and Common Dynamics [RD, CD]

Finally, one may want to impose restrictions on the autoregressive structure in (2.4), implying that the purchase reinforcement effects are the same across brands. For example, the restriction that the attraction of brand i at time t only depends on its own lagged market shares M_{it} corresponds with the restriction $\alpha_{pji} = 0$ for $j \neq i$ in (2.4). The corresponding multivariate model, representing an attraction model with Restricted Dynamics [RD], then becomes

$$\begin{aligned} \ln M_{it} - \ln M_{It} = \tilde{\mu}_i + \sum_{j=1}^{I} \sum_{k=1}^{K} \tilde{\beta}_{kji} \ln x_{kjt} + \\ \sum_{j=1}^{I} \sum_{p=1}^{P} \left(\alpha_{pi} \ln M_{j,t-p} - \alpha_{pI} \ln M_{I,t-p} + \sum_{k=1}^{K} \tilde{\beta}_{pkji} \ln x_{kj,t-p} \right) + \eta_{it}, \end{aligned} \tag{2.13}$$

for $i = 1, \ldots, I-1$, where we again save on notation by using α_{pi} instead of α_{pii}. Note that now the α_{pI} parameters are the same across the $(I-1)$ equations and hence that

these restrictions should be imposed when estimating the model parameters. To illustrate, Chen *et al.* (1994) additionally impose that $P = 1$ and $\alpha_{1i} = \gamma$, which yields the estimable version of the attraction model in (2.3) which assumes that the purchase reinforcement effects are the same across brands. For further reference, we will call this last restriction the Common Dynamics [CD] restriction. It turns out that under this restriction it is possible to derive the dynamical properties of the attraction model. This restriction will be used in Chapter 4 where we explicitly consider the dynamics in market shares, that is, we will consider the long-run and the short-run effects of the marketing mix on shares.

The above discussion shows that various attraction models, which are considered in the relevant literature and in practice for modeling and forecasting market shares, are nested within the general attraction model in (2.4). In Table 2.1 we summarize the different parameter restrictions and we mention which studies use these restrictions. In the literature a model specification is usually selected a priori. However, the fact that the models are nested automatically suggests that an empirical model selection strategy can be based on a general-to-simple strategy, see also Section 2.5.

2.3 Interpretation

As the market shares get modeled through the attraction specification, and as this implies a reduced form of the model where parameters represent the impact of marketing efforts on the logarithm of relative market shares, the parameter estimates themselves are not easy to interpret. To facilitate an easier interpretation, one usually resorts to elasticities. In fact, it turns out that the reduced-form parameters are sufficient to identify these (cross-)elasticities.

For model (2.4), the instantaneous elasticity of the k-th marketing instrument of brand j on the market share of brand i is given by

$$\frac{\partial M_{it}}{\partial x_{kjt}} \frac{x_{kjt}}{M_{it}} = \beta_{kij} - \sum_{r=1}^{I} M_{rt} \beta_{krj}, \tag{2.14}$$

see Cooper (1993). To show that these elasticities are identified, one can rewrite them such that they only depend on the reduced-form parameters, that is,

$$\frac{\partial M_{it}}{\partial x_{kjt}} \frac{x_{kjt}}{M_{it}} = (\beta_{kji} - \beta_{kjI})(1 - M_{it}) - \sum_{r=1 \wedge r \neq i}^{I-1} M_{rt} (\beta_{kjr} - \beta_{kjI}), \tag{2.15}$$

see (2.6). Under Restricted Competition, these elasticities simplify to

$$\frac{\partial M_{it}}{\partial x_{kjt}} \frac{x_{kjt}}{M_{it}} = (\delta_{i=j} - M_{jt}) \beta_{kj}, \tag{2.16}$$

Table 2.1: Attraction model specifications used in the literature

Model	Lag	Restrictions on*				Literature
		Dynamics	Covariance matrix	Exogenous	Lagged exogenous	
I	1	RD	NR	RC	NI	Leeflang and Reuyl (1984) Danaher (1994)
II	1	CD	NR	RC	NI	Naert and Weverbergh (1981) Brodie and Bonfrer (1994) Brodie and de Kluyver (1984) Chen *et al.* (1994) Kumar (1994)
III	0	-	NR	RC	NI	Chen *et al.* (1994) Ghosh *et al.* (1984)
IV	1	CD	NR	RE	NI	Naert and Weverbergh (1981) Brodie and de Kluyver (1984) Leeflang and Reuyl (1984) Chen *et al.* (1994) Kumar (1994)
V	0	-	NR	RE	NI	Chen *et al.* (1994)

* RD=restricted dynamics, CD=common dynamics, RC=restricted competition, RE=restricted effects, NR=no restrictions, NI=not included

where $\delta_{i=j}$ is the Kronecker δ which has a value of 1 if i equals j and 0 otherwise. Under Restricted Effects, we simply have

$$\frac{\partial M_{it}}{\partial x_{kjt}} \frac{x_{kjt}}{M_{it}} = (\delta_{i=j} - M_{jt})\beta_k. \tag{2.17}$$

It is easy to see that the elasticities converge to zero if the market share goes to 1. From a marketing perspective, this seems rather plausible. If a brand controls almost the total market, its marketing efforts will have little if any effect on its market share. Secondly, in case the market share is an increasing function of instrument X, then if X goes to infinity the elasticity will go to 0. These two properties may seem straightforward, but among the best known market share models, the attraction model is the only model satisfying these properties, see also Cooper (1993). Whether the above two properties hold in a practical attraction model depends on the specific transformation of variables used, although the MCI and the MNL specification both lead to elasticities satisfying these properties.

2.4 Parameter estimation

In this section we discuss two methods for parameter estimation, and we show that they are equivalent. The first method is rather easy, whereas the second (which seems to be commonly applied) is more difficult.

2.4.1 Base brand approach

To estimate the parameters in attraction models, we consider the $(I-1)$-dimensional set of linear equations which results from log-linearizing the attraction model given in (2.7). In general, these equations can be written in the following form

$$\begin{array}{lcllll} y_{1t} & = & w_{1t}'b_1 & + z_{1t}'a & + & \eta_{1t} \\ y_{2t} & = & w_{2t}'b_2 & + z_{2t}'a & + & \eta_{2t} \\ \vdots & = & \vdots & + \quad \vdots & + & \vdots \\ y_{I-1,t} & = & w_{I-1,t}'b_{I-1} & + z_{I-1,t}'a & + & \eta_{I-1,t}, \end{array} \tag{2.18}$$

where $y_{it} = \ln M_{it} - \ln M_{It}$, $\eta_t = (\eta_{1t}, \ldots, \eta_{I-1,t})' \sim \mathrm{N}(\mathbf{0}, \tilde{\Sigma})$, and where w_{it} are k_i-dimensional vectors of explanatory variables with regression coefficient vector b_i, which is different in each equation, and where z_{it} are n-dimensional vectors of explanatory variables with regression coefficient vector a which is the same across the equations, $i = 1, \ldots, I-1$. Each (restricted) version of the general attraction model discussed in Section 2.2.2 can be written in this format.

To discuss parameter estimation, it is convenient to write (2.18) in matrix notation. We define $y_i = (y_{i1}, \ldots, y_{iT})'$, $W_i = (w_{i1}, \ldots, w_{iT})'$, $Z_i = (z_{i1}, \ldots, z_{iT})'$ and $\eta_i = (\eta_{i1}, \ldots, \eta_{iT})'$ for $i = 1, \ldots, I-1$. In matrix notation, (2.18) then becomes

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{I-1} \end{pmatrix} = \begin{pmatrix} W_1 & \mathbf{0} & \ldots & \mathbf{0} & Z_1 \\ \mathbf{0} & W_2 & \ldots & \mathbf{0} & Z_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & W_{I-1} & Z_{I-1} \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_{I-1} \\ a \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_{I-1} \end{pmatrix} \tag{2.19}$$

or

$$y = X\gamma + \eta \tag{2.20}$$

with $\eta \sim \mathrm{N}(\mathbf{0}, (\tilde{\Sigma} \otimes \mathbf{I}_T))$, where $\otimes$ denotes the familiar Kronecker product.

One method for parameter estimation of (2.20) is ordinary least squares [OLS]. Generally, however, this leads to consistent but inefficient estimates, where the inefficiency is due to the (possibly neglected) covariance structure of the disturbances. Only if the explanatory variables in each equation are the same, or in the unlikely case that $\tilde{\Sigma}$ is a diagonal matrix, and provided that there are no restrictions on the regression parameters ($w_{it} = \mathbf{0}$ for all i, t), OLS provides efficient estimates, see Judge *et al.* (1985, Chapter 12), among others. Therefore, one should better use generalized least squares [GLS] methods to estimate the model parameters. As the covariance matrix of the disturbances is usually unknown, one has to opt for a feasible GLS procedure, where we use the OLS estimator of the covariance matrix of the disturbances. This procedure is known as Zellner's (1962) seemingly unrelated regression [SUR] estimation method. Under the assumption of normality, an iterative SUR estimation method will lead to the maximum likelihood [ML] estimator of the model parameters, see Zellner (1962).

To estimate the parameters in attraction models, and to facilitate comparing various models, we favor ML estimation. The log of the likelihood function of (2.20) is given by

$$\ell(\gamma, \tilde{\Sigma}) = -\frac{T(I-1)}{2}\ln(2\pi) + \frac{T}{2}\ln|\tilde{\Sigma}^{-1}| - \frac{1}{2}(y - X\gamma)'(\tilde{\Sigma}^{-1} \otimes \mathbf{I}_T)(y - X\gamma). \tag{2.21}$$

The parameter values which maximize this log likelihood function are consistent and efficient estimates of the model parameters.

For the FE-MCI model without any parameter restrictions in (2.7), the ML estimator corresponds with the OLS estimator, as the explanatory variables are the same across equations. In that case,

$$\hat{\gamma}_{OLS} = (X'X)^{-1}X'y \tag{2.22}$$

such that $\hat{\gamma}_{OLS} = (\hat{b}_{OLS,1}, \ldots, \hat{b}_{OLS,I-1}, \hat{a}_{OLS})'$, see (2.19), and

$$\hat{\tilde{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T} \hat{\eta}_t \hat{\eta}_t', \tag{2.23}$$

where $\hat{\eta}_t$ consists of stacked $\hat{\eta}_{it} = y_{it} - w'_{it}\hat{b}_{OLS,i} - z'_{it}\hat{a}_{OLS}$.

For the attraction models with restrictions on the regression parameters, that is, for the RC model in (2.10), the RE model in (2.12), and the RD model in (2.13), one can opt for the iterative SUR estimator which converges to the ML estimator. Starting with the OLS-based estimator for $\tilde{\Sigma}$ in (2.23), one constructs the feasible GLS estimator

$$\hat{\gamma}_{SUR} = (X'(\hat{\tilde{\Sigma}}^{-1} \otimes \mathbf{I}_T)X)^{-1}X'(\hat{\tilde{\Sigma}}^{-1} \otimes \mathbf{I}_T)y, \tag{2.24}$$

that is the SUR estimator, see Zellner (1962). Next, we replace the estimate of the covariance matrix $\hat{\tilde{\Sigma}}$ by the new estimate of $\tilde{\Sigma}$, that is (2.23), where $\hat{\eta}_t$ now consists of stacked $\hat{\eta}_{it} = y_{it} - w'_{it}\hat{b}_{SUR,i} - z'_{it}\hat{a}_{SUR}$, to obtain a new SUR estimate of γ. This routine is repeated until the estimates for γ and $\tilde{\Sigma}$ have converged. Under the assumption of normally distributed disturbances, the final estimates are the ML estimates of the model, that is, they maximize the log likelihood function (2.21).

A little more involved are the restrictions on the $\tilde{\Sigma}$ matrix. To estimate the attraction model under the restriction (2.8), one can either directly maximize the log likelihood function (2.21) with $\tilde{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_{I-1}^2) + \sigma_I^2 \mathbf{i}_{I-1}\mathbf{i}'_{I-1}$ using a numerical optimization algorithm like Newton-Raphson or one can again use an iterative SUR procedure. In the latter approach, the new estimate of $\tilde{\Sigma}$ is obtained by maximizing

$$\ell(\tilde{\Sigma}) = -\frac{T(I-1)}{2}\ln(2\pi) + \frac{T}{2}\ln|\tilde{\Sigma}^{-1}| - \frac{1}{2}\hat{\eta}'(\tilde{\Sigma}^{-1} \otimes \mathbf{I}_T)\hat{\eta}, \tag{2.25}$$

where $\hat{\eta}$ are the residuals from the previous SUR regression. Again, we need a numerical optimization routine to maximize (2.25). Especially in cases where there are many brands, the optimization of (2.25) can become cumbersome. It can however be shown, see Appendix 2.A, that the optimization can be reduced to numerically maximizing a concentrated likelihood over just σ_I^2 where one uses

$$\hat{\sigma_i}^2 = \frac{\hat{\eta}'_i\hat{\eta}_i}{T} - \hat{\sigma}_I^2 \quad \text{for } i = 1, \ldots, I-1, \tag{2.26}$$

where $\hat{\eta}_i = (\hat{\eta}_{i1}, \ldots, \hat{\eta}_{iT})'$. Given an estimate of σ_I^2, this relationship can be used to obtain estimates of $\sigma_1^2, \ldots, \sigma_{I-1}^2$.

Finally, in all the above cases the standard errors for the estimated regression parameters γ are to be estimated by

$$\widehat{\text{Var}}(\hat{\gamma}) = (X'(\hat{\tilde{\Sigma}}^{-1} \otimes \mathbf{I}_T)X)^{-1}, \tag{2.27}$$

where one should include the appropriate ML estimator for $\tilde{\Sigma}$. When taking the square roots of the diagonal elements of this matrix, one obtains the appropriate standard errors.

2.4.2 Log-centering approach

The above estimation routine is based on the reduced-form model, which is obtained from reducing the system of equations using the base-brand approach. An alternative method is the, what is called, log-centering method advocated by Cooper and Nakanishi (1988). We will now show that this method is equivalent to the above method, although a bit more complicated.

The log-centering approach is based on the following transformation. After taking the natural logs for the I model equations, the log of the geometric mean market share over the brands is subtracted from all equations. The reduced-form model is now specified relative to the geometric mean. So instead of reducing the system of equations by using a base brand, this methodology reduces the system by the "geometric average brand". Note that the reduced-form model in this case still contains I equations.

To demonstrate the equivalence of parameters obtained through the log-centering technique of Cooper and Nakanishi (1988) and those using the base-brand approach, we show that there exists an exact relationship between these sets of parameters. The parameters for the base-brand specification can uniquely be determined from the parameters for the log-centering specification and vice versa. Given the 1-to-1 relationship the likelihoods are the same, that is, the discussed feasible GLS estimator yields the same maximum value of the likelihood as we can use the invariance principle of maximum likelihood, see for example Greene (1993, page 115). All that needs to be shown is the 1-to-1 relationship between the parameters in the two specifications.

Consider a general attraction specification, that is

$$A_{it} = \exp(\mu_i + \varepsilon_{it}) \prod_{j=1}^{I} \prod_{k=1}^{K} z_{kjt}^{\beta_{kji}}, \tag{2.28}$$

where z_{kjt} may contain any kind of explanatory variable, such as lagged market shares, promotion and price. The market shares are again defined by

$$M_{it} = \frac{A_{it}}{\sum_{j=1}^{I} A_{jt}}. \tag{2.29}$$

Written in a vector notation the model for the natural logarithm of attraction becomes

$$\begin{aligned} \ln A_t := \begin{pmatrix} \ln A_{1t} \\ \vdots \\ \ln A_{It} \end{pmatrix} &= \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_I \end{pmatrix} + \sum_{k=1}^{K} \begin{pmatrix} \beta_{k11} & \beta_{k21} & \dots & \beta_{kI1} \\ \beta_{k12} & \beta_{k22} & \dots & \beta_{kI2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1I} & \beta_{k2I} & \dots & \beta_{kII} \end{pmatrix} \begin{pmatrix} \ln z_{k1t} \\ \vdots \\ \ln z_{kIt} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{It} \end{pmatrix} \\ &= \mu + \sum_{k=1}^{K} B_k \ln z_{kt} + \varepsilon_t. \end{aligned} \tag{2.30}$$

The definition of market share in (2.29) implies that $\ln M_{it} = \ln A_{it} - \ln \sum_{j=1}^{I} A_{jt}$. In a vector notation this gives

$$\ln M_t := \begin{pmatrix} \ln M_{1t} \\ \vdots \\ \ln M_{It} \end{pmatrix} = \ln A_t - \mathbf{i}_I \ln \sum_{j=1}^{I} A_{jt}, \tag{2.31}$$

where $\mathbf{i}_I$ denotes a $(I \times 1)$ unity vector.

As the model in (2.31) cannot be estimated directly due to the nonlinear dependence of $\ln(\sum_{j=1}^{I} A_{jt})$ on the model parameters, a reduced-form model should be considered. The log-centering method now subtracts the average of the log market shares from the equations to give a reduced-form specification. The dependent variable in this system of equations is now

$$\begin{aligned} \begin{pmatrix} \ln M_{1t} \\ \vdots \\ \ln M_{It} \end{pmatrix} - \begin{pmatrix} \frac{1}{I} \sum_{j=1}^{I} \ln M_{jt} \\ \vdots \\ \frac{1}{I} \sum_{j=1}^{I} \ln M_{jt} \end{pmatrix} &= \begin{pmatrix} 1 - \frac{1}{I} & -\frac{1}{I} & \dots & -\frac{1}{I} \\ -\frac{1}{I} & 1 - \frac{1}{I} & \dots & -\frac{1}{I} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{I} & -\frac{1}{I} & \dots & 1 - \frac{1}{I} \end{pmatrix} \ln M_t \\ &= H^{(lc)} \ln M_t, \end{aligned} \tag{2.32}$$

where $H^{(lc)}$, with rank $I - 1$, denotes the transformation matrix corresponding to the log-centering approach. The reduced-form model then becomes

$$H^{(lc)} \ln M_t = H^{(lc)} \ln A_t - H^{(lc)} \mathbf{i}_I \ln \sum_{j=1}^{I} A_{jt}, \tag{2.33}$$

which equals

$$H^{(lc)} \ln M_t = H^{(lc)} \mu + \sum_{k=1}^{K} H^{(lc)} B_k \ln z_{kt} + H^{(lc)} \varepsilon_t \tag{2.34}$$

as $H^{(lc)} \mathbf{i}_I = \mathbf{0}_{I \times I}$. Due to the reduced rank of $H^{(lc)}$, the system in (2.34) contains I equations, but it only has $I - 1$ independent equations.

Alternatively, the base-brand approach in Section 2.4.1 gives as the dependent variables in the reduced-form model

$$\begin{aligned} \begin{pmatrix} \ln M_{1t} \\ \vdots \\ \ln M_{I-1,t} \end{pmatrix} - \begin{pmatrix} \ln M_{It} \\ \vdots \\ \ln M_{It} \end{pmatrix} &= \begin{pmatrix} 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix} \ln M_t \\ &= H^{(bb)} \ln M_t, \end{aligned} \tag{2.35}$$

with $H^{(bb)}$ as the relevant transformation matrix. As $H^{(bb)}\mathbf{i}_I = \mathbf{0}_{I-1\times I}$, the reduced-form model becomes

$$H^{(bb)} \ln M_t = H^{(bb)} \ln A_t = H^{(bb)}\mu + \sum_{k=1}^{K} H^{(bb)} B_k \ln z_{kt} + H^{(bb)}\varepsilon_t, \tag{2.36}$$

which is to be compared with (2.34). This system contains only $I-1$ equations.

The 1-to-1 relation between the parameters in the two approaches follows from the fact that the equation $CH^{(lc)} = H^{(bb)}$ yields a unique solution C, given by

$$C = \begin{pmatrix} 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -1 \\ 1 & 1 & \dots & 1 & 1 \end{pmatrix}. \tag{2.37}$$

Hence, the matrix C relates the "log-centered" parameters to the "base-brand" parameters. The inverse transformation from the base-brand specification to the log-centered specification follows from applying the Moore-Penrose inverse of C, denoted by C^+, that is,

$$C^+ = \begin{pmatrix} 1-\frac{1}{I} & -\frac{1}{I} & \dots & -\frac{1}{I} \\ -\frac{1}{I} & 1-\frac{1}{I} & \dots & -\frac{1}{I} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{I} & -\frac{1}{I} & \dots & 1-\frac{1}{I} \\ -\frac{1}{I} & -\frac{1}{I} & \dots & -\frac{1}{I} \end{pmatrix}. \tag{2.38}$$

Note that the matrix C^+ satisfies $H^{(lc)} = C^+ H^{(bb)}$.

The above shows that the transformations yield equivalent parameters. For example, assume that the log-centered form of the model is estimated, giving estimates of $H^{(lc)}\mu$, $H^{(lc)}B_k$ and $H^{(lc)}\Sigma H^{(lc)\prime}$. By multiplying the estimated system of equations by C we get $CH^{(lc)}\mu$, $CH^{(lc)}B_k$ and $CH^{(lc)}\Sigma H^{(lc)\prime}C'$ as model coefficients. Using the invariance principle of maximum likelihood and the relation $CH^{(lc)} = H^{(bb)}$, these coefficients are the maximum likelihood estimates of $H^{(bb)}\mu$, $H^{(bb)}B_k$ and $H^{(bb)}\Sigma H^{(bb)\prime}$. These coefficients are exactly the same as the coefficients used in the base-brand specification, see (2.36). Using the inverse of C, the procedure can be used the other way around. We can also obtain estimates of the coefficients in a log-centered specification from the estimates in a base-brand specification by multiplying them with C^+.

In our opinion, the main reason to prefer taking a base brand to reduce the model is that the statistical analysis of the resulting model is more straightforward as compared to the log-centering technique. Recall that the log-centered reduced-form model contains I equations whereas the base brand reduced-form model only has $I-1$ equations. One of the

equations in the log-centered specification is however redundant. This redundancy leads to some difficulties in the estimation and interpretation, as estimation usually requires the (inverse) covariance matrix of the residuals. In the log-centering case the residuals are linearly dependent, and the covariance matrix is therefore non-invertible. Further, direct interpretation of the coefficients obtained from the base-brand approach is easier as each coefficient only concerns two brands, while a coefficient in the log-centering approach always involves all brands.

Another advantage of using the base-brand approach concerns markets where the number of brands changes over time. In this case the "geometric average" brand may consist of different number of brands across weeks. The variability in the market share of this average brand will fluctuate with the numbers of brands available. Using this average brand as a base brand, as proposed in the log-centering approach, will therefore introduce complicated forms of heteroscedasticity. If a brand is available during the entire sample period, the base-brand approach can be straightforwardly applied without introducing heteroscedasticity. If such a brand is not available, a different base brand can be considered for different weeks. This will also introduce some heteroscedasticity, but of a more manageable form than would be the case for the log-centering approach.

2.5 Model selection

Attraction models are often considered for forecasting market shares. It is usually assumed that, by imposing in-sample specification restrictions, the out-of-sample forecasting accuracy will improve. Exemplary studies are Brodie and Bonfrer (1994), Danaher (1994), Naert and Weverbergh (1981), Leeflang and Reuyl (1984), Kumar (1994) and Chen *et al.* (1994), among others. A summary of the relevant studies is given in Brodie *et al.* (2001). A common characteristic of these studies, an exception being Chen *et al.* (1994), is that they tend to compare one or two specific forms of the attraction model with various more naive models. In this section we consider the question of obtaining the best (or a good) choice for the specification from the wide range of possible attraction specifications.

There are of course many possible approaches to obtain a suitable attraction specification. One could consider a set of popular specifications and select the optimal model using an information criterion, like the BIC (Schwarz, 1978), or use statistical tests to determine the "best" model. In a Bayesian setting one could even derive posterior probabilities for the proposed models. One may select the model with the highest posterior probability or one can combine several models. For example, to construct forecasts, one can use the posterior probabilities to weight forecasts generated by the different models. Another strategy is to start with a general model and try to simplify it using statistical

test. In this chapter we opt for this general-to-simple model selection strategy, following Hendry (1995).

The starting point of the model selection strategy is the most extended attraction model, that is, model (2.7) without any restrictions. Of course, in practice the size of the model is governed by data availability and sample size. The first step of a model selection strategy concerns fixing the proper lag order P of the model. It is well known that an inappropriate value of P leads to inconsistent and inefficient estimates. To perform valid inference on the restrictions on the explanatory variables and covariance matrix it is therefore necessary to first determine the appropriate lag order. Furthermore, imposing incorrect restrictions on the explanatory variables and covariance matrix may lead to selecting an incorrect lag order. Lag order selection may be based on the BIC criterion. Another strategy may be a sequential procedure, where one starts with a large value of P and tests for the significance of the $\tilde{\beta}_{Pkji}$ and $\tilde{\alpha}_{Pji}$ parameters and imposes these restrictions when they turn out to be valid. These tests usually concern many parameter restrictions and may therefore have little power. Instead, one may therefore base the lag order determination on Lagrange Multiplier [LM] tests for serial correlation in the residuals, see Lütkepohl (1993) and Johansen (1995, p. 22). The advantage of these tests is that they concern less parameter restrictions and hence have more power. We would recommend to start with a model of order 1 and increase the order with 1 until the LM tests do not indicate the presence of any serial correlation.

Once P is fixed, we propose to test the validity of the various restrictions on (2.7) as proposed in Section 2.2.2. We test for the validity of restriction (2.8) on the covariance matrix $\tilde{\Sigma}$ [RCM] in model (2.7). Additionally, we test in model (2.7) for restricted dynamics [RD], common dynamics [CD], and, for each explanatory variable k, for restricted competition [RC], for restricted effects [RE] (2.12) and even for the absence of this variable. Finally, we propose to test for the significance of the lagged explanatory variables in the general model.

Next, we recommend to perform an overall test for all restrictions which were not rejected in the individual tests. If this joint test is not rejected, all restrictions are imposed, and this results in a final model that can be used for forecasting. However, if the joint test indicates rejection, one may want to decide to relax some restrictions, where the p-values of the individual tests can be used to decide which of these restrictions have to be relaxed. Note that apart from the lag order selection stage we perform the individual tests in the general model and that we do not directly impose the restrictions if not rejected. Hence, the model selection approach in this stage does not depend on the sequence of the tests. Furthermore, as we use a general-to-specific strategy, we do not *a priori* exclude model specifications.

To apply our general-to-simple model selection strategy, we have to test for restrictions on the covariance matrix $\tilde{\Sigma}$ and on the other model parameters (collected in γ) in (2.7). To

test these parameter restrictions, we opt for Likelihood Ratio [LR] tests, see for example Judge *et al.* (1985, p. 475). Denoting the ML estimates of the parameters under the null hypothesis by $(\hat{\gamma}_0, \hat{\tilde{\Sigma}}_0)$ and the ML estimates under the alternative hypothesis by $(\hat{\gamma}_a, \hat{\tilde{\Sigma}}_a)$, then

$$\text{LR} = -2(\ell(\hat{\gamma}_0, \hat{\tilde{\Sigma}}_0) - \ell(\hat{\gamma}_a, \hat{\tilde{\Sigma}}_a)) \underset{asy}{\sim} \chi^2(\nu), \tag{2.39}$$

where $\ell(\cdot)$ denotes the log-likelihood function as defined in Section 2.4 and where ν is the number of parameter restrictions.

2.6 Forecasting

There has been considerable research on forecasting market shares using the market share attraction model. Most studies discuss the effect on the forecasts of the estimation technique used in combination with the parametric model specification, see for example Leeflang and Reuyl (1984), Brodie and de Kluyver (1984) and Ghosh *et al.* (1984), among others. More recent interest has been on the optimal model specification under different conditions, see, for example, Kumar (1994) and Brodie and Bonfrer (1994). The available literature, however, is not very informative as to how forecasts of market shares should be generated. In this section we show that forecasting market shares turns out not to be a trivial exercise and that in order to obtain unbiased forecasts one has to use simulation methods.

Furthermore, in empirical applications it should be recognized that parameter values are obtained through estimation. The true parameter values are usually unknown, and parameter values are at best obtained through unbiased estimators of the true values. In a linear model this parameter uncertainty can be ignored when constructing unbiased forecasts. However, in nonlinear models this may not be true. Recently, Hsu and Wilcox (2000) addressed this issue in the context of a multinomial logit framework. These authors study the stochastic prediction of market shares using the multinomial logit model. In their paper they stress that in order to obtain accurate forecasts the uncertainty in parameter estimates should be taken into account. Upon using Monte Carlo experiments, they demonstrate that indeed the forecast accuracy improves when uncertainty is included.

The market share attraction model differs from the multinomial logit model in an important aspect. The multinomial logit model obtains the market share as aggregated brand choice probabilities. For a large number of households the market shares are therefore deterministic. In a market share attraction model we however have two sources of uncertainty. We have the intrinsic uncertainty due to the stochastic nature of the market shares and we have the uncertainty induced by parameter uncertainty. Even when the model parameters are known it is not straightforward to obtain market share forecasts.

In Section 2.6.1 we present how to obtain forecasts without considering parameter uncertainty. In Section 2.6.2 we discuss the case where parameter uncertainty is taken into account.

2.6.1 Forecasting market shares

To provide some intuition why forecasting in a market share attraction model is not a trivial exercise, consider the following. The attraction model ensures logical consistency, that is, market shares lie between 0 and 1 and they sum to 1. These restrictions imply that the model parameters can be estimated from a multivariate reduced-form model with $I-1$ equations. The dependent variable in each of the $I-1$ equations is the natural logarithm of a relative market share. More formally, it is $\ln m_{it} \equiv \ln \frac{M_{it}}{M_{It}}$, for $i = 1, 2, \ldots, I-1$. The base brand I can be chosen arbitrarily.

Of course, one is usually interested in predicting M_{it} and not in the logs of the relative market shares. It is then important to recognize that, first of all, $\exp(\mathrm{E}[\ln m_{it}])$ is not equal to $\mathrm{E}[m_{it}]$ and that, secondly, $\mathrm{E}[M_{it}/M_{It}]$ is not equal to $\mathrm{E}[M_{it}]/\mathrm{E}[M_{It}]$, where E denotes the expectation operator. Therefore, unbiased market share forecasts cannot be obtained by routinized transformations of forecasts of log relative market shares, see also Fok and Franses (2001) for similar statements in the context of forecasting market shares from models for sales.

To forecast the market share of brand i at time t, one needs to consider the relative market shares

$$m_{jt} = M_{jt}/M_{It} \quad \text{for } j = 1, 2 \ldots, I, \tag{2.40}$$

as $m_{1t}, \ldots, m_{I-1,t}$ form the dependent variables (after log transformation) in the reduced-form model (2.7). As $M_{It} = 1 - \sum_{j=1}^{I-1} M_{jt}$, we have that

$$\begin{aligned} M_{It} &= \frac{1}{1 + \sum_{j=1}^{I-1} m_{jt}} \\ M_{it} &= M_{It} m_{it} = \frac{m_{it}}{1 + \sum_{j=1}^{I-1} m_{jt}} \quad \text{for } i = 1, 2, \ldots, I-1. \end{aligned} \tag{2.41}$$

Note that $m_{It} = M_{It}/M_{It} = 1$ and hence (2.41) can be summarized as

$$M_{it} = \frac{m_{it}}{\sum_{j=1}^{I} m_{jt}} \quad \text{for } i = 1, 2, \ldots, I. \tag{2.42}$$

As the relative market shares m_{it}, $i = 1, \ldots, I-1$ are log-normally distributed by assumption, see (2.7), the probability distribution of the market shares involves the inverse of the sum of log-normally distributed variables. The exact distribution function of the market shares is therefore complicated. Moreover, correct forecasts should be based on

the expected value of the market shares, and unfortunately, for this expectation there is no simple algebraic expression. Appropriate forecasts therefore cannot be obtained from the expectations directly.

If we ignore parameter uncertainty for the moment, we need to calculate the expectations of the market shares given in (2.42). This cannot be done analytically. However, we can calculate the expectations using simulations. The relevant procedure works as follows. We use model (2.7) to simulate relative market shares for various disturbances η randomly drawn from a multivariate normal distribution with mean 0 and covariance matrix $\tilde{\Sigma}$. In each run, we compute the market shares where parameter values and the realization of the disturbance process are assumed to be given. The market shares averaged over a number of replications now provide their unbiased forecasts. Notice that we only need the parameters of the reduced-form model in the simulations.

To be more precise about this simulation method, consider the following. The one-step ahead forecasts of the market shares are simulated as follows, first draw $\eta_t^{(l)}$ from $\mathrm{N}(0, \tilde{\Sigma})$, then compute

$$m_{it}^{(l)} = \exp(\tilde{\mu}_i + \eta_{it}^{(l)}) \prod_{j=1}^{I} \left(\prod_{k=1}^{K} x_{kjt}^{\tilde{\beta}_{kji}} \prod_{p=1}^{P} \left(M_{j,t-p}^{\tilde{\alpha}_{pji}} \prod_{k=1}^{K} x_{kj,t-p}^{\tilde{\beta}_{pkji}} \right) \right), \; i = 1, \ldots, I-1, \quad (2.43)$$

with $m_{It}^{(l)} = 1$ and finally compute

$$M_{it}^{(l)} = \frac{m_{it}^{(l)}}{\sum_{j=1}^{I} m_{jt}^{(l)}} \quad \text{for } i = 1, \ldots, I, \quad (2.44)$$

where $l = 1, \ldots, L$ denotes the simulation iteration and where the FE-MCI specification is used, see (2.4). Every vector $(M_{1t}^{(l)}, \ldots, M_{It}^{(l)})'$ generated this way amounts to a draw from the joint distribution of the market shares at time t. Using the average over a sufficiently large number of draws we calculate the expected value of the market shares. By the weak law of large numbers we have

$$\operatorname*{plim}_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} M_{it}^{(l)} = \mathrm{E}[M_{it}]. \quad (2.45)$$

For finite L the mean value of the generated market shares is an unbiased estimator of the market share. The estimate may differ from the expected market share, but this difference is only due to simulation error and this error will rapidly converge to zero if L gets large. Of course, the value of L can be set at a very large value, depending on available computing power.

The lagged market shares in (2.7) are of course only available for one-step ahead forecasting and not for multiple-step ahead forecasting. Hence, one has to account for

the uncertainty in the lagged market share forecasts. One can now simply use simulated values for lagged market shares, thereby automatically taking into account the uncertainty in these lagged variables. Note that we assume that the marketing efforts of all market players are known. It is possible to also model these efforts and use the estimated model to obtain market share forecasts that also account for that uncertainty. The models describing the marketing efforts can be used to simulate future values of the levels of the marketing instruments. To take into account the uncertainty of future marketing efforts for forecasting market shares, we use the simulated efforts instead of forecasted efforts to obtain draws from the joint distribution of market shares in (2.43) and (2.44).

2.6.2 Parameter uncertainty

As the model parameters are estimated and parameter estimators are random variables, one should take into account their associated uncertainty, see also Hsu and Wilcox (2000). When estimated parameters are used for forecasting in combination with a nonlinear model, we should also take into account the uncertainty of these estimates. In linear models, the uncertainty can be ignored. To see this, consider the model $y = X\beta + \varepsilon$, $\varepsilon \sim N(0, 1)$. The OLS estimate of β, denoted by $\hat{\beta}$, is a stochastic variable as it is a function of the data y. The uncertainty in this estimate however is irrelevant for generating unbiased forecasts as $\mathrm{E}[X\hat{\beta}] = X\mathrm{E}[\hat{\beta}] = X\beta = \mathrm{E}[y]$. In nonlinear models this is in general not the case. Consider $y = g(X, \beta) + \varepsilon$, where $g(\cdot, \cdot)$ is a nonlinear function. In general, $\mathrm{E}[g(X, \hat{\beta})] \neq g(X, \mathrm{E}[\hat{\beta}]) = g(X, \beta) = \mathrm{E}[y]$. Therefore $g(X, \hat{\beta})$ is not an unbiased estimator of y, even when $\hat{\beta}$ is an unbiased estimator of β. For some special cases there are closed-form expressions for obtaining unbiased forecasts in a nonlinear models using estimated parameters. For example, Finney (1941) and Bradu and Mundlak (1970) consider the expectation of log-normal random variables and forecasting in log-normal regression, respectively. Even for the rather simple case of log-normal regression, the expressions derived for the forecasts are very involved. It is very likely that in our case, where we have a multivariate model with a nonlinear dependence between the disturbances and the dependent variable, a technique based on the work of Finney (1941) is not feasible.

To take account of the stochastic nature of the estimator, we again have to rely on simulation. Unfortunately, the relevant distribution of the parameters is not known. To overcome this difficulty, we propose to use parametric bootstrapping to draw parameters from their distribution. Summarizing all parameters in θ, we sample θ using the following scheme: (i) Use the estimated parameters $\hat{\theta}$, the realizations of the exogenous variables and the first P observed realizations as starting values to generate artificial realizations of the market shares; (ii) Reestimate the model based on this artificial data. The thus obtained parameters $\hat{\theta}^{(l)}$, $l = 1, \ldots, L$, where L denotes the number of draws, can be seen as draws from the small sample distribution of $\hat{\theta}$. For every bootstrap realization of $\hat{\theta}^{(l)}$

we calculate the conditional expectation $\mathrm{E}[M_{it}|\ \hat{\theta}^{(l)}],\ i = 1, \ldots, I$ using the simulation technique in Section 2.6.1. The average of the forecasts over all generated parameter vectors constitutes unbiased forecasts of the market shares under uncertain parameters. That is, we calculate $\mathrm{E}[M_{it}]$ as $\frac{1}{L}\sum_{l=1}^{L}\mathrm{E}[M_{it}|\ \hat{\theta}^{(l)}]$. It is not necessary to use many simulation rounds conditional on the parameters. Theoretically it suffices to use one round for every $\hat{\theta}^{(l)}$.

In a classical setting we have to rely on bootstrapping techniques to account for parameter uncertainty. A Bayesian analysis of market share models would have the advantage that it provides a more natural approach to account for parameter uncertainty. To obtain the posterior distribution of the parameters of the market share attraction model one can rely on Markov chain Monte Carlo [MCMC] methods, see Casella and George (1992) for a simple introduction and Paap (2002) for a recent survey. As byproduct of this sampler we can obtain forecasts which account for parameter uncertainty. See Chapter 4 for a discussion of the market share attraction model in a Bayesian setting. Although we do not consider forecasting in that chapter, the sampling scheme derived there can easily be used to generate forecasts.

2.7 Marketing implications

We now develop some empirical intuition on the adverse effects of not using unbiased forecasts and of not using a correctly specified model. To illustrate the effects of the forecasting method, consider

$$\begin{aligned} A_{it} = \exp(\mu_i + \varepsilon_{it}) \text{ and } M_{it} &= \frac{A_{it}}{A_{1t} + A_{2t} + A_{3t}}\ , i = 1, 2, 3, \\ (\mu_1, \mu_2, \mu_3)' &= (2, -1, 1), \\ (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})' &\sim N(0, \begin{pmatrix} 2 & 0.5 & -0.5 \\ 0.5 & 1 & -0.5 \\ -0.5 & -0.5 & 1 \end{pmatrix}). \end{aligned} \tag{2.46}$$

A tempting (but naive and incorrect) way to get market share forecasts relies on forecasts of log market shares relative to a base brand, that is, $\ln m_{it} = \ln(A_{it}/A_{3t}) = \mu_i - \mu_3 + \varepsilon_{it} - \varepsilon_{3t}$, where here brand 3 is chosen as the base brand. Therefore, $\widehat{\ln m_{it}} = \mu_i - \mu_I$. From these log relative market share forecasts, market share forecasts could be obtained from $\widehat{M}_{it} = \exp(\widehat{\ln m_{it}})/\sum_{j=1}^{3}\exp(\widehat{\ln m_{jt}})$. With this, the market share forecasts would become $\widehat{M}_{1t} = 0.705$, $\widehat{M}_{2t} = 0.035$ and $\widehat{M}_{3t} = 0.260$.

These forecasts are of course biased as the order of the expectation, the exponent and the division operator can not be interchanged! Indeed, unbiased forecasts follow from

$$\mathrm{E}[M_{it}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{A_{it}}{A_{1t} + A_{2t} + A_{3t}}\phi(\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})d\varepsilon_{1t}d\varepsilon_{2t}d\varepsilon_{3t}, \tag{2.47}$$

where $\phi(\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})$ denotes the joint (normal) density function of the random attractions. Evaluating the integrals using simulation yields $\mathrm{E}[M_{1t}] = 0.616$, $\mathrm{E}[M_{2t}] = 0.048$ and $\mathrm{E}[M_{3t}] = 0.336$, and these correct forecasts clearly differ from those reported above.

To illustrate the empirical relevance of the correct forecasting procedure, consider a marketing manager who has to decide whether or not to feature or display her brand in the next week. This decision could be based on a comparison of the costs of the promotion and the expected additional profit. Profit can be defined as $\Pi = (p-c)M \times S$, with p the price, c the unit cost of the product, M the market share, and S the market size. Assuming no category expansion due to the promotion, the expected additional profit equals $\Delta\Pi = (p-c)(\mathrm{E}[M|\text{ prom.}] - \mathrm{E}[M|\text{ no prom.}]) \times S$, with $\mathrm{E}[M|$ prom.$]$ and $\mathrm{E}[M|$ no prom.$]$ the expected market shares with and without a promotion, respectively. To compare the additional profits under the two forecasting methods, we will consider the percentage difference of the additional profits, that is, $(\Delta\Pi^n - \Delta\Pi^s)/\Delta\Pi^s$, with n denoting naive forecasts and s denoting simulated forecasts. Note that this measure does not depend on the price, unit cost or the market size.

To illustrate the effects for a realistic situation, we use the model selected for the first case considered in Section 2.8. This model includes as explanatory variables prices, display, feature, these variables one period lagged, and lagged market shares. For the period for which the promotion decision is made, we fix the prices, lagged prices and lagged market shares to their average value. We also assume the absence of displays and features of competitors, to save space. Of course, the exercise could be extended in various directions. Table 2.2 shows that using the naive forecasts yields quite substantive errors in the expected additional profits. For the own effects, indicated in bold face in Table 2.2, the naive method overestimates the effects for three brands and underestimates the effect for the fourth (base) brand. The cross effects in the off-diagonal elements concern a change in profit of brand A due to a promotion of brand B. These effects are measured highly inaccurately when estimated using the naive method, with differences in the expected additional profit up to 20%.

Finally, to emphasize the importance of proper specification for attraction models, we consider a simple model with only one explanatory variable, say, the logarithm of the price for each brand. We consider three versions of this model, that is, (i) no restrictions on the competitive structure, (ii) restricted competition and (iii) restricted effects. We generate data based on price series and estimated parameters from the ERIM data base on Lite Tuna, see Section 2.8, using each of these versions. For each data generating process, the three models are estimated. Table 2.3 shows the forecasting accuracy for each of these models. On the brand level we measure the forecasting performance using the Root Mean Squared Prediction Error [RMSPE]. On the market level this measure is however less suitable as market shares, and market share forecasts, sum to one, the forecasts error will sum to zero. As a market level measure we use the log of the determinant of the residual

Table 2.2: Percentage bias in the effect of a feature or display. A positive percentage indicates an overstatement of the effect estimated by the naive method.

	Display				Feature			
	Del Monte	Heinz	Hunts	Rest	Del Monte	Heinz	Hunts	Rest
Del Monte	**1.93**	4.86	−10.49	−21.18	**2.27**	2.98	−9.80	−20.84
Heinz	11.43	**5.50**	16.68	3.45	12.14	**3.11**	16.42	5.44
Hunts	−11.22	9.48	**7.30**	−18.09	−11.44	6.88	**7.36**	−16.20
Rest	−23.51	−2.66	−18.81	**−5.70**	−23.54	−5.36	−18.33	**−4.04**

covariance matrix, where one brand is omitted. Due to the properties of the forecasting errors and the determinant operator this measure is invariant for the choice of the omitted brand. A large determinant corresponds to large forecasting errors.

We find that imposing restrictions while they are not valid is worse than the other way around. That is, estimating a model with full competition when in fact the true model has restricted competition yields a better performing model than when a model with restricted effects were estimated. Moreover, standard econometric theory tells us that inappropriate parameter restrictions imply omitted variable bias. Consequently, parameter estimates cannot be trusted. In general marketing-mix elasticities will be biased in case a too restrictive specification is used, whereas we will obtain unbiased estimates if too few restrictions are imposed. However, proper restrictions increase efficiency, and subsequent statistical analysis of parameters can be based on greater accuracy.

2.8 Illustration

For our empirical work we consider the so-called ERIM database of the University of Chicago. Each of the available datasets contains information on market shares and marketing instruments of a specific product category in either Sioux Falls (market 1) or Springfield (market 2), USA, all collected by A.C. Nielsen. The data span 124 weeks from July 1986 to December 1988. We have information on 14 data sets containing market shares, prices, and two promotional 0/1 dummy variables (display and feature). The data sets concern seven product categories, containing two, three or four brands in two markets, see the first three rows of Table 2.4.

For each of these markets we consider our in-sample model selection strategy. Model selection and estimation is based on the first 111 observations, so as to have 13 observations (a quarter of a year) for out-of-sample forecasting. Table 2.4 shows the results of model

Table 2.3: Forecast rank based on the Root Mean Squared Prediction Error for each brand (first 3 sets of columns) and on the determinant of the covariance matrix of the forecast errors (last 3 columns)

	Data Generating Process											
	Brand 1			Brand 2			Brand 3			All brands		
	NR	RC	RE	NR	RC	RE	NR	RC	RE	NR	RC	RE
Model						*Tuna market 1*						
NR	2	3	3	2	3	3	1	2	3	1	2	3
RC	3	2	2	3	2	2	2	1	2	2	1	2
RE	1	1	1	1	1	1	3	3	1	3	3	1
Model						*Tuna market 2*						
NR	1	2	3	1	2	3	1	2	3	1	2	3
RC	2	1	2	2	1	2	2	1	2	2	1	2
RE	3	3	1	3	3	1	3	3	1	3	3	1

RC=restricted competition, RE=restricted elasticities, NR=no restrictions.

selection. The fourth row of Table 2.4 shows the resulting lag order P of the models. The optimal value of P turns out to be 1 for all data sets, except for Sugar market 2, where P appears to be two. The final six rows show the restrictions (if any) on the dynamic structure, the covariance matrix, and the exogenous variables resulting from the model selection strategy. We notice that in most cases (11 out of 14), models with common dynamics [CD] are preferred. Also, the restricted covariance matrix restriction is preferred in 10 out of 14 cases. This is quite interesting as this restriction is rarely (if ever) imposed in practice! Lagged exogenous variables are relevant in 11 out of 14 cases, which is also in contrast with many models considered in the literature, see Table 2.1. The main conclusion is that no single type of model is preferred for all 14 cases. Hence, our model selection strategy arrives at a wide variety of models.

The last 13 observations are used to assess the out-of-sample forecasting accuracy of the different models. The forecasts of the models, discovered by our model selection strategy, are to be compared with the unbiased forecasts generated from the five alternative attraction models in Table 2.1. In practice, marketing actions of competitors are not known to a brand manager. These marketing efforts therefore also have to be forecasted. Klapper and Herwartz (2000) suggest that, for market share forecasting, simple mod-

Table 2.4: Attraction models selected by the general-to-simple strategy

ategory	Catsup		Peanut butter		Stick margarine		Sugar		Tissue		Tuna		Tube margarine	
arket	1	2	1	2	1	2	1	2	1	2	1	2	1	2
umber of brands	4	4	4	4	4	4	2	2	4	4	3	3	4	4
ag order (P)	1	1	1	1	1	1	2	1	1	1	1	1	1	1
						Valid restrictions[a]								
ynamics	CD	CD	RD	NR	CD	CD	CD	CD	CD	RD	CD	CD	CD	CD
ovariance matrix	RCM	RCM	RCM	RCM	RCM	RCM	NR	NR	RCM	RCM	NR	NR	RCM	RCM
rice	RE	RC	NR	NR	NR	NR	RE	RC	RC	NR	RE	NR	RC	NR
isplay	RE	RE	NR	NR	RE	NI	RE	RC	RE	NR	NR	NI	NR	NI
eature	RC	NR	RC	RE	RE	RC	RC	NI	RC	RC	RE	RE	NR	RC
agged Exogenous	IN	IN	IN	IN	IN	IN	IN	IN	NI	IN	NI	IN	IN	NI

RD=restricted dynamics, CD=common dynamics, RC=restricted competition, RE=restricted elasticities,
RCM=restricted covariance matrix, NR=no restrictions, and NI=not included IN=included.

els for marketing instruments suffice. We follow their recommendations and use AR(2) models for the prices and rely on logit models for feature and display. The logit models have as explanatory variables the lagged feature and display indicators of all brands. To accommodate for the uncertainty in these actions we base our market share forecasts on simulated marketing efforts.

Table 2.5: Rank based on 1- and 2-step ahead forecasts using the simulation method (1 is best, 6 is worst)

	Model*					
	I	II	III	IV	V	VI
1-step ahead forecasting						
Number of times best model	1	0	2	2	2	7
Average rank	3.07	3.86	3.93	3.71	3.79	2.64
2-step ahead forecasting						
Number of times best model	0	1	2	3	1	7
Average rank	3.57	3.79	3.21	3.93	3.36	3.14

* Models I-V are given in Table 2.1. Model VI is selected by our model selection strategy and details are given in Table 2.4.

Table 2.5 evaluates the out-of-sample performance of the six considered models. The appropriate evaluation criterion is again the log of the determinant of the covariance matrix of the forecast errors. The models selected by our strategy (model VI) deliver the best forecasts in 7 out of 14 cases for both horizons, where for each forecast 25,000 replications are used. Furthermore, the average rank across models is lowest. Additionally, for each model we calculate the ratio of the forecast error and the forecast error of the best performing model. The average of this ratio over all markets gives an indication of the improvement in forecasting performance. For models I to VI, these ratios equal 1.046, 1.055, 1.073, 1.050, 1.066, and 1.028, thereby indicating that on average the model selected by our strategy performs 2.8% worse than the best model, whereas the other models perform about 5% (or more) worse.

2.9 Concluding remarks

In this chapter we have gone through part of the econometrics involved in analyzing market share attraction models. We believe that a systematic strategy enhances the possibility to compare various empirical findings and to understand deficiencies in case model forecasts turn out to be inaccurate. Key issues that we have discussed are model selection and forecasting. We have shown that the commonly applied forecasting method yields biased forecasts and that the use of a systematic model selection procedure can further enhance the performance of an empirical model.

There are a few more issues that we feel need concern. First of all, one may want to allow for the event of new brands entering the market or old brands leaving it. In Chapter 3 we discuss techniques for doing so. Another area of research involves the analysis of possibly differing short-run and long-run effects of marketing efforts, see Dekimpe and Hanssens (1995b) and Paap and Franses (2000), among others. In Chapter 4 we discuss this issue in the context of the market share attraction model.

2.A Estimation of restricted covariance matrix

Recall the log likelihood function (2.25)

$$\ell(\tilde{\Sigma}) = -\frac{T(I-1)}{2}\ln(2\pi) + \frac{T}{2}\ln|\tilde{\Sigma}^{-1}| - \frac{1}{2}\hat{\eta}'(\tilde{\Sigma}^{-1}\otimes \mathbf{I}_T)\hat{\eta}, \tag{2.48}$$

where $\tilde{\Sigma} = \text{diag}(\sigma_1^2,\ldots,\sigma_{I-1}^2) + \sigma_I^2 \mathbf{i}_{I-1}\mathbf{i}_{I-1}'$. For $i = 1,\ldots,I-1$ it holds that

$$\begin{aligned}
\frac{\partial \ell(\tilde{\Sigma})}{\partial \sigma_i} &= \left(\frac{\partial \ell(\tilde{\Sigma})}{\partial \text{vec}(\tilde{\Sigma}^{-1})}\right)' \frac{\partial \text{vec}(\tilde{\Sigma}^{-1})}{\partial \sigma_i} \\
\frac{\partial \ell(\tilde{\Sigma})}{\partial \text{vec}(\tilde{\Sigma}^{-1})} &= \frac{T}{2}\frac{\partial \ln|\tilde{\Sigma}^{-1}|}{\partial \text{vec}(\tilde{\Sigma}^{-1})} - \frac{1}{2}\frac{\partial \hat{\eta}'(\tilde{\Sigma}^{-1}\otimes \mathbf{I}_T)\hat{\eta}'}{\partial \text{vec}(\tilde{\Sigma}^{-1})} \\
&= \frac{T}{2}\text{vec}(\tilde{\Sigma}) - \frac{1}{2}\text{vec}\left(\begin{pmatrix}\hat{\eta}_1' \\ \vdots \\ \hat{\eta}_{I-1}'\end{pmatrix}(\hat{\eta}_1,\ldots,\hat{\eta}_{I-1})\right) \\
&= \frac{1}{2}\text{vec}[T\tilde{\Sigma} - \begin{pmatrix}\hat{\eta}_1' \\ \vdots \\ \hat{\eta}_{I-1}'\end{pmatrix}(\hat{\eta}_1,\ldots,\hat{\eta}_{I-1})]
\end{aligned} \tag{2.49}$$

and

$$\begin{aligned}
\frac{\partial \text{vec}(\tilde{\Sigma}^{-1})}{\partial \sigma_i} &= \text{vec}(-\tilde{\Sigma}^{-1}\frac{\partial \tilde{\Sigma}}{\partial \sigma_i}\tilde{\Sigma}^{-1}) = -(\tilde{\Sigma}^{-1})_{ii}^2 e_{i,I-1} \\
\frac{\partial \ell(\tilde{\Sigma})}{\partial \sigma_i} &= \frac{1}{2}\text{tr}[-T\tilde{\Sigma}(\tilde{\Sigma}^{-1})_{ii}^2 e_{i,I-1} + \begin{pmatrix}\hat{\eta}_1' \\ \vdots \\ \hat{\eta}_{I-1}'\end{pmatrix}(\hat{\eta}_1,\ldots,\hat{\eta}_{I-1})(\tilde{\Sigma}^{-1})_{ii}^2 e_{i,I-1}] \\
&= \frac{1}{2}[-T(\tilde{\Sigma})_{ii}(\tilde{\Sigma}^{-1})_{ii}^2 + \hat{\eta}_i'\hat{\eta}_i(\tilde{\Sigma}^{-1})_{ii}^2] \\
&= \frac{1}{2}(\tilde{\Sigma}^{-1})_{ii}^2[\hat{\eta}_i'\hat{\eta}_i - T(\sigma_i^2+\sigma_I^2)],
\end{aligned} \tag{2.50}$$

where e_{ik} is a zero vector of size $(k\times 1)$ with the i-th element equal to 1. Solving the last equation given $\hat{\sigma}_I^2$ yields

$$\hat{\sigma_i}^2 = \frac{\hat{\eta}_i'\hat{\eta}_i}{T} - \hat{\sigma}_I^2. \tag{2.51}$$

The concentrated likelihood is obtained by inserting (2.51) into the likelihood (2.48). The concentrated likelihood now has to be optimized over just one parameter, that is σ_I.

Chapter 3

Analyzing the effects of a brand introduction on competitive structure using a market share attraction model

3.1 Introduction

The introduction of a new brand in an existing market can have a large impact on the competitive structure, which concerns the market shares, the marketing instrument elasticities (as established by consumer behavior) and the use of marketing-mix instruments by brand managers. For example, such an introduction may trigger intensified price competition, thereby possibly also affecting the relative effectiveness of price promotions for the incumbent brands. There may also be consumer reactions which can lead to a change in the relative effectiveness of pricing strategies. Note that similar changes could occur if a brand is removed from the market. Naturally, the effects of a changing number of brands is not confined to price competition, as the same arguments would hold for any other marketing instrument, like for example display or distribution. Furthermore, after correcting for the effects of the marketing-mix variables, the new brand may turn out to win share at the cost of only a few competitors, instead of drawing share proportionally from all incumbent brands.

In the literature, there are several studies of the effects of the entry of a new brand on the competitive structure. These studies can be broadly divided in two types of approaches. One approach takes a non-cooperative game-theoretic view, while the other is predominantly based on empirical research using time-series or panel data models.

The first type of studies takes a normative viewpoint, that is, there is a focus on how one should respond to an entry in an optimal way. An example is Basuroy and Nguyen (1998), who derive the theoretical conditions under which a market share attraction model is appropriate for equilibrium analysis. Within the context of this model, these authors demonstrate that the entry of a new brand would establish price decreases, which would hold true for fixed and expanding markets. For fixed markets they further show that incumbent brands would be inclined to lower marketing expenditures, while in expanding markets these expenditures would be set at higher levels. Other examples of similar approaches can be found in Karnani (1985), Cubbin and Domberger (1988), and Gruca *et al.* (1992, 2001). A key feature of these studies is that there is usually no focus on empirical data. A notable exception is Shankar (1997) who studies the marketing-mix reactions by pioneers to entry. In the studied market, the entry changes the competition from a monopoly to a duopoly, with as a consequence that in the duopoly market, different competitive games can be played. Next, optimal responses to entry are derived for each case and empirical data is then used to find the actual game played in a pharmaceutical market. Shankar (1997) concludes that the results found in for example Gruca *et al.* (1992) only hold under certain competitive games, while it also has to be assumed that marketing-mix elasticities are constant.

Good examples of the second type of research, which is more data-based, are Bowman and Gatignon (1996) and Chintagunta (1999). This research is explicitly based on observed market situations. Bowman and Gatignon (1996) study the effect of the order of entry on market shares and the effectiveness of marketing instruments. They show that the order of entry negatively influences the effectiveness of promotion and that it lowers price sensitivity. The main effects of the order of entry on own market share are found to be small, while in contrast there are strong effects of the order of entry on the effectiveness of marketing efforts. Bowman and Gatignon (1996) do not consider the effects of an entrant on the incumbent brands, while it would not be unlikely that a brand introduction also affects the effectiveness of marketing-mix variables of these other brands. Note that they do assume dependence between marketing-mix elasticities and the number of competitors, but they abstain from testing for changes in these elasticities, nor do they consider cross effects.

An example of a study that does consider changes in the marketing-mix effectiveness of incumbent brands is Chintagunta (1999), where the effects of entry are studied in the context of an individual choice model. A random effects multinomial logit model is used where brand intercepts are modeled by brand locations in attribute space with household-specific importance weights. A new brand introduces an additional brand position in the attribute space. As a consequence of this entry, several changes to the competitive structure may occur. First of all, locations of extant brands or importance weights may change, or both. It is shown that a brand introduction has a substantial impact on the

importance weights assigned to attributes. Only minor changes occur in brand positions and in the sensitivity to marketing activities. Chintagunta (1999) documents that, due to a new brand introduction, price sensitivity tends to increase while promotional sensitivity tends to decrease.

As can be understood from the discussion above, Bowman and Gatignon (1996) and Chintagunta (1999) only describe the demand side of the market. That is, these studies focus on the effects on elasticities, where these elasticities may change due to brand repositioning or changes in preference. The behavior on the supply side, by retailers and manufacturers, is not studied. An example of an empirical study, which does consider the effects of new entry on this side of the market is Robinson (1988). He presents an analysis of the reactions to an entrant in 115 different cases, where he shows that the most common reaction pattern to entry is no reaction or only a reaction with a single marketing instrument. However, as Basuroy and Nguyen (1998) suggest in a response to these findings, there is a need for further empirical analysis to support the theoretical results.

This brings us to the contribution of this chapter to the literature on the effects of market entry. We put forward several empirical methods to examine both sides of the market. We suggest statistical methods to validate the various predictions from normative studies. Our techniques are designed for weekly scanner time series data on market shares. We focus on market shares as we are interested in the relative performance of brands when a new brand is introduced, even when this introduction would lead to an increase in category sales. Naturally, our methods can be redesigned if one intends to focus on sales only. Such an approach could turn out to be useful in a category in which expansion effects play an important role. However, in a sales model, one needs to model seasonal effects and the category expansion or category contraction explicitly.

In order to analyze changes in actual behavior of a brand manager, that is, changes in the use of instruments like price and display, one needs methods to test for structural changes in time series variables. Upon doing so, we build on the findings in Srinivasan *et al.* (2000), who document that, except for possible level shifts, marketing time series data seem to be stationary. Hence, we also assume stationarity of all time series under scrutiny. Next, to examine possible changes in the effects of marketing-mix efforts, within the context of a market share attraction model, we propose new methods. First of all, we introduce a new estimation method that can handle a changing number of brands in the observation period. Next, we propose a method to test whether parameter values differ across the subsamples. To demonstrate the value added of our approach we perform a simulation study where we compare our method with various, naive, alternatives.

To summarize, this chapter contributes to the literature by the development of measurement tools to examine the effects of the entry and exit of brands, given the availability of a sample of the relevant time series data. As the exit of a brand mirrors an entry, we

only focus on the latter for brevity. Naturally, our methods can readily be adapted to the exit case.

It should be mentioned that we condition our analysis on the observed entrant's strategy. Hence, we study the effects of this strategy on a specific existing competitive structure, and we therefore cannot consider the effects of different strategies. There are papers where the entrant's strategy is explicitly considered, see for example Gatignon *et al.* (1990), Shankar (1999) and Shankar *et al.* (1999). However, when one were to apply our methods to a range of data sets, one might make generalizing statements about the observed behavior and its consequences.

We apply our methods to a set of weekly time series observations concerning the detergent category. For this market we observe the introduction of a new brand at about one-third of the sample. Upon application of our methods, we do not find supportive evidence for the hypothesis that prices are lowered after the introduction of a new brand. Next, we find that not many marketing efforts are increased, after the introduction of a new brand, where we should bear in mind of course that we consider only one market. We further find that part of the competitive structure changes after an introduction, thereby providing an incentive to further examine consumer response to brand introduction.

The outline of this chapter is as follows. In Section 3.2, we discuss the testing approach verbally, that is, without explicit formulas. In Section 3.3, we briefly discuss the attraction model, and we discuss parameter estimation in the attraction model, while taking into account the introduction of the new brand. Technical details are relegated to Appendix 3.A. In Section 3.4, we present a testing procedure to assess whether the introduction of the new brand results in a different competitive structure among the incumbent brands, where we put the technicalities in Appendix 3.B. Section 3.5 we present the results of a simulation experiment in which we compare the approach suggested in Section 3.4 with various alternatives. In Section 3.6, we discuss the testing procedure for breaks in the level of marketing instruments. The testing and estimation procedures are illustrated in Section 3.7 for the detergent category. We conclude this chapter in Section 3.8 with some remarks.

3.2 Testing approach

In this section, we outline the ideas behind our empirical approach, without laying out any technical issues. Our analysis of the effect of a brand introduction on the competitive structure is guided by the notion that part of the changes in market shares might directly be attributed to the marketing efforts of the new brand. This is in contrast to standard structural break analysis, where there is no endogenous effect at the time of the break. Hence, the question at hand is whether all the observed changes are due to the

marketing efforts of the new brand or whether the consumers have changed their behavior too. A third effect can be that incumbent brands adapt their marketing strategies. For example, as a reaction to the introduction, brand managers may decide to change their marketing strategy. Combinations of these effects are of course also possible.

Testing for changes in the use of marketing instruments is relatively straightforward. To test whether brands make less or more use of their marketing instruments, we will analyze breaks in the levels of the marketing instruments. This can be pursued using relatively standard time series techniques, although we will see below that one has to take into account that not all variables amount to continuous data.

Testing for changes in the competitive structure is more complicated, due to the fact that changes in the marketing mix and the competitive structure can occur simultaneously. One has to control for the direct effects of the new brands' marketing mix and for the effects of competitive response. For example, consider a market with two brands with an equal market share. The introduction of a third brand may cause that one brand loses more market share than the other. However, from this observation we cannot infer that the first brand is affected more by brand entry, as the second brand for example could have lowered its prices as a reaction to the entry, thereby obtaining market share from the first brand. Also, if we observe a price cut, we cannot conclude that all differences between the two incumbent brands can be assigned to this price change.

In sum, a substantive empirical analysis of the effects of a brand introduction calls for a model that jointly captures the pre-introduction and the post-introduction period. Separate models for the pre- and post-entry period are not very informative. Indeed, only in a combined model, it is possible to perform statistical tests on the constancy of parameters or on changes in the competitive structure. If one is only interested to see if there might be an effect at all, one can use two independent models. Technically speaking, all model parameters are then allowed to change, that is, the brand intercepts, parameters concerning all marketing instruments and the covariance matrix, and due to this, it is difficult, if not impossible, to find which aspect of the competitive structure has really changed.

As another strategy, one might be inclined to test for changes in a model where the new entrant is simply not included. This however is also not a good idea. The reason for this is that marketing instruments for the new brand generally have an indirect effect on the performance of the incumbent brands. So, even if the competitive structure among the incumbent brands remains constant, one could find changing parameters due to the effect of the marketing instruments of the new brand.

We therefore propose an alternative method, and our strategy can now be summarized as follows. We propose a model that concerns the periods before and after the entry. Below we will see that parameter estimation in such a model is not straightforward. The main complication is that market share models are developed for a constant number of

brands throughout the sample. To test for changes in a market, we need a model that incorporates the influence of the new brand by describing both periods and all brands. This chapter deals with the empirical analysis of such a model. The next three sections all deal with the (changing) competitive structure within the context of an attraction model. In Section 3.6, we turn to testing for changes in the use of the marketing mix.

3.3 Attraction Model

The competitive structure can be studied using either sales or shares. We choose to analyze market shares and to measure the effectiveness of marketing instruments of different brands using the familiar market share attraction model (Naert and Weverbergh, 1981; Leeflang and Reuyl, 1984; Cooper and Nakanishi, 1988; Bronnenberg *et al.*, 2000). An advantage of using market shares is that these are in general not influenced by category expansion or contraction. Unlike sales, market shares are therefore usually stationary series, see for example Srinivasan *et al.* (2000), which, by the way, also facilitates the statistical analysis. With a brand entry in an expanding market, sales of incumbent brands may stay constant and, consequently, their market shares may fall. It may also happen that category sales do not increase, and in that case the incumbent brands' market shares also fall. Hence, out of the two, it seems that market shares are the most interesting variable to consider. An additional advantage is that market shares usually do not show seasonal fluctuations as all brands are similarly affected by seasonal effects.

Below we discuss the representation and parameter estimation of the attraction model, for the special case where a brand is introduced during the sample period. We focus on the introduction of a single brand, but our method can easily be extended to the simultaneous introduction of multiple brands, to multiple independent brand introductions and to the exit of a brand. These extensions are not considered here to save space.

We first assume that the competitive structure does not change due to brand entry. We focus on an attraction model with constant parameters, where only the number of brands changes during the observation period. First, we need to introduce some notation. Suppose that the sample spans weeks 1 to T. Without loss of generality, denote the new brand as brand 1 and the time of the introduction of this brand as $T_1,\ 1 < T_1 < T$, where T_1 is known. The total number of brands in the market, after the introduction, will be denoted by I. So, before the introduction we have brands $i = 2, 3, \ldots, I$ and after the introduction we have the brands $i = 1, 2, \ldots, I$.

Following Cooper and Nakanishi (1988), the overall attraction of brand $i = 2, \ldots, I$ at time $t = 1, \ldots, T_1 - 1$, that is, prior to the introduction, is defined by

$$A_{it} = \exp(\mu_i + \varepsilon_{it}) \prod_{j=2}^{I} \prod_{k=1}^{K} x_{kjt}^{\beta_{kji}}, \tag{3.1}$$

see also Chapter 2. As usual, x_{kjt} denotes the k-th marketing instrument of brand j at time t, and the parameter associated with the effect of this instrument on brand i is denoted by β_{kji}. The μ_i are intercepts and the ε_{it} denote error terms. After the introduction of brand 1, the attraction of brand $i = 1, \ldots, I$ at time $t = T_1, \ldots, T$ is defined by

$$A_{it} = \exp(\mu_i + \varepsilon_{it}) \prod_{j=1}^{I} \prod_{k=1}^{K} x_{kjt}^{\beta_{kji}}. \tag{3.2}$$

Note that (3.2) allows the marketing instruments of brand 1 to influence the attractions of all incumbent brands in the post-introduction period. Further note that, for the moment, we assume that the effects of the marketing instruments for the incumbent brands are constant over the entire sample, as we assume the same parameters across (3.1) and (3.2). This assumption will be relaxed in Section 3.4, where we consider parameter estimation in a more flexible model, where the incumbent brands' parameters are allowed to change.

As usual, we assume a normal distribution for the errors. For the pre-introduction period we have $(\varepsilon_{2t}, \ldots, \varepsilon_{It})' \sim N(0, \Sigma^{(1)})$ and for the post-introduction period we have $(\varepsilon_{1t}, \ldots, \varepsilon_{It})' \sim N(0, \Sigma^{(2)})$, where the $^{(1)}$ and $^{(2)}$ superscripts denote parameters before and after the introduction, respectively. For the moment, we also assume that the covariances of the unexplained attractions among the incumbent brands do not change due to brand introduction. Therefore, $\Sigma^{(2)}$ can be partitioned as

$$\Sigma^{(2)} = \left(\begin{array}{c|ccc} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1I} \\ \hline \sigma_{12} & & & \\ \vdots & & \Sigma^{(1)} & \\ \sigma_{1I} & & & \end{array}\right). \tag{3.3}$$

The interpretation of this restriction is as follows. The covariance can be seen as a measure of brand similarity. If two brands have a strong positive correlation, consumers must evaluate these brands as similar. By assuming a constant covariance, we assume that the unobserved part of the relative positioning of the incumbent brands, that is, the part that is not explained by observed marketing instruments, remains unchanged.

According to the market share theorem (Bell *et al.*, 1975), a brand's market share is defined by its relative attraction. The market share of a brand is equal to its attraction relative to the total attraction of the market, that is

$$\begin{aligned} M_{it} &= \frac{A_{it}}{\sum_{j=2}^{I} A_{jt}}, \; i = 2, \ldots, I, \; t = 1, \ldots, T_1 - 1, \\ M_{it} &= \frac{A_{it}}{\sum_{j=1}^{I} A_{jt}}, \; i = 1, \ldots, I, \; t = T_1, \ldots, T, \end{aligned} \tag{3.4}$$

where M_{it} denotes the market share of brand i at time t. This completes the model.

As can been seen from (3.1), (3.2) and (3.4), the attraction model is nonlinear in its parameters. Fortunately, the model can be linearized to allow for relatively straightforward estimation, see also Section 2.2. By considering market shares relative to the market share of a base brand, for example brand I, and by taking natural logs we obtain a multi-equation model for log relative market shares which is linear in the parameters. This reduced-form model is defined in terms of relative parameters, that is in terms of $\tilde{\mu}_i = \mu_i - \mu_I, \tilde{\beta}_{kji} = \beta_{kji} - \beta_{kjI}$ and $\eta_{it} = \varepsilon_{it} - \varepsilon_{It}$, where $(\eta_{2t}, \ldots, \eta_{It})' \sim N(0, \tilde{\Sigma}^{(1)})$, for $t = 1, \ldots, T_1 - 1$ and $(\eta_{1t}, \ldots, \eta_{It})' \sim N(0, \tilde{\Sigma}^{(2)})$, for $t = T_1, \ldots, T$. It turns out that only these differences in parameters can be identified. In Appendix 3.A, we provide the technical details of parameter estimation and the identification.

3.4 Testing for shifts

In this section we discuss how various kinds of shifts in the competitive structure can be translated into the context of the attraction model. We discuss how one can test for these shifts. This section focuses on changes in aggregate consumer behavior. In Section 3.6, we will discuss testing for changes in the use of marketing instruments.

The previous discussion of the attraction model considered a stable competitive structure among incumbent brands. More technically stated, the parameters $\tilde{\mu}_i$ and $\tilde{\beta}_{kji}$ for $i = 2, \ldots, I-1$, $j = 2, \ldots, I$ and $k = 1, \ldots, K$ and the covariance matrix of the attractions of the incumbent brands are assumed constant during the sample period. In practice, one is often interested in actually testing these restrictions. For example, one might be interested in testing whether the competitive structure concerning the pricing strategy amongst the incumbent brands is affected by the new brand. For example, if the new brand has a very competitive price positioning, then consumers might become more price-sensitive, possibly also due to an increase in price competition amongst incumbent brands.

There are several interesting hypotheses that can be tested. For example, one can test whether shifts in relative market shares have taken place among the incumbent brands that cannot be attributed to changes in the level of the marketing instruments. This hypothesis corresponds to the parameter restriction $\tilde{\mu}_i^{(1)} = \tilde{\mu}_i^{(2)}$, $i = 2, \ldots, I-1$. One can also focus on the effect of a single specific marketing instrument, like the relative price. In that case, the test concerns $\tilde{\beta}_{kji}^{(1)} = \tilde{\beta}_{kji}^{(2)}$, $i = 2, \ldots, I-1$ and $j = 2, \ldots, I$ for a specific marketing instrument k. Finally, we can test for constant variance of the unexplained attractions of the incumbent brands. As stated before, the covariances can be interpreted as measuring the similarity of brands. Testing for the constancy of the covariance structure can therefore be interpreted as testing for constancy of the overall brand positioning.

To test the various hypotheses, we rely on the Likelihood Ratio [LR] test principle. To perform the tests, the parameters in both the restricted and the unrestricted models need to be estimated. In the restricted model, we impose that before and after the brand introduction the parameters are equal. For the unrestricted model, we allow some parameters to change. Note that we can also test hypotheses without estimating the unrestricted model by applying a Lagrange Multiplier test. However, in general we are not only interested in whether parameters change, but also in the actual parameter values before and after the brand introduction, or at least in the direction of change, and hence the need to be able to estimate unrestricted models.

A less restricted specification of the attraction model, allowing only a few parameters to change, can also be estimated using the strategy outlined in Appendix 3.A. In case of a possible shift in the brand-specific constant attractions or a shift in marketing instrument effectiveness, only the regressor design matrix needs to be reorganized. If the covariance matrix is allowed to change, then the estimation procedure for the covariance matrix of the attractions also changes. The most flexible case where all parameters, including the covariance matrix, change is discussed in Appendix 3.B. Estimation routines for more restricted versions are easily obtained from that discussion.

In case all parameters are allowed to change, the pre- and post-introduction models are completely independent. In this case the models for the pre- and post-introduction periods can be estimated independently using the usual estimation techniques. However, in the more realistic case that one wants to test for specific changes in the competitive structure, these standard estimation techniques do not work as the model for the pre-introduction period is no longer independent of the model for the post-introduction period. The estimation routine to be used for such models directly follows from the previous discussion by reorganizing the design matrices.

To test whether parameters indeed change after the introduction of a new brand, the value of the log likelihood at the restricted model parameter estimates is compared with the log likelihood evaluated at the parameter estimates for the more flexible model. The specific expressions for the likelihood functions are given in the appendices. In general, denote the value of the log likelihood at the restricted estimates as ℓ^r and the value of the log likelihood at the parameter estimates for unrestricted model as ℓ^u. Under the hypothesis of constant competitive structure, the difference $-2(\ell^r - \ell^u) \sim \chi^2_J$, where J equals the number of restricted parameters in the restricted model. If the log likelihood difference is sufficiently large, one concludes that the imposed restrictions are not valid, meaning that part of the competitive structure changed as a reaction to the brand introduction.

For the special case where the covariance matrix is assumed to be constant, one would be tempted to add to the model a number of dummy variables and interactions of the dummy variables with the marketing-mix variables. However, parameter estimation for this "dummy with interaction" model is not straightforward as the number of model

equations change due to the introduction. However one could choose to ignore the new brand and estimate a competitive model for the remaining brands. This approach is theoretically less elegant, but this is probably one of the ways the analysis of the effects of a brand introduction is done in practice. Therefore, in the next section, we will study the relative effectiveness of our suggested approach to various alternatives using a simulation study.

3.5 Simulation study

As an alternative to the method we suggest, one could choose to ignore some of the technicalities and test for changes in the competitive structure using other techniques. One of the possibilities is to ignore the newly introduced brand and estimate a competitive model for the incumbent brands only. Dependent on the chosen functional form of the attraction model it may matter if the researcher does or does not take into account the marketing mix of the new brand. For example if in the competitive structure at hand there are cross-price effects, the price of the new brand will have an effect on the relative shares of the incumbent brands.

For the general case where cross effects are present, the approach where the marketing-mix variables of the new brand are ignored can easily be rejected on theoretical grounds. In this case the resulting system of equations will be subject to the "omitted variables problem". That is, unless the cross effects are zero or there is no correlation between the marketing instruments, the parameter estimates for the post-introduction period will be biased. Proper testing in this system of equations will therefore be unnecessarily complicated in all practical settings.

Ignoring the market shares of the new brand amounts to omitting an equation in a system of (dependent) equations. The statistical tests for changes in the competitive structure associated with this system will be less powerful compared to a system in which all brands are considered. For example, an often imposed restriction on the competitive structure is that certain marketing instruments have the same effect on all attractions. The equation corresponding to the new brand then also contains information on this effect. Omission of this equation leads to more uncertainty in the parameter estimates, thereby leading to less powerful tests.

There are two options for the covariance matrix of the log relative market shares in an alternative testing approach. First, one could decide to ignore the correlation between market shares all together and to study the market shares as independent series. An advantage of ignoring the covariance structure is that this makes it easy to also take into account the newly introduced brand. However, with or without considering the new brand, such an approach will obviously give erroneous results in the case that (log

relative) market shares do show a strong correlation. Note that even when the attractions are independent, the log relative market shares will show correlation. Second, one can assume that the covariance structure does not change due to the introduction. In cases where the covariance structure does change, this approach will of course give biased results.

Summarizing, we can identify four ways to test for changes in the competitive structure, that is, (i) use our method proposed in Section 3.4, (ii) ignore the covariance matrix, the new brand's market shares, but use its marketing mix, (iii) ignore the covariance matrix, but use the new brand's market shares and marketing mix, (iv) account for a (constant) covariance matrix, ignore the market shares of the new brand, but do use the new brand's marketing mix.

To evaluate the relative performance of these four approaches, we consider a simulation experiment. For this experiment we consider a simple attraction model with price as the only marketing instrument. To obtain reasonable price series, we use the prices of 5 brands from the category that will be used in the empirical part of this chapter. In this hypothetical market four brands are observed for 37 weeks. In week 38, a new brand is introduced. After this introduction, we observe the market for another 96 weeks. The parameters of the model before the introduction, that is, the brand intercepts and the price effects are randomly drawn from the standard normal distribution, $N(0, 1)$. For the covariance matrix of the attraction errors we use the unit matrix. Note that in the reduced-form model for the log relative market shares this corresponds to a non-diagonal covariance matrix. After the introduction of the new brand some of the parameters may change. In our simulations, we add a draw from $N(0, \kappa^2)$ to the model parameters, where κ controls the magnitude of the change.

First, we consider a scenario in which only the price effect changes after the introduction. For simplicity we assume that there are no cross-attraction effects of price. That is, the price of a brand does not affect the attraction of another brand. To assess the relative performance of the different testing methods, we simulate market shares using our attraction model under different (randomly drawn) parameter settings. We then use the different testing procedures to test for a change in the price effects. The first panel in Table 3.1 gives the fraction of rejections of the null of no change in price effects of 1,000 simulations for the case where only the price effects may change. In case of no actual change in price effects ($\kappa = 0$), we should want to reject the null in 5% of the cases. From Table 3.1 we see that for all four tests the size of the test is close to 5%. For this simple case the power of the four tests does not differ much. Our method, however, performs slightly better than the three alternatives, in terms of proper size and higher power.

If we drop the assumption of a constant covariance matrix, the testing procedures do differ substantially. In a second simulation experiment we consider the case where the covariance matrix changes after the introduction. After a brand introduction one may expect increased variance in market shares possibly caused by increased uncertainty in

Table 3.1: Power and size of test on constancy of price effects versus magnitude of change (tests at the 5% significance level)

Test method[1]	Magnitude of change in price effects (κ) 0	0.05	0.1	0.15	0.25	0.5	0.75
Constant brand intercepts and constant covariance matrix							
(i)	0.053	0.348	0.725	0.874	0.967	0.996	0.997
(ii)	0.075	0.323	0.678	0.834	0.961	0.990	0.997
(iii)	0.078	0.326	0.679	0.837	0.960	0.990	0.996
(iv)	0.057	0.337	0.724	0.873	0.963	0.997	0.997
Changing covariance matrix[2]							
(i)	0.063	0.259	0.628	0.830	0.951	0.991	0.997
(ii)	0.023	0.137	0.449	0.706	0.907	0.980	0.994
(iii)	0.021	0.132	0.443	0.694	0.901	0.980	0.994
(iv)	0.012	0.139	0.501	0.767	0.934	0.986	0.995

[1] The test methods are defined as:
(i) Test method proposed in Section 3.4
(ii) Test based on an estimation method where the market shares of the new brand and the covariance of the attractions are ignored.
(iii) Test based on an estimation method where all market shares are used but where the covariance of the attractions are ignored.
(iv) Test based on an estimation method where the new brand's market shares are ignored, but where the covariance is taken into account.

[2] The variances of the attractions are doubled after the brand introduction.

household preferences. Such a change will cause the covariance matrix of the attractions to change. In the simulation experiment we capture this by multiplying the variances by a factor 2. In the second panel of Table 3.1 we present the simulation results under this scenario. The four methods now do differ substantially. First, the three alternative tests are quite undersized, that is, when there is in fact no change ($\kappa = 0$) the test rejects the null in fewer than 5% of the cases. In case there is a change in the price parameters, the alternative tests reject the null in fewer cases compared to the testing method we propose. With our method, the probability of correctly detecting a change in the price parameters is 10% higher.

As a final simulation experiment we consider the case where, due to the brand introduction, the covariance matrix and the brand intercepts may change next to the pricing parameter. We consider testing the constancy of the pricing parameters under four different scenarios. The covariance matrix of the attractions before the introduction is again set to unity, after the introduction the variances may double. For the brand intercepts we set $\kappa = 0.1$. For each of the four scenarios, we consider the number of rejections of the null hypothesis when the pricing parameters are in fact constant (the column "size" in Table 3.2) and the case where the pricing parameter changes from -2 to -1 for each brand (the column labeled "power" in Table 3.2). If necessary, and possible, we control for the changes in the brand intercepts and the covariance matrix when testing.

Table 3.2: Power and size of test on constancy of price effects under four different scenarios (tests at the 5% significance level)[1]

	Test method[2]	Σ constant Size	Σ constant Power	Σ changing Size	Σ changing Power
μ constant	(i)	0.056	0.281	0.072	0.246
	(ii)	0.067	0.204	0.019	0.088
	(iii)	0.067	0.208	0.021	0.077
	(iv)	0.058	0.279	0.017	0.121
μ changing	(i)	0.070	0.193	0.070	0.206
	(ii)	0.119	0.196	0.025	0.082
	(iii)	0.122	0.208	0.023	0.079
	(iv)	0.075	0.181	0.006	0.052

[1] Pricing effectiveness is set to -2, in the case where the effectiveness changes it equals -1 after the brand introduction. Σ, the attraction covariance matrix, is set to unity, and doubled after the introduction in case of a change in Σ. Finally the possible change in the brand intercepts corresponds to $\kappa = 0.1$.

[2] Test methods are as defined in Table 3.1.

The results in Table 3.2 clearly show that only for the case where neither the brand intercepts nor the covariance matrix change, all methods are correctly sized. For the

other cases, the alternative tests have quite large size distortions, with the exception of test (iv) for the case where only the brand intercepts change. This implies that, when using standard critical values, these tests cannot be used for valid inference. It is possible to obtain correct critical values for each alternative test under each scenario. However, this will require extensive simulation and the critical value will probably be different for each data set and model specification. Concerning the power of the tests, we see that our method generally has the best power. The largest differences are found when both the covariance matrix and the brand intercepts change. The alternative methods correctly identify the change in pricing parameters in at most 8% of the cases, whereas our proposed method correctly rejects the null hypothesis in over 20% of the cases.

Summarizing, we find that the methods, alternative to ours, work reasonably well in the case where the researcher knows that there are no changes in the competitive structure besides the set of parameters being tested. In all other cases, these tests have size distortions and low power compared with our method. In practice, one is therefore more likely to fail to reject the null hypothesis of no change when in fact the competitive structure has changed. Such a finding will also be reported in Section 3.7 for an actual data set.

3.6 Testing for breaks in marketing efforts

In the previous section we have discussed the testing of changes in the competitive structure. Such changes can mainly be attributed to the consumers. As discussed earlier, brand managers oftentimes also react to an entry. Next to testing whether the competitive structure has changed as a consequence of a brand introduction, one may also be interested in testing whether the incumbent brands adapt their marketing strategy as a response to the new competitor. For example, one may want to test whether the incumbent brands change the frequency of their displays and whether they make more use of price cuts. These tests may be used to provide some empirical validity to normative studies. For example, Gruca *et al.* (1992) suggest that in response to an entry, non-dominant brands having a market share smaller than 50% should lower prices and reduce their marketing efforts such as display and feature. The tests presented in this section can be used to empirically validate these conjectures.

To analyze the use of marketing instruments which are measured on a continuous interval, like price and advertising spending, we recommend the use of a Chow breakpoint test (Chow, 1960) in combination with an autoregressive model of order P [AR(P)] to correct for possible autocorrelation. This correction is important as only testing for a shift in the mean of the series is not sufficient. For example, consider a brand manager issuing a temporary price cut every other week. As a reaction to the brand introduction, s/he

might decide to have price cuts for two subsequent weeks, followed by two weeks of regular pricing, again followed by two weeks with a price cut, and so on. The average number of weeks with a price cut before and after the introduction is equal in this example. However, there is a structural break in the pattern of price cuts. Only testing for the mean will not identify this break, only if the dynamic structure of the process is taken into account is it possible to find such breaks. A rather simple way to identify the dynamics in the marketing process is the use of an AR(P) model. For our purposes, it does not seem necessary to consider very complicated models for the marketing instruments, see for example Klapper and Herwartz (2000) who show that, when forecasting market shares, simple models for marketing instruments work best.

The Chow test for a structural break is based on the sum of the squared errors (SSE) of three regressions. Let SSE_F denote the SSE of an AR model estimated on the full sample of T observations. Let SSE_1 denote the SSE of the same model estimated on the sample up to (but not including) the breakpoint T_1, and SSE_2 denote the SSE for the remaining sample. The test statistic is now defined as

$$F = \frac{(SSE_F - SSE_1 - SSE_2)/(1+P)}{(SSE_1 + SSE_2)/(T - P - 2(1+P))}. \tag{3.5}$$

under the null of no structural break the test statistic has an $F(1+P, T-P-2(1+P))$ distribution. In the case of a structural break, the separate models will fit the data much better than the model for the total sample implying that $SSE_1 + SSE_2$ will be much smaller than SSE_F. The test statistic F will therefore be large leading to a rejection of the null hypothesis of constant parameters.

Many marketing measures, like display and feature, are usually measured as 0/1 variables. The use of a Chow test in combination with an AR model is not useful for these variables, as the AR model typically assumes continuous data. We therefore use the logit model to analyze these 0/1 time series, and we explicitly include dynamics. The probability of $y_t = 1$, say a feature at time t, is modeled as

$$\Pr[y_t = 1] = F(\alpha + \sum_{p=1}^{P} \phi_p y_{t-p}) \tag{3.6}$$

where P is the maximum lag used and $F(\cdot)$ is the usual logit transformation, that is,

$$F(z) = \frac{\exp(z)}{1 + \exp(z)}. \tag{3.7}$$

With a binary series it is not possible to directly apply the Chow test as the residuals from a logit model are not easily defined. A measure like the sum of squared errors is therefore not easily obtained. However, we can use a test based on the Likelihood Ratio

principle. Let ℓ_r denote the log likelihood of a logit model estimated on the total sample assuming constant parameters. Further, let ℓ_u denote the log likelihood of a logit model with two unrelated sets of parameters, one for the periods before the brand introduction and one for the periods following the introduction. For the binary series, we use the test statistic $LR = -2(\ell_r - \ell_u)$ to test for a possible structural break. Under the null of no break the statistic has a χ^2_{1+P} distribution.

For the data we use for illustration in Section 3.7, some marketing instruments cannot be modeled by an AR(P) or a logit model. These instruments concern weighted indicators and are measured on the [0, 1] interval. These variables for example give the fraction of stores having the corresponding product on display. As these variables have many zero values and cannot be smaller than zero, straightforwardly fitting an AR model would lead to inappropriate inference. For these series, we use the Tobit model (Tobin, 1958), that is,

$$\begin{aligned} y_t^* &= \alpha + \sum_{p=1}^{P} \phi_p y_{t-p} + \varepsilon_t, \quad \text{where } \varepsilon_t \sim N(0, \sigma^2), \\ y_t &= \begin{cases} 0 & \text{if } y_t^* \leq 0 \\ y_t^* & \text{otherwise.} \end{cases} \end{aligned} \tag{3.8}$$

Testing for a structural break for this model can be done along the same lines as for the logit model.

3.7 Illustration

The use of an attraction model in case the sample contains a brand introduction is illustrated in this section using a data set on detergent. We also illustrate the use of some statistical tests for changes in the use of marketing instruments or for changes in competitive structure. The data set concerns twelve brands of liquid detergent, covering 134 weeks. One of these twelve brands ("Surf") is introduced in week 38.

As explanatory variables for the market shares of the brands in this market we have the price of each brand, denoted by P_{it}, which is the actual price paid by consumers. Furthermore, we have the fraction of stores having the product on display, the fraction of stores having featured the product and the fraction of stores in which a coupon could be redeemed. These variables are denoted by D_{it}, F_{it} and C_{it} respectively.

Table 3.3 gives an overview of various data characteristics. The first columns of this table give the average market share before and after the introduction of Surf. The average market share of the new brand is quite substantial. In the period after the introduction, the average share of this new brand is almost 12.5%. Many brands lose approximately 32%

of their market share to the entrant (Bold, Cheer, Dynamo, Era and Fab). On the other hand, some brands, Tide and Oxydol, have even gained market share. Proportionally, Solo has lost most market share (53%). Given the discussion earlier, we cannot conclude from these summary statistics that one brand appeared more sensitive to the introduction than another brand. The decrease or increase in market share could for example be caused by an (in-)efficient marketing plan to respond to the entry. More specifically, we are interested in testing whether the loss (or increase) of market share can be explained by the competition with the new brand, changes in the marketing strategy or by changes in the competitive structure.

Table 3.3: Data characteristics (averages over time) of detergent market

	Market share		Price		Display		Feature		Coupon	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Surf	–	12.47	–	5.04	–	1.76	–	0.99	–	6.16
Bold	5.49	3.38	5.80	5.82	0.25	0.11	0.37	0.37	5.53	3.27
Cheer	14.51	9.94	4.77	5.10	1.75	0.07	1.13	0.29	4.97	4.19
Dynamo	2.63	1.75	5.12	5.14	0.53	0.46	0.25	0.47	3.20	2.28
Era	8.79	6.11	6.15	6.18	0.39	0.46	0.41	0.40	5.27	2.52
Fab	2.17	1.49	5.21	5.40	0.02	0.05	0.22	0.21	3.79	2.34
Oxydol	10.02	10.26	5.13	5.21	0.03	0.71	0.04	0.48	4.93	4.91
Solo	2.97	1.39	6.23	6.20	0.03	0.22	0.06	0.05	5.27	1.87
Wisk	14.06	12.02	5.22	5.20	0.56	2.55	1.51	2.13	8.27	8.24
Yes	2.47	2.01	5.15	4.18	3.37	1.71	1.16	0.60	3.69	2.27
All	3.83	3.63	3.68	3.78	–[1]	0.06	0.20	0.17	2.30	3.04
Tide	30.88	33.37	5.13	5.06	1.12	1.53	1.14	1.14	7.06	7.39

[1] Marketing instrument was not used in this period.

Marketing instruments

We first analyze the series of marketing instruments separately. As discussed in Section 3.6, we use Chow tests to see whether brand managers have adapted their marketing strategy in a response to the introduction. As prices are measured on a continuous scale, we use AR models to capture the dynamics. For the display, feature and coupon variables

we use a Tobit model to capture the fact that these variables represent a percentage of stores.

Table 3.4 shows the p-values of Chow tests on the stability of the price, display, feature and coupon series of the twelve incumbent brands in the focal market, where an AR(2) model is used to correct for serial correlation in prices and where a Tobit model with one lag is used for display, feature and coupons. These test results and the summary statistics in Table 3.3 give an indication of the effect of the entrant on the use of marketing instruments. For example, the Chow tests show that Yes and Cheer have significantly changed their prices. Indeed, Table 3.3 shows that Yes has dramatically decreased its price by almost 19% and that Cheer increased its price with 7%. The other brands seem to have kept a constant price. Cheer and Oxydol have changed the use of displays, Cheer has decreased displays while for Oxydol we observe an increase. All uses displays after the introduction but did not use them before. This change can therefore also be called significant, but note that, also after the introduction, All makes only little use of this instrument. Most brands make little use of this instrument, while the new brand uses display quite often. Brands are also not often featured in this market. Changes in the use of feature are therefore not large. Again, only Cheer seems to have decreased their use of feature. Coupons are relatively often used in the observed detergent market. The new brand again heavily uses this instrument, although the other brands have decreased the use of coupons (except All and Tide). We observe significant decreases in the use of coupons for Bold, Era, Fab and Solo. A significant increase is only observed for All.

For our data set it is interesting to consider Cheer as it is manufactured by Proctor and Gamble, while the new entrant is a Lever Bros. brand. For Cheer, we observe a significant increase in price and significant decreases in the use of display and feature. These changes in the use of marketing instruments are not consistent with what is expected based on the literature. For example, Gruca *et al.* (1992) show that as a response to entry, dominant brands (market share greater than 50%) should reduce prices and increase advertising spending, while non-dominant brands should reduce prices and reduce the use of other marketing variables. But, while we observe a decrease in the use of display and feature, we do not find a decrease in prices for Cheer.

In general, we find that if a brand changes the use of a non-price marketing instrument, it usually decreases its use. Shankar (1997) pointed out that the results in Gruca *et al.* (1992) are based on the assumption of a particular competitive game and, maybe even more important, on the assumption of no consumer reactions. Our model allows for changing consumer reactions, and this might perhaps generate the findings. In sum, however, we find empirical support for the conclusions in Robinson (1988), which is that the most common reaction to entry is no reaction or a reaction with just a single instrument.

Table 3.4: p-values of Chow-tests on structural breaks in the marketing efforts using AR(2) models for prices and Tobit models with AR(1) structure for display, feature and coupon.

	Price	Display	Feature	Coupon
Bold	0.967	0.225	0.265	0.031
Cheer	0.012	0.009	0.050	0.285
Dynamo	0.465	0.408	0.623	0.154
Era	0.234	0.756	0.404	0.021
Fab	0.105	0.892	0.768	0.019
Oxydol	0.232	0.001	0.253	0.136
Solo	0.602	0.802	0.759	0.001
Wisk	0.830	0.391	0.601	0.267
Yes	0.005	0.728	0.970	0.310
All	0.158	-[1]	0.814	0.028
Tide	0.098	0.597	0.350	0.572

[1] Instrument used after the introduction of Surf, but not before.

An attraction model

To further analyze the effect of the entry of Surf on the detergent market and to find out whether the entry affected the competitive structure, we consider a market share attraction model. This way, we can compare the effectiveness of marketing instruments before and after the introduction. For this illustration we use the basic attraction framework in (3.1) and (3.2). To save parameters, we assume that the price of brand i does not influence the *attraction* of brand j, $i \neq j$. Note that this does not imply that the *market share* of brand j is also independent of the marketing instruments of brand i, see Section 2.2.2. In (3.1) and (3.2), this restriction gives $\beta_{kji} = 0$ for $i \neq j$. This restriction implies that all remaining marketing instrument parameters are now identified, that is, $\beta_{kjj} = \tilde{\beta}_{kjj}$ and $\beta_{kII} = -\tilde{\beta}_{kIj},\ j = 1, \ldots, I-1$. The competitive structure for the effects of display and feature are even more restricted as we additionally assume an equal effect for every brand. This restriction is necessary because all brands make little use of these two instruments. There simply is not enough information in our data set to estimate all brand-specific effects.

The display, feature and coupon variables cannot be directly included in the attraction model, as in that case zero display or zero use of feature would lead to a zero market share, see (3.1). We apply an exponential transformation to these variables, so instead of D_{it}, F_{it} and C_{it} we include $\exp(D_{it}), \exp(F_{it})$ and $\exp(C_{it})$ as marketing instruments in the attraction specification. For the moment, we use independent models before and after the introduction of Surf. The brand intercepts, effectiveness of the marketing instruments as well as the covariance of the unexplained attractions are allowed to change due to the introduction.

Table 3.5 contains the estimates of the brand intercepts and marketing instrument parameters for the reduced forms of these two models, see (3.9) and (3.10). The covariance structures before and after the introduction are also estimated independently. However to save space, we do not report the estimates of these matrices. The main conclusion from a comparison between the covariance matrices is that in general the variance of the unexplained attractions increased after the introduction. As a result of the introduction, the uncertainty about market shares apparently has increased.

The parameters in Table 3.5 give the effect of the marketing instruments on the attraction of the brands. The intercepts give the relative position of a brand to the chosen base brand Tide. Significant parameters for marketing-mix variables have the expected sign. Price decreases the attraction of a brand. Display, feature and coupons increase the attractiveness. The introduction of the new brand shows to have a large impact on the model parameters. Due to the introduction, all the intercept signs have changed, hence after the introduction all brands have lost share relative to Tide. In fact, Tide is one of the two incumbent brands with larger market share after the introduction, see Table 3.3. The price and coupon parameters also show major changes. The display and feature parameters seem to be relatively constant.

Tests for constancy

To statistically test whether there are significant structural changes in the competitive structure, we perform the tests discussed in Section 3.4 on the different parts of the model. First, we test whether the effects of measured marketing efforts are affected by the introduction. First, we perform the tests independently for each marketing instrument while allowing the remaining model parameters to change at the moment of introduction. For the detergent data set, the results show that the competition concerning price and coupons has changed due to the introduction of Surf (p-values both very close to zero). The p-values concerning feature and display are 0.178 and 0.696 respectively. The effectiveness of coupons has changed quite dramatically after the introduction of Surf. Not only do the coupons of Surf have a large impact on its market share, the effectiveness of coupons

Table 3.5: Parameter estimates for an attraction model, where pre- and post-introduction periods are estimated independently (standard errors in parentheses, significant parameters in boldface)

	Display		Feature	
	Pre	Post	Pre	Post
All brands	**0.027**	**0.020**	**0.055**	**0.033**
	(0.012)	(0.007)	(0.010)	(0.009)

	Intercept		Price		Coupon	
	Pre	Post	Pre	Post	Pre	Post
Surf		**−4.743**		**−3.093**		**0.108**
		(1.782)		(0.500)		(0.010)
Bold	6.442	**−6.971**	−2.128	**−2.407**	**0.090**	**0.244**
	(4.050)	(2.700)	(1.656)	(0.212)	(0.009)	(0.027)
Cheer	**19.173**	−2.362	**−9.844**	**−4.477**	**0.072**	**0.116**
	(4.159)	(2.402)	(2.382)	(1.116)	(0.016)	(0.010)
Dynamo	**9.489**	**−13.608**	**−4.676**	0.821	**0.136**	**0.253**
	(3.428)	(4.195)	(0.954)	(2.340)	(0.020)	(0.046)
Era	**12.585**	−2.862	**−4.992**	**−3.937**	**0.059**	**0.126**
	(3.055)	(2.953)	(0.851)	(1.378)	(0.007)	(0.017)
Fab	6.998	**−7.754**	**−3.468**	−2.700	**0.206**	**0.321**
	(3.849)	(2.854)	(1.679)	(1.421)	(0.037)	(0.056)
Oxydol	1.040	−3.196	1.507	**−3.945**	**0.105**	**0.110**
	(3.943)	(2.548)	(1.673)	(1.116)	(0.013)	(0.007)
Solo	12.496	**−18.679**	−5.790	3.501	**0.098**	**0.332**
	(6.309)	(5.207)	(3.016)	(2.709)	(0.040)	(0.058)
Wisk	**15.954**	**−6.590**	**−7.246**	**−1.825**	**0.025**	**0.056**
	(3.090)	(2.206)	(1.023)	(0.880)	(0.009)	(0.008)
Yes	2.991	**−10.935**	−1.102	−1.152	**0.095**	**0.315**
	(4.068)	(1.967)	(1.857)	(0.807)	(0.045)	(0.057)
All	21.218	**−6.974**	−14.275	**−2.786**	**0.108**	**0.156**
	(10.681)	(2.299)	(7.956)	(1.147)	(0.039)	(0.017)
Tide	0*	0*	2.799	**−5.131**	**0.068**	**0.052**
	–	–	(1.666)	(0.971)	(0.008)	(0.008)

* Restricted for identification

also increased for all other brands (except Tide). Note that the usage of coupons did not change much, see Table 3.3.

The changes in the effectiveness of price have different signs across the brands, see Table 3.5. For many brands we see a reduction in the price effectiveness. For three brands we observe an increase in price effectiveness (Bold, Oxydol and Tide), the change appears significant for Oxydol and Tide. Note that these three brands are all owned by Proctor and Gamble and that the entrant is a Lever Bros. brand. Apparently, Proctor and Gamble had the opportunity to better use its price. The effectiveness of display and feature does not change. As these two instruments are not used very often, we can only comment on the change in the average effectiveness. It may be the case that for some individual brands the effectiveness did change. However we cannot test for these changes using the current data.

Imposing the constancy of display and feature, we now test whether the intercepts and/or the covariance matrix change. The constancy of the intercepts and the constancy of the covariance matrix are both rejected (p-values indistinguishable from zero). This final result indicates that the relative positioning of the brands, as perceived by the consumer, has changed. Finally, we test if the restrictions hold jointly, that is, we test for the constancy of feature and display. This restriction cannot be rejected (p-value 0.244).

To summarize, for this data set we find that the use of marketing instruments is less sensitive to brand introduction than would have been expected, given the prevalent literature. Based on game-theoretic analysis using the market share attraction model (for example Basuroy and Nguyen, 1998; Gruca *et al.*, 1992), we would have expected more breaks in the marketing-mix variables. But, with Robinson (1988), we conclude that competitive response in practice can be limited. One of the reasons might be that in practical situations the evaluation of consumers and their sensitivity to marketing instruments may also change in reaction to a brand introduction. In game-theoretic analysis, these consumer reactions are usually ignored. The consumers, however, so we seem to find, do show to react strongly to the introduction, or to the unobserved changes in the market. We find that the effectiveness of coupons strongly increased. The effectiveness of price to influence market share has decreased for many brands. However, for two P&G brands, we observe a significant increase in the price effectiveness. The covariance matrix of the unexplained attractions also changed, indicating that the (unobserved) brand positioning was also affected by the introduction. A final conclusion is that we cannot assume constancy of the brand intercepts.

Finally we compare the results obtained with our testing procedure with the those that were discussed and analyzed using simulated data in Section 3.5. We have performed the same tests using the three alternative testing methods. The p-values for the tests are presented in Table 3.6. The first row of Table 3.6 gives the earlier presented p-values. Note

that, although our method strongly suggests the rejection of a constant covariance matrix, none of the other methods allows for a change in the covariance matrix. Concerning price, we see that if the covariance of the attractions is not accounted for, we would conclude that the competitive structure for pricing had not changed due to the introduction of Surf. The different tests for display, feature and coupon would lead to the same conclusion. However, test (iv) would lead us to believe that some change in display effectiveness had occurred (p-value 0.074). Imposing the constancy of display and feature competition, we again test for constancy of the brand intercepts. The tests where the covariance matrix is ignored again lead to different conclusions. These two tests would lead us to believe that the brand intercepts have not changed. The final joint tests for constancy of feature and display do not lead to different conclusions.

Table 3.6: Comparison of testing methods for the detergent category (p-values of tests of no change in competitive structure)

Test method*	price	display	feature	coupon	brand intercepts	Σ	display & feature
(i)	0.000	0.696	0.178	0.000	0.000	0.000	0.244
(ii)	0.146	0.202	0.354	0.000	0.093	–	0.347
(iii)	0.126	0.209	0.330	0.000	0.076	–	0.337
(iv)	0.004	0.074	0.620	0.000	0.002	–	0.196

* See Table 3.1 for a description of the test methods

3.8 Conclusion and discussion

In this chapter we proposed methods to empirically analyze the effects of a brand introduction on the competitive structure, where we focus on weekly observed market shares and marketing instruments. We suggested a number of statistical tests that can be used to judge whether or not the brand introduction affects the competitive structure among incumbent brands. Tests for the constancy of the marketing strategies themselves were also presented. If incumbent brands respond to the introduction by means of increased promotion or price cuts, the market shares of different brands might change, but this does not automatically imply that the competitive structure changes. Hence we argued that

only changes in elasticities or cross elasticities correspond to a structural change in the competitive structure.

To be able to statistically test for these kinds of structural breaks, we developed a model for competition in which the pre- and post-introduction periods can be estimated simultaneously. In this chapter we choose to relate competition to market shares. The familiar market share attraction model can then be used as a basis for the two period model. In such a model we can allow for all kinds of different structural changes. Furthermore we have shown that various alternative methods that ignore some of the technicalities involved with this testing approach do not perform as good as the method we have proposed.

In the illustration, we stressed the importance of jointly estimating the two periods by showing that at least some part of the competitive structure remains unchanged. If all available data are used, these constant parameters can be estimated more accurately. For our illustrative data set, we do not find supportive evidence for the hypothesis that prices decrease after an introduction, as has been suggested in game-theoretic studies. An explanation for this could be that the reaction of consumers to the introduction is ignored in these studies. Another explanation could be that brand managers do not respond optimally.

The presented methodology can be used to examine the consequences of brand introduction on competitive structure. Future empirical work should indicate whether generalizing statements can be made about this issue.

3.A Parameter estimation for the case of constant parameters

The linear equivalents of (3.1) and (3.2) are

$$\begin{aligned} \ln M_{it} - \ln M_{It} &= \tilde{\mu}_i + \sum_{j=2}^{I}\sum_{k=1}^{K} \tilde{\beta}_{kji} \ln x_{kjt} + \eta_{it}, \\ i &= 2, \ldots, I-1,\ t = 1, \ldots, T_1 - 1, \end{aligned} \tag{3.9}$$

which is an $I-2$ equations model, and

$$\begin{aligned} \ln M_{it} - \ln M_{It} &= \tilde{\mu}_i + \sum_{j=1}^{I}\sum_{k=1}^{K} \tilde{\beta}_{kji} \ln x_{kjt} + \eta_{it}, \\ i &= 1, \ldots, I-1,\ t = T_1, \ldots, T, \end{aligned} \tag{3.10}$$

which is an $I-1$ equations model, where

$$\begin{aligned} \tilde{\mu}_i &= \mu_i - \mu_I \\ \tilde{\beta}_{kji} &= \beta_{kji} - \beta_{kjI} \\ \eta_{it} &= \varepsilon_{it} - \varepsilon_{It}, \end{aligned} \tag{3.11}$$

for $i = 1, \ldots, I-1,\ j = 1, \ldots, I,\ k = 1, \ldots, K$ and $t = 1, \ldots, T$. As market shares are only determined by relative attractions, the levels of the parameters themselves are not identified. In fact, one can only identify the parameters relative to a base brand.

The distributional assumptions on ε_{it} imply that the distributions for the disturbances in the reduced-form model become $\eta_t^{(1)} = (\eta_{2t}, \ldots, \eta_{It})' \sim N(0, \tilde{\Sigma}^{(1)})$, for $t = 1, \ldots, T_1 - 1$ and $\eta_t^{(2)} = (\eta_{1t}, \ldots, \eta_{It})' \sim N(0, \tilde{\Sigma}^{(2)})$, for $t = T_1, \ldots, T$ where $\tilde{\Sigma}^{(1)} = L_{I-2}\Sigma^{(1)}L'_{I-2}$ and $\tilde{\Sigma}^{(2)} = L_{I-1}\Sigma^{(2)}L'_{I-1}$ with $L_k = (\mathbf{I}_k | -\mathbf{i}_k)$, with $\mathbf{I}_k$ a k-dimensional identity matrix and $\mathbf{i}_k$ a k-dimensional unit vector. The $^{(1)}$ and $^{(2)}$ notation is used to indicate vectors, matrices or parameters concerning the pre- or post-introduction period, respectively. The covariance matrix of the reduced-form disturbances after the introduction ($\tilde{\Sigma}^{(2)}$) can be partitioned similarly as $\Sigma^{(2)}$, that is,

$$\tilde{\Sigma}^{(2)} = \left(\begin{array}{c|ccc} \tilde{\sigma}_{11} & \tilde{\sigma}_{12} & \ldots & \tilde{\sigma}_{1,I-1} \\ \hline \tilde{\sigma}_{12} & & & \\ \vdots & & \tilde{\Sigma}^{(1)} & \\ \tilde{\sigma}_{1,I-1} & & & \end{array}\right). \tag{3.12}$$

To facilitate the discussion of parameter estimation, we cast the reduced-form model in a compact matrix notation. To this end, we introduce the following design matrices

$$\begin{aligned} X_t^{(1)} &= (\mathbf{0}_{I-2}, \mathbf{I}_{I-2}) \otimes (1, \mathbf{0}'_K, \ln x'_{2t}, \ldots, \ln x'_{It}),\ t = 1, \ldots, T_1 - 1 \text{ and} \\ X_t^{(2)} &= \mathbf{I}_{I-1} \otimes (1, \ln x'_{1t}, \ln x'_{2t}, \ldots, \ln x'_{It}),\ t = T_1, \ldots, T, \end{aligned} \tag{3.13}$$

where $\mathbf{0}_s$ denotes a $s \times 1$ vector of zeros, $x_{it} = (x_{1it}, \ldots, x_{Kit})'$ and $\otimes$ is the Kronecker product. Furthermore, we introduce as vectors of dependent variables

$$\begin{aligned} y_t^{(1)} &= (y_{2t}, \ldots, y_{I-1,t})',\ t = 1, \ldots, T_1 - 1 \\ y_t^{(2)} &= (y_{1t}, \ldots, y_{I-1,t})',\ t = T_1, \ldots, T, \end{aligned} \tag{3.14}$$

where $y_{it} = \ln M_{it} - \ln M_{It}$. In matrix notation, the reduced-form equations (3.9) and (3.10) now become

$$\begin{pmatrix} y_1^{(1)} \\ \vdots \\ y_{T_1-1}^{(1)} \\ y_{T_1}^{(2)} \\ \vdots \\ y_T^{(2)} \end{pmatrix} = \begin{pmatrix} X_1^{(1)} \\ \vdots \\ X_{T_1-1}^{(1)} \\ X_{T_1}^{(2)} \\ \vdots \\ X_T^{(2)} \end{pmatrix} \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \vdots \\ \tilde{\beta}_{I-1} \end{pmatrix} + \begin{pmatrix} \eta_1^{(1)} \\ \vdots \\ \eta_{T_1-1}^{(1)} \\ \eta_{T_1}^{(2)} \\ \vdots \\ \eta_T^{(2)} \end{pmatrix} \tag{3.15}$$

where

$$\begin{aligned} \tilde{\beta}_i &= (\tilde{\mu}_i, \tilde{\beta}_{11i}, \ldots, \tilde{\beta}_{K1i}, \tilde{\beta}_{12i}, \ldots, \tilde{\beta}_{K2i}, \ldots\ldots, \tilde{\beta}_{1Ii}, \ldots, \tilde{\beta}_{KIi})', \\ \eta_t^{(1)} &= (\eta_{2t}, \ldots, \eta_{I-1,t})',\ \text{for } t = 1, \ldots, T_1 - 1 \text{ and} \\ \eta_t^{(2)} &= (\eta_{1t}, \ldots, \eta_{I-1,t})',\ \text{for } t = T_1, \ldots, T. \end{aligned} \tag{3.16}$$

For further reference, we denote (3.15) as $y = X\tilde{\beta} + \eta$.

As the disturbances are assumed to be independent over time, η is normally distributed with mean 0 and covariance matrix Ω, which is defined by

$$\Omega = \begin{pmatrix} \mathbf{I}_{T_1-1} \otimes \tilde{\Sigma}^{(1)} & 0 \\ 0 & \mathbf{I}_{T-T_1+1} \otimes \tilde{\Sigma}^{(2)} \end{pmatrix}. \tag{3.17}$$

In case the covariance matrix Ω is known, the estimate of $\tilde{\beta}$ can be found by the GLS estimator

$$\widehat{\tilde{\beta}} = [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}y. \tag{3.18}$$

In general, the covariance matrix Ω is not known. In that case, we recommend the use of an iterative estimation technique similar to SUR estimation (Zellner, 1962). Based on an estimate $\Omega_{(n)}$ of Ω, we use (3.18) to obtain the estimate $\tilde{\beta}_{(n)}$ of $\tilde{\beta}$, where the subscript (n) denotes the parameter estimate in the n-th iteration of the estimation routine. Using $\tilde{\beta}_{(n)}$ we get a new estimate $\Omega_{(n+1)}$ by maximizing the reduced-form model log-likelihood

$$\begin{aligned} \ln \mathcal{L} = -\frac{(T_1 - 1)(I - 2) + (T - T_1 + 1)(I - 1)}{2} \ln 2\pi \\ - \frac{1}{2} \ln |\Omega| - \frac{1}{2}(y - X\tilde{\beta}_{(n)})'\Omega^{-1}(y - X\tilde{\beta}_{(n)}) \end{aligned} \tag{3.19}$$

with respect to the elements of Ω. The special structure of Ω in (3.17) establishes that

$$\Omega^{-1} = \begin{pmatrix} \mathbf{I}_{T_1-1} \otimes \tilde{\Sigma}^{(1)^{-1}} & 0 \\ 0 & \mathbf{I}_{T-T_1+1} \otimes \tilde{\Sigma}^{(2)^{-1}} \end{pmatrix}, \tag{3.20}$$

and

$$|\Omega| = |\tilde{\Sigma}^{(1)}|^{T_1-1} |\tilde{\Sigma}^{(2)}|^{T-T_1+1}. \tag{3.21}$$

Finally, denote the residuals based on $\tilde{\beta}_{(n)}$ by $\eta_{(n)} = y - X\tilde{\beta}_{(n)}$. Maximizing (3.19) is equivalent to maximizing

$$-(T_1 - 1) \ln |\tilde{\Sigma}^{(1)}| - (T - T_1 + 1) \ln |\tilde{\Sigma}^{(2)}| - \eta_{(n)}{}' \Omega^{-1} \eta_{(n)}. \tag{3.22}$$

This maximization can be done by any general purpose maximization routine like for example the BFGS method. The maximizer of (3.22) gives $\Omega_{(n+1)}$, a new estimate of Ω.

The iterative estimation scheme starts with taking $\Omega_{(1)}$ equal to the identity matrix. By iterating over (3.18) and the maximization of (3.22), we obtain the Feasible Generalized Least Squares estimates $\widehat{\tilde{\beta}}$ and $\widehat{\tilde{\Sigma}}^{(l)}$ $(l = 1, 2)$, of $\tilde{\beta}$ and $\tilde{\Sigma}^{(l)}$, respectively.

3.B Parameter estimation in unrestricted models

In this section we discuss the estimation procedure for an attraction model with a brand introduction for the case where all parameters are allowed to change due to the introduction. Note that in this case one may also estimate models for the pre- and post-introduction periods separately. However, when estimating the model under the restriction that some part of the competitive structure remains fixed, the two periods are no longer independent. Each hypothesis presented in Section 3.4 corresponds to certain parameter restrictions in the model below.

Introduce the following reorganized design matrix for the pre-introduction period

$$\bar{X}_t^{(1)} = \mathbf{I}_{I-2} \otimes (1, \ln x'_{2t}, \ldots, \ln x'_{It}),\ t = 1, \ldots, T_1 - 1 \tag{3.23}$$

The design matrix for the period after the introduction remains unchanged, see $X_t^{(2)}$ in (3.13). In matrix notation, the reduced-form model with changing parameters is

$$\begin{pmatrix} y_1^{(1)} \\ \vdots \\ y_{T_1-1}^{(1)} \\ y_{T_1}^{(2)} \\ \vdots \\ y_T^{(2)} \end{pmatrix} = \begin{pmatrix} \bar{X}_1^{(1)} & 0 \\ \vdots & \vdots \\ \bar{X}_{T_1-1}^{(1)} & 0 \\ 0 & X_{T_1}^{(2)} \\ \vdots & \vdots \\ 0 & X_T^{(2)} \end{pmatrix} \begin{pmatrix} \tilde{\beta}_2^{(1)} \\ \vdots \\ \tilde{\beta}_{I-1}^{(1)} \\ \tilde{\beta}_1^{(2)} \\ \tilde{\beta}_2^{(2)} \\ \vdots \\ \tilde{\beta}_{I-1}^{(2)} \end{pmatrix} + \begin{pmatrix} \eta_1^{(1)} \\ \vdots \\ \eta_{T_1-1}^{(1)} \\ \eta_{T_1}^{(2)} \\ \vdots \\ \eta_T^{(2)} \end{pmatrix}, \tag{3.24}$$

where $\tilde{\beta}_i^{(1)},\ i = 2, \ldots, I-1$ denotes all parameters concerning the log relative market share of brand i before the introduction, and $\tilde{\beta}_i^{(1)},\ i = 1, \ldots, I-1$ denotes the parameters after the introduction. Only after the introduction we have an effect of the marketing mix of the entrant, therefore the dimension of $\tilde{\beta}_i^{(1)}$ is smaller than the dimension of $\tilde{\beta}_i^{(2)}$. For further reference, we denote (3.24) as $y = \bar{X}\tilde{\beta}^* + \eta$. Note that there are no restrictions on the parameters before and after T_1. The covariance matrix of η is

$$\bar{\Omega} = \begin{pmatrix} \mathbf{I}_{T_1-1} \otimes \tilde{\Sigma}^{(1)} & 0 \\ 0 & \mathbf{I}_{T-T_1+1} \otimes \tilde{\Sigma}^{(2)} \end{pmatrix}. \tag{3.25}$$

It is important to note that now $\tilde{\Sigma}^{(1)}$ is not a submatrix of $\tilde{\Sigma}^{(2)}$. Based on an estimate $\bar{\Omega}_{(n)}$ of $\bar{\Omega}$, the estimator $\tilde{\beta}^*_{(n)}$ for $\tilde{\beta}^*$ is the standard GLS estimator

$$\tilde{\beta}^*_{(n)} = [\bar{X}'\bar{\Omega}^{-1}_{(n)}\bar{X}]^{-1}\bar{X}'\bar{\Omega}^{-1}_{(n)}y. \tag{3.26}$$

Again, denote the residuals by $\eta_{(n)} = y - \bar{X}\tilde{\beta}^*_{(n)}$. A new estimate of $\bar{\Omega}$ can be obtained by maximizing

$$-(T_1 - 1)\ln|\tilde{\Sigma}^{(1)}| - (T - T_1 + 1)\ln|\tilde{\Sigma}^{(2)}| - \eta_{(n)}{}'\bar{\Omega}^{-1}\eta_{(n)}. \tag{3.27}$$

The last part, ${\eta_{(n)}}'\bar{\Omega}^{-1}\eta_{(n)}$, can be rewritten as

$$\eta_{(n)}^{(1)'}(I_{T_1-1}\otimes\tilde{\Sigma}^{(1)^{-1}})\eta_{(n)}^{(1)}+\eta_{(n)}^{(2)'}(I_{T_1-1}\otimes\tilde{\Sigma}^{(2)^{-1}})\eta_{(n)}^{(2)}. \tag{3.28}$$

Using this equation and the fact that $\tilde{\Sigma}^{(1)}$ and $\tilde{\Sigma}^{(2)}$ are assumed to be completely independent, we can analytically give the optimizers of (3.27), which are

$$\begin{aligned}\tilde{\Sigma}_{(n+1)}^{(1)} &= \frac{1}{T_1-1}\sum_{t=1}^{T_1-1}\eta_{(n),t}^{(1)}\eta_{(n),t}^{(1)'}\\ \tilde{\Sigma}_{(n+1)}^{(2)} &= \frac{1}{T-T_1+1}\sum_{t=T_1}^{T}\eta_{(n),t}^{(2)}\eta_{(n),t}^{(2)'}\end{aligned} \tag{3.29}$$

where $\eta_{(n),t}^{(j)}$ is defined as the vector of residuals in the reduced-form model at time t based on the parameter estimates in the n-th iteration, and where $j=1$ for $t=1,\ldots,T_1-1$ and $j=2$ for $t=T_1,\ldots,T$. If we can assume that the covariance structure remains unchanged, so that $\tilde{\Sigma}^{(1)}$ is a submatrix of $\tilde{\Sigma}^{(2)}$, the estimates in (3.29) are not appropriate. Instead we should numerically maximize (3.27) over the elements of the largest covariance matrix $(\tilde{\Sigma}^{(2)})$.

Chapter 4

Explaining dynamic effects of the marketing mix on market shares

4.1 Introduction

In recent literature on market structures it has been shown that marketing efforts, such as for example temporary price promotions, do not have permanent effects on sales or market shares. A prerequisite for permanent effects of temporary promotions is non-stationarity of sales or market shares. Srinivasan *et al.* (2000), Nijs *et al.* (2001), and Pauwels *et al.* (2002), among others, have shown that almost all sales series for fast moving consumer goods are stationary. Lal and Padmanabhan (1995), Dekimpe and Hanssens (1995a), and Franses *et al.* (2001) report similar results for market shares. Hence, to study dynamic effects of the marketing mix, one needs to examine the cumulative effect of a temporary promotion on current and future market shares. Only when the cumulative effect is positive a promotion is worthwhile.

In this chapter we put forward a method which allows us to directly estimate the potentially differing short-run and long-run marketing-mix effects on market shares. The short-run effect is defined as the instantaneous effect of a promotion on current market shares. The long-run effect is defined as the cumulative effect of a temporary promotion on current and future market shares, see also Pauwels *et al.* (2002)[1]. If a promotion has positive carry-over effect, the long-run effect will exceed the short-run effect. The long-run effect will be zero if the positive direct effect of a promotion is exactly balanced by negative carry-over effects.

[1]Note that this definition differs from the usual approach in the marketing literature. There, the common definition of the long-run effect is the effect of a temporary promotion on market shares in the distant future. However, as discussed above, such a permanent effect is hardly found. Indeed, in case of stationarity only permanent changes of the marketing mix will affect market shares in the long run.

The long-run and short-run effects of the marketing mix usually differ across brands and markets. Differences in promotional intensities, price structures or market concentration may lead to different market structures, see also Mela *et al.* (1998), Bronnenberg *et al.* (2000) and Srinivasan *et al.* (2000). In this chapter we aim to understand the potentially differing long-run and short-run marketing-mix effects on market shares. For that purpose, we link both the short-run and the long-run effects to brand-specific and category-specific characteristics in a second level of our model.

As market shares are in between zero and one, and also as they sum to unity, models for these dependent variables are a little more complicated than basic regression models. One might now be inclined to circumvent this complication by modeling own brand sales and category sales, and simply dividing the outcomes. Fok and Franses (2001), however, show that this might complicate matters even more as these two components of market shares are not independent. Additionally, depending on the model specification for category sales, this approach would also not guarantee that shares will sum to one. Another motivation to consider market shares is that subsequent models allow us to link market share elasticities directly to model parameters.

A useful model for market shares, when measured at, say, the weekly level, is the so-called market share attraction model. This model has theoretically sound properties and it is also easy to analyze in practice, see Cooper and Nakanishi (1988) for an early introductory book on this model. In Chapter 2 we have presented a review of its econometric aspects. In this chapter we will also consider this model.

We introduce into the marketing literature two new modifications of the attraction model. The first modification amounts to explicit expressions of long-run and short-run effects for the reduced-form attraction model, which is typically used to estimate the relevant parameters. This may seem like a trivial issue, but as we will demonstrate in Section 4.2.2, it is not. The second modification concerns the introduction of a second level in our model. That is, we propose to simultaneously analyze the attraction model for I_c brands in category $c = 1, \ldots, C$. This can lead to a multitude of parameters. For parsimony, but also for interpretation purposes, we therefore correlate some of these parameters with brand-specific and category-specific variables. The resultant model is a Hierarchical Bayes Attraction model. As such, we extend on a similar route taken by the rigorous study in Nijs *et al.* (2001), who instead consider a two-step approach and focus on (category) sales, whereas we put everything into a single model and consider brands' market shares. We also use explicit measures of dynamic effects, instead of derivative measures such as the impulse-response function. That said, the empirical results obtained from this chapter can be seen as adding to the knowledge base created by Nijs *et al.* (2001).

The outline of this chapter is as follows. In Section 4.2, we put forward our new two-step attraction model in so-called error-correction format, which allows us to analyze dynamic marketing-mix effects across categories. In Section 4.3, we apply our Hierarchical

Bayes Attraction model to weekly data for two to four brands in seven different categories at two locations. One of our conclusions is that only display and feature promotions are likely to have a higher long-run effect than short-run effect, while price elasticities are, in absolute value, higher for the short run. That is, price promotions often have negative carry-over effects, while display and feature tend to have positive carry-over effect. In Section 4.4, we conclude with a discussion of managerial and modeling implications.

4.2 Attraction models with dynamic effects

The basic attraction model contains two components. The first component is a specification for the (unobserved) attraction of a specific brand, which can depend on current and past marketing-mix instruments and past market shares or past attraction. The second component defines the market share by dividing own attraction by the sum of the attractions of all brands in a category. Together this leads to a reduced-form attraction model with parameters that can be estimated using the relevant data.

In this section we put forward an attraction model specification with a reduced-form model that can be converted into error-correction format. This error-correction model [ECM] enables us to disentangle long-run from short-run effects of marketing-mix variables on market shares in a direct way. The derivation of these effects is discussed in Section 4.2.2. In Section 4.2.3, we discuss our Hierarchical Bayes [HB] specification which is used to summarize the information on long-run and short-run effects for a large number of categories. By considering multiple markets in a single model, we can provide empirical generalizations concerning the dynamic effects of elements of the marketing mix in a statistically efficient way.

Before we discuss our complete model, we first review some notational issues in Section 4.2.1, where we confine ourselves to a single category to save notation. In this section we furthermore consider various dynamic specifications of the attraction model.

4.2.1 Preliminaries

To model market shares we define the attraction of brand i, out of I brands in a single category, at time t by A_{it}. We assume that attraction can be described by

$$A_{it} = \exp(\mu_i + \varepsilon_{it}) x_{it}^{\alpha_i} x_{i,t-1}^{\beta_i} M_{i,t-1}^{\rho}, \qquad \text{for } i = 1, \ldots, I, \tag{4.1}$$

where $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{It})' \sim N(0, \Sigma)$, and where M_{it} denotes the market share of brand i, and x_{it} a marketing instrument. This dynamic attraction specification is often successfully applied to model market shares, see among others Leeflang and Reuyl (1984); Cooper and Nakanishi (1988); Kumar (1994) and Bronnenberg *et al.* (2000). It is however difficult to

directly disentangle short-run effects and long-run effects of the marketing instruments on attractions and market shares using this parameterization. Another specification that has been suggested is based on including lagged attraction instead of lagged market share as an explanatory variable. However it turns out that such a specification is equivalent to (4.1) and has the same interpretation problems. Carpenter *et al.* (1988) and Hanssens *et al.* (1989) first transform the marketing instruments using an AR model to yield so-called effective advertising or promotion. The transformed instruments are then included in the attraction model. The main advantage of this approach is the reduction in parameters. The disadvantage is however that the short-run and long-run effects cannot be directly captured into single parameters. In the remainder of this section we therefore continue with specification (4.1) and show that it is possible to assign short-run and long-run effects of x_{it} to separate parameters. To keep notation simple, we present our model assuming there is a single marketing instrument. An extension to more explanatory variables is straightforward and will be presented in Section 4.2.3.

The second component of an attraction model amounts to the definition of market share as the relative attraction of a brand in the market, that is,

$$M_{it} = \frac{A_{it}}{\sum_{j=1}^{I} A_{jt}}. \tag{4.2}$$

Components (4.1) and (4.2) together lead to estimable reduced-form models.

The market share attraction model can be linearized by considering the natural logs of the market shares relative to a base brand. The model then reduces to

$$\begin{aligned} \ln \frac{M_{it}}{M_{It}} &= \ln M_{it} - \ln M_{It} = \ln A_{it} - \ln A_{It} \\ &= \mu_i - \mu_I + \alpha_i \ln x_{it} - \alpha_I \ln x_{It} + \beta_i \ln x_{i,t-1} - \beta_I \ln x_{I,t-1} \\ &\quad + \rho(\ln M_{i,t-1} - \ln M_{I,t-1}) + \varepsilon_{it} - \varepsilon_{It}, \end{aligned} \tag{4.3}$$

for $i = 1, \ldots, I-1$ and where brand I is chosen as the base brand, see Section 2.4. For parameter identification the choice of the base brand turns out to be arbitrary. The reduced-form specification in (4.3) shows that not all parameters in (4.1) can be identified. First, only the differences across the brand intercepts $\tilde{\mu}_i \equiv \mu_i - \mu_I$ are identified. Next, we can only identify the covariance structure of $\tilde{\varepsilon}_{it} \equiv \varepsilon_{it} - \varepsilon_{It}$, that is, the covariance matrix of $(\tilde{\varepsilon}_{1t}, \ldots, \tilde{\varepsilon}_{I-1,t})'$ denoted by $\tilde{\Sigma}$. The parameters in the resulting system of $I-1$ equations can now easily be estimated.

4.2.2 Short-run and long-run effects

In this chapter we consider the short-run and long-run effects that are implied by the dynamic structure of the model. This is in contrast with studies by, for example, Mela

et al. (1998) and Jedidi *et al.* (1999) where dynamics enter through the model parameters. In these studies the preferences and marketing sensitivity of households may change as a consequence of (intensified) promotional activities. In this case the long-run effect is defined as the impact of a promotion on the future accounting for changes in individuals behavior. In this chapter we take a different approach and consider (aggregated) household behavior to be constant. The dynamics in market shares are directly caused by feedback loops in household behavior.

It is well known that it is difficult to interpret the parameters in autoregressive distributed lag models as they combine short-run and long-run effects of the explanatory variables on the dependent variable. We will now show how to rewrite the attraction model in the interpretable error-correction format. To our knowledge we are the first to use the error-correction model in the context of the market share attraction model. In the marketing literature the error-correction model, has been used by for example Franses (1994) and Paap and Franses (2000) for new product sales and brand choice, respectively.

Consider again the attraction specification in (4.1) which leads to the $I-1$ equations in (4.3). This equation is an Autoregressive Distributed Lag [ADL(1,1)] model for the variable $(\ln M_{it} - \ln M_{It})$. To determine the dynamic effects of lagged x_{it} and x_{It} on the market shares we solve (4.3) for $(\ln M_{it} - \ln M_{It})$ by repeated substitution until the first observation. The solution is

$$\begin{aligned}\ln M_{it} - \ln M_{It} &= \rho^t(\ln M_{i0} - \ln M_{I0}) \\ &+ \sum_{\tau=0}^{t-1} \rho^\tau(\tilde{\mu}_i + \alpha_i \ln x_{i,t-\tau} - \alpha_I \ln x_{I,t-\tau} + \beta_i \ln x_{i,t-\tau-1} - \beta_I \ln x_{I,t-\tau-1} + \tilde{\varepsilon}_{i,t-\tau}). \quad (4.4)\end{aligned}$$

The long-run market shares follow from (4.4) by taking $t \to \infty$. Under the stationary condition $|\rho| < 1$, the influence of the market shares at time 0 disappears over time as $\lim_{t\to\infty} \rho^t = 0$. If we further set the explanatory variables at fixed values over time, that is, $x_{it} = x_i$ and $x_{It} = x_I$ for all t, the long-run market shares are now given by

$$\ln M_i - \ln M_I = \frac{\tilde{\mu}_i}{1-\rho} + \frac{\alpha_i + \beta_i}{1-\rho} \ln x_i - \frac{\alpha_I + \beta_I}{1-\rho} \ln x_I + \sum_{\tau=0}^{\infty} \rho^\tau \tilde{\varepsilon}_{i,t-\tau}. \quad (4.5)$$

As $\mathrm{E}[\tilde{\varepsilon}_{it}] = 0$ for all t, the long-run expectation of $(\ln M_i - \ln M_I)$, given x_i and x_I, equals

$$\mathrm{E}[\ln M_i - \ln M_I | x_i, x_I] = \frac{\tilde{\mu}_i}{1-\rho} + \frac{\alpha_i + \beta_i}{1-\rho} \ln x_i - \frac{\alpha_I + \beta_I}{1-\rho} \ln x_I, \quad (4.6)$$

and the long-run conditional variance is given by

$$\mathrm{Var}[\ln M_i - \ln M_I | x_i, x_I] = \sum_{\tau=0}^{\infty} \rho^{2\tau} \mathrm{Var}[\tilde{\varepsilon}_{i,t-\tau}] = \frac{\mathrm{Var}[\tilde{\varepsilon}_{it}]}{1-\rho^2}. \quad (4.7)$$

This implies that, in the long-run, market shares are determined by the following attraction specification

$$A_i = \exp\left(\frac{\mu_i}{(1-\rho)} + \eta_i\right) x_i^{\gamma_i} \qquad \text{for } i = 1, \ldots, I, \tag{4.8}$$

where $\gamma_i = (\alpha_i + \beta_i)/(1-\rho)$ and $(\eta_1, \ldots, \eta_I)' \sim N(0, \frac{1}{(1-\rho^2)}\Sigma)$. Interestingly, as the long-run market shares correspond to an attraction model, we can use the standard results in Cooper and Nakanishi (1988) to compute long-run (cross)-elasticities. For example, the long-run elasticity of x_i is given by

$$\frac{\partial M_i}{\partial x_i}\frac{x_i}{M_i} = \gamma_i(1 - M_i), \tag{4.9}$$

which can provide useful information for managers who need to decide on the marketing mix.

It follows immediately from (4.4) that under stationarity ($|\rho| < 1$) the effect of a temporary change of x_i at time τ has no permanent impact on market shares as the term ρ^τ will be zero for large τ. Only a permanent change in the value of x_i will have a permanent long-run effect on the market shares. The long-run effect on the log relative market shares is measured by the parameter γ_i. A temporary change of x_i does however have a short-run effect on market shares. The direct short-run effect is measured by α_i. To disentangle the long-run effects from the short-run effects of x_i on market shares, that is, to allow for directly estimating these effects, it is convenient to rewrite (4.3) in error-correction format, see Hendry *et al.* (1984), that is,

$$\begin{aligned}\Delta(\ln M_{it} - \ln M_{It}) = \tilde{\mu}_i + \alpha_i \Delta \ln x_{it} - \alpha_I \Delta \ln x_{It} + \\ (\rho - 1)[\ln M_{i,t-1} - \ln M_{I,t-1} - \gamma_i \ln x_{i,t-1} + \gamma_I \ln x_{I,t-1}] + \tilde{\varepsilon}_{it},\end{aligned} \tag{4.10}$$

where $\gamma_i = (\alpha_i + \beta_i)/(1-\rho)$ and Δ denotes the first-differencing operator, that is, $\Delta y_t = y_t - y_{t-1}$. The short-run, or instantaneous, effects are given by α_i as

$$\frac{\partial \ln M_{it} - \ln M_{It}}{\partial \ln x_{it}} = \alpha_i. \tag{4.11}$$

The long-run relation between x_{it} and M_{it} is put in the so-called error-correction term and hence long-run effects of $\ln x_{it}$ on $\ln M_{it}$ are given by γ_i. That is, this parameter gives the marginal effect of a permanent change in $\ln x_{it}$ on the log relative market shares in the long-run. The parameter $(\rho - 1)$ is often called the adjustment parameter and determines the speed of convergence to the long-run relation. It can be shown that γ_i in error-correction model (4.10) is also equal to the cumulative effect of a temporary change

in $\ln x_{it}$ on current and future log relative market shares, that is, under stationarity the following property holds

$$\sum_{\tau=0}^{\infty} \frac{\partial(\ln M_{i,t+\tau} - \ln M_{I,t+\tau})}{\partial \ln x_{it}} = \gamma_i. \tag{4.12}$$

The error-correction model is nonlinear in some parameters. This is no problem as we can estimate (4.3) and transform the estimates to the parameters of (4.10). If one uses a Seemingly Unrelated Regression [SUR] estimator to estimate the parameters, standard errors can be obtained using the Delta method, see Greene (1993, p. 297). In this chapter we will use Bayesian methods, and hence parameter uncertainty naturally follows from our sampling output, but we will return to this issue further below.

It is perhaps of interest to mention that the reduced-form model in (4.10) can also be derived from an alternative starting point. Consider the attraction component,

$$A_{it} = \exp(\mu_i + \varepsilon_{it}) x_{it}^{\alpha_i} x_{i,t-1}^{\beta_i} A_{i,t-1}^{\rho}, \tag{4.13}$$

where now lagged attraction instead of lagged market share enters the specification. If we take logarithms on both sides, we obtain a system of I equations

$$\ln A_{it} = \mu_i + \rho \ln A_{i,t-1} + \alpha_i \ln x_{it} + \beta_i \ln x_{i,t-1} + \varepsilon_{i,t}. \tag{4.14}$$

If we solve for A_{it}, the long-run expectation of the attraction for brand i is given by $\mathrm{E}[A_i|x_i] = (\mu_i + \gamma_i \ln x_i)/(1-\rho)$ and the long-run variance is $\mathrm{Var}[A_i] = \mathrm{Var}[\varepsilon_{it}]/(1-\rho^2)$. Hence, starting from (4.13) the long-run attractions also satisfy (4.8). The error-correction specification for the attractions is given by

$$\Delta \ln A_{it} = \mu_i + \alpha_i \Delta \ln x_{it} + (\rho - 1)(\ln A_{i,t-1} - \gamma_i \ln x_{i,t-1}) + \varepsilon_{it}, \tag{4.15}$$

which implies that we can make direct statements about the attractions, and not only on the market shares.

So far, we have only considered first order attraction model specifications. Extensions to higher order attraction models are straightforward. The resulting error-correcting specifications are similar to (4.10) but with extra lagged values of $\Delta \ln M_{it}$, $\Delta \ln M_{It}$, $\Delta \ln x_{it}$ and $\Delta \ln x_{It}$. Furthermore, the model can be extended to allow for more flexible cross-effects of marketing instruments. Above, we have assumed that the marketing instruments of brand i do not impact the attraction of brand j ($j \neq i$), that is, we have imposed the Restricted Competition restriction as presented in Section 2.2.2. In case cross-attraction effects are allowed for the marketing instruments of all brands will enter (4.10). However, this will lead to a large number of parameters and will complicate the analysis of the long

and short-run effects of the marketing mix. Therefore we will restrict ourselves to the Restricted Competition specification.

Extending the model to allow for more flexible effects of lagged market shares may also seem feasible. One may for example want to allow the ρ-parameter to differ across brands. This will however yield a model for which it is impossible to derive the dynamical properties. For this model it is not possible to solve the system of equations for the market shares, like is done in (4.4). However, as we have shown in Section 2.8, in practice one often cannot reject the hypothesis of equal ρ-parameters for all brands.

4.2.3 Hierarchical Bayes

Now we turn to an analysis of the error-correction attraction model with a large number of categories. Let $A_{it}(c)$ and $M_{it}(c)$ denote the attraction and market share, respectively, of brand i in product category c in week t. In this section we consider multiple marketing instruments. Let $x_{ikt}(c)$ denote the k-th explanatory variable of brand i in category c in week t. An attraction specification of brand i in category c is given by

$$A_{it}(c) = \exp(\mu_{ic} + \varepsilon_{it}(c)) M_{i,t-1}^{\rho_c} \prod_{k=1}^{K} \left(x_{ikt}(c)^{\alpha_{ick}} x_{ik,t-1}(c)^{\beta_{ick}} \right) \tag{4.16}$$

for $i = 1, \ldots, I_c$, $t = 1, \ldots, T_c$ and $c = 1, \ldots, C$ with $\varepsilon_t(c) \sim N(0, \Sigma_c)$, where we now allow for multiple marketing instruments indexed by $k = 1, \ldots, K$. This attraction specification corresponds to a similar set of $I_c - 1$ linear equations as given in (4.3). We allow for the fact that categories may differ in the number of brands and the number of observed periods. The linear equations can be written in the error-correction model in a similar as (4.10), that is,

$$\begin{aligned} \Delta(\ln M_{it}(c) - \ln M_{I_c,t}(c)) = \tilde{\mu}_{ic} + \sum_{k=1}^{K} (\alpha_{ikc} \Delta \ln x_{ikt}(c) - \alpha_{I_c,kc} \Delta \ln x_{I_c,kt}(c)) + \\ (\rho_c - 1) \Big(\ln M_{i,t-1}(c) - \ln M_{I_c,t-1}(c) + \\ \sum_{k=1}^{K} [-\gamma_{ick} \ln x_{ik,t-1}(c) + \gamma_{I_c,ck} \ln x_{I_c,k,t-1}(c)] \Big) + \tilde{\varepsilon}_{it}(c), \end{aligned} \tag{4.17}$$

for $i = 1, \ldots, I_c - 1$ and $t = 1, \ldots, T_c$ and where $\gamma_{ikc} = (\alpha_{ikc} + \beta_{ikc})/(1 - \rho_c)$.

To relate the short- and long-run elasticity parameters to explanatory variables, we define K-dimensional vectors $\alpha_{ic} = (\alpha_{i1c}, \ldots, \alpha_{iKc})'$ and $\gamma_{ic} = (\gamma_{i1c}, \ldots, \gamma_{iKc})'$. The long-run and short-run effects of the marketing mix will obviously differ across brands and across categories. Some of these differences can be attributed to observable characteristics of the brand and the category, such as the size of a brand and the average use of a marketing

instrument. Another part of the effects of the marketing mix cannot be explained, either by the fact that it is specific to the brand or that there are characteristics that we do not observe. In sum, we propose to describe the short-run and long-run effects parameters by

$$\alpha_{ic} = \lambda_1' z_{ic} + \eta_{ic} \tag{4.18}$$

$$\gamma_{ic} = \lambda_2' z_{ic} + \nu_{ic}, \tag{4.19}$$

where z_{ic} is an L-dimensional vector containing an intercept and $L-1$ explanatory variables for brand i in category c, like promotion frequency of brand i in category c, a market leader dummy and so forth. The $L \times K$ matrices λ_1 and λ_2 give the effects of the brand characteristics on the short-run and long-run parameters, respectively. The error terms η_{ic} and ν_{ic} are assumed as uncorrelated across brands and normally distributed with mean 0 and covariance matrix Σ_η and Σ_ν, respectively. Note that there are $\sum_{c=1}^{C} I_c$ vectors α_{ic} and γ_{ic}.

To estimate the parameters in the model (4.17) with (4.18)–(4.19), we use a Bayesian approach. Bayesian estimation provides exact inference in finite samples. To obtain posterior results we use the Gibbs sampling technique of Geman and Geman (1984) which is a Markov Chain Monte Carlo [MCMC] technique. In the Appendix we derive the likelihood function of the model together with the full conditional posterior distributions which are necessary in the Gibbs sampler.

Another estimation strategy which is often applied in practice, is a two-step procedure in which, first, individual market-level models are estimated and, in a second stage regression, the parameters from the market-level models are related to brand and market characteristics, see for example Nijs *et al.* (2001). This method is however theoretically less elegant as the uncertainty in the first-level parameter estimates is not correctly accounted for in the second stage. In finite samples, this may lead to underestimation of the uncertainty in the parameter estimates in the second stage. At the end of our empirical section, we will briefly discuss the relevant differences between our Hierarchical Bayes approach and the two-step approach.

4.3 Empirical results

In this section we first discuss the data and the variables in the two components in our HB-ECM-attraction model. Then we elaborate on a few prior conjectures about the signs of the correlations in the second component of our model. Finally, we present the estimation results.

4.3.1 Data and variables

For the empirical part of this chapter we consider the so-called ERIM database of the GSB of the University of Chicago. The data concern seven different categories in two geographical areas. So we have 14 different markets. For each category we have weekly observations of the market shares and of the marketing efforts of the major national brands and a rest category. On average, we have 123 weekly observations for each category. Two markets concerning sugar have just two brands. The tuna category has three brands and the remaining five categories (catsup, peanut butter, stick margarine, tube margarine and tissues) each have four brands. We model the market shares of all 50 brands simultaneously using our HB model.

As explanatory variables for the market shares in the first model component we use a dummy variable for coupon promotion, a dummy variable for the combination of feature and display promotion and the actual price. The price parameters therefore describe price elasticities and not promotional price elasticities. The dummy variables cannot directly enter our attraction specification (4.1), as in that case weeks with no promotion would by definition have zero market shares. Instead, we use an exponential transformation for these two 0/1 marketing instruments. Finally, we use a lag order of 1 to capture the dynamics in the markets, which effectively leads to the model discussed in Section 4.2.3. Previous studies show that this lag order is sufficient to capture the dynamics, see Table 2.1 for an overview of lag orders used in the literature and Table 2.4 for the lag orders selected for the same data set using the model selection strategy suggested in Chapter 2.

For the second model component, where we correlate the long-run and short-run effects of the marketing instruments with category- and brand-specific variables, we construct five covariates. Four of these covariates are brand-specific, these are, relative price, coupon intensity, display/feature promotion intensity and a 0/1 dummy variable for the market leader. The market leader is set as the brand having the largest market share averaged over time. The coupon and promotion intensity variables equal the observed weekly frequency of the use of coupons and promotions, respectively. Finally, the relative price is defined as the average price divided by the maximum average price in the market. The brand that, on average, has the highest price therefore has a relative price equal to one. Note that it is important to use a relative measure of price effectiveness as some categories are more expensive than others.

The fifth and final covariate is defined at the category level and it measures market concentration. As the concentration index, we take the so-called entropy measure, that is,

$$\mathrm{CI}_c = \sum_{i=1}^{I_c} \bar{M}_i(c) \ln \bar{M}_i(c), \tag{4.20}$$

where $\bar{M}_i(c)$ denotes the average market share of brand i in market c. If all market power is concentrated in one brand, the concentration index equals 0. The index decreases when power is spread over more brands.

4.3.2 Some a priori conjectures

As said, with our model we aim to provide empirical results that might add to the knowledge base, created in Nijs *et al.* (2001), see also Raju (1992) and Jedidi *et al.* (1999). Concerning the effects of price, we conjecture that a higher marketing-mix intensity has a positive effect on instrument effectiveness, see Nijs *et al.* (2001). Hence, for example, more promotions increase the effects of price changes. There are no strong theoretical reasons why these increases should differ across the long-run and short-run impact of the instruments.

Next, more market concentration would have a positive impact on the absolute price elasticity. Hence, the more concentrated is the market, the more negative is the price effect. Nijs *et al.* (2001) report a significant impact of market structure for short-run effects, and an insignificant effect for those in the longer run.

Finally, for the leading brands one would expect that marketing-mix elasticities are smaller in absolute sense. Evidence for this conjecture is found in Bolton (1989) and Srinivasan *et al.* (2001), where it is shown that brands with smaller market shares tend to have larger price elasticities.

Concerning the dynamic properties of display and feature promotion we are not aware of any direct evidence. However, van Heerde *et al.* (2000) do find some dynamic effects of display and feature promotion. They report differences in the dynamic effects of price under four types of support (no support, feature only, display only, feature and display support). Indirectly this implies that display and feature also have dynamic effects. From the tables in van Heerde *et al.* (2000) one can conclude that display and feature have positive carry-over effects. In our setting we therefore also expect the long-run effect of display and feature to be larger than the corresponding short-run effect.

4.3.3 Estimation results

We estimate our model using MCMC techniques, where we use 10,000 iterations as burn-in. Of the next 100,000 iterations, we retain each tenth draw to obtain an approximately random sample from the posterior distribution. Our posterior results are based on the resulting 10,000 draws.

Figure 4.1 shows a histogram of the posterior means per brand of the short-run effects (α_{ikc}), the long-run effects (γ_{ikc}) and the differences between these two effects ($\gamma_{ikc} - \alpha_{ikc}$) for each of the marketing instruments and for all brands. Note that, in the classical

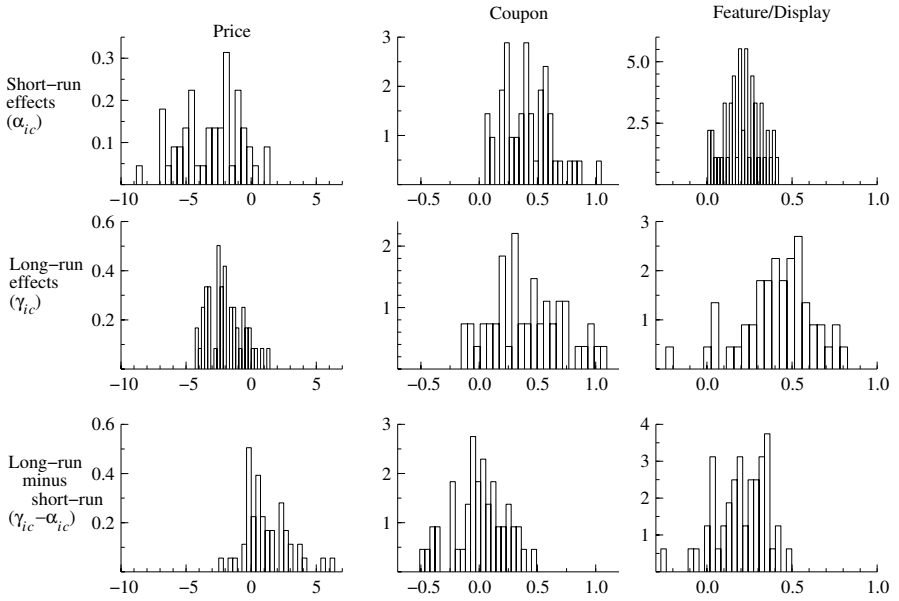

Figure 4.1: Histograms of posterior means of marketing-mix effectiveness, for all fifty brands

sense, these effects are not parameters of the model. They should be seen as latent variables. The histograms show the distribution of the expected values of these latent variables conditional on the market characteristics and the observed market shares. As expected, most of the mean price effects are negative and most of the mean coupon and display/feature promotion effects are positive. The posterior mean short-run effects over all brands are -3.127, 0.414, and 0.213 for price, coupon promotion, and feature/display, respectively. The long-run posterior mean effects equal -1.974 for price, 0.416 for coupon promotion and 0.415 for feature/display. Interestingly, the variation of the long-run effect of price is smaller than the corresponding short-run effect, while for feature/display we find the opposite outcome. For feature/display, it holds that the mean long-run effects tend to be larger than the mean short-run effects. For price, we find the opposite, that is, the short-run effects tend to be larger (in absolute size) than the long-run effects. On average, the long-run and short-run effects of coupon promotion seem to be equal in size. Whether these eyeball impressions stand a statistical test will be seen below.

Figure 4.2 shows how the short-run effects are related to the long-run effects. For all three variables, we notice a positive correlation between the short-run and long-run effects. That is, brands for which a marketing instrument has a large short-run effect (in

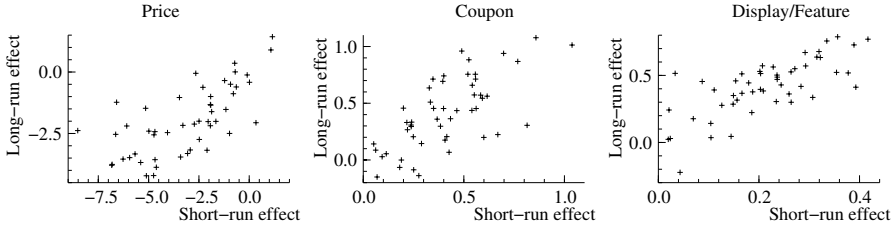

Figure 4.2: Scatter plots of long-run effects versus short-run effects (posterior means per brand), for all fifty brands. Short-run effects are given by α_{ic}, long-run effects equal γ_{ic}, see (4.17)

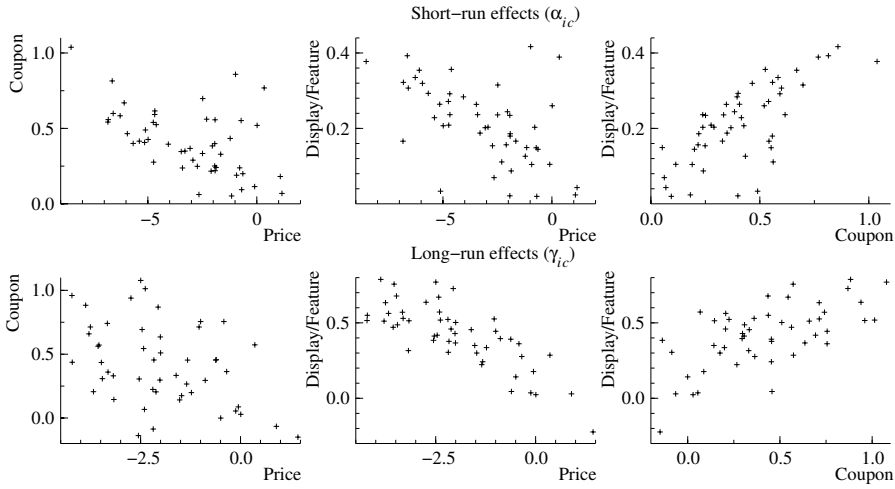

Figure 4.3: Scatter plots of posterior means of marketing-mix effectiveness

absolute sense), the corresponding long-run effectiveness is also large, on average. This correlation seems strongest for price and the combination of feature and display.

Figure 4.3 shows scatter plots of the mean posterior effects for different combinations of marketing instruments. For some combinations we find strong correlations. It is noteworthy to mention that there is strong correlation between the effectiveness of feature/display and coupon at the short run and price and feature/display at the long run.

In Table 4.1, we present the posterior estimates of the effects of covariates on the marketing effectiveness. For the short-run effects, we find substantive interactions between the price effect and various brand characteristics. Higher priced brands and brands that

Table 4.1: Posterior means of the effects of covariates on short-run and long-run effects of the marketing mix (λ_1 and λ_2 in (4.18) and (4.19)), posterior standard deviation between brackets.

	Marketing-mix effectiveness					
	Price		Coupon		Display/Feature	
	Short-run effects (λ_1)					
Intercept	-5.952***	(2.010)	0.918***	(0.233)	0.173	(0.158)
Feature/Display intensity	-3.138*	(1.566)	-0.090	(0.210)	-0.069	(0.129)
Coupon intensity	-3.210**	(1.378)	-0.045	(0.169)	0.054	(0.112)
Relative price	-0.222**	(0.115)	-0.027*	(0.015)	0.004	(0.009)
Market concentration	-5.242**	(1.979)	0.372*	(0.227)	-0.062	(0.159)
Market leader	1.030	(0.836)	-0.004	(0.103)	-0.074	(0.065)
	Long-run effects (λ_2)					
Intercept	-3.241**	(1.592)	1.185***	(0.340)	0.472*	(0.263)
Feature/Display intensity	-0.756	(1.107)	-0.435	(0.301)	-0.194	(0.213)
Coupon intensity	-2.632***	(0.960)	-0.062	(0.232)	0.257	(0.183)
Relative price	-0.215***	(0.072)	-0.032	(0.021)	0.005	(0.016)
Market concentration	-2.462	(1.567)	0.384	(0.326)	0.051	(0.267)
Market leader	0.813	(0.534)	-0.127	(0.131)	-0.087	(0.111)
Pr[|Long run| > |Short run|]	0.283		0.544		0.872	

*, **, *** Zero not contained in 90%, 95% or 99% highest posterior density region, respectively

more often issue coupons or are featured tend to have stronger price effects. Relative price also has an effect on the impact of coupons. Coupons of higher priced brands are less effective. Moreover, higher market concentration tends to lead to stronger price effects and higher coupon effectiveness. This corresponds to what we hypothesized above.

For the long-run parameters, we only find strong results for price effects. The signs of these effects are similar to those for the short-run. A high relative price or a high coupon intensity is correlated with a strong price effect. For the long-run effects we do not find

substantive interactions with market concentration. This final result corresponds with the findings in Nijs *et al.* (2001).

In the final row of Table 4.1, we present the posterior probability that the absolute long-run effect of a marketing instrument exceeds the absolute value of the short-run effect. For price there is only a 28.3% probability that the long-run effect will exceed the short-run effect. For coupon this probability is 54.4% and for display/feature promotion it is 87.2%. These probabilities of course correspond well to the bottom row of graphs in Figure 4.1. Hence, price changes mainly impact the market shares in the short run, while promotions seem to have more of a longer-run impact.

Finally, when we compare our results with those obtained from the commonly used two-step procedure in which first individual market-level models are estimated and where the resulting parameters are then regressed on market and brand characteristics, we find that the signs of the estimated parameters in the second step are the same for both methods. However, the significance levels of the estimates differ substantially. As the uncertainty in market-level parameters in the two-step procedure is underestimated, we find more significant second-stage parameter estimates for the two-step method. Details can be obtained from the authors.

4.4 Conclusion

In this chapter we have put forward a new and useful model for describing market shares. The first novelty was that we considered those attraction models which entail easy to estimate long-run and short-run effects of marketing-mix instruments. As a consequence, and that is the second novelty, we could explicitly link the long-run and short-run effects with category and brand characteristics in a second level. Our resultant error-correction Hierarchical Bayes attraction model was applied to fifty brands covering seven product categories. The main results were that prices exercise mainly a short-run impact, while feature promotions have a larger long-run effect. Furthermore, which is line with the results in Nijs *et al.* (2001) who focused on category sales, we found that a more intensive use of marketing instruments, and also a higher level of market concentration, leads to stronger price effects, both in the long run and in the short run.

The model in this chapter is essentially a rather natural, and statistically proper, framework to establish generalizing statements about dynamic effects of marketing instruments on market shares. The model resembles a logit structure, and hence one possible extension of our model could be in describing the choice between brands by households. Next, the model can also be used to analyze marketing-mix effectiveness of new to introduce brands, and also to determine optimal price levels. Finally the model we have proposed can provide the basis of a study of (optimal) competitive effects. The long-run

and short-run effects we have derived in this chapter are based on the assumption of no competitive reaction. In practice this will of course not be the case. Our model could be used to perform a scenario analysis of different competitive reactions.

4.A Bayes estimation

Define $Y_{it}(c) = \ln M_{it}(c) - \ln M_{It}(c)$ and $X_{it}(c) = (\ln x_{i1t}(c), \ldots, \ln x_{iKt}(c))'$. Equation (4.17) can now be written as

$$\Delta Y_{it}(c) = \tilde{\mu}_{ic} + \Delta X_{it}(c)'\alpha_{ic} - \Delta X_{I_c,t}(c)'\alpha_{I_c,c} + \delta_c(Y_{i,t-1}(c) - X_{it}(c)'\gamma_{ic} + X_{I_c,t}(c)'\gamma_{I_c,c}) + \tilde{\varepsilon}_{it}(c), \quad (4.21)$$

for $i = 1, \ldots, I_c - 1$, where $\delta_c = 1 - \rho_c$, $\alpha_{ic} = (\alpha_{i1c}, \ldots, \alpha_{iKc})'$ and $\gamma_{ic} = (\gamma_{i1c}, \ldots, \gamma_{iKc})'$ for $i = 1, \ldots, I_c$. Furthermore, define $\tilde{\mu}_c = (\tilde{\mu}'_{1c}, \ldots, \tilde{\mu}'_{I_c-1,c})'$, $\alpha_c = (\alpha'_{1c}, \ldots, \alpha'_{I_c,c})'$, $\gamma_c = (\gamma_{1c}, \ldots, \gamma_{I_c,c})'$ and the error terms

$$\tilde{e}_{it}(c)(\tilde{\mu}_{ic}, \delta_c, \alpha_c, \gamma_c) = \Delta Y_{it}(c) - \tilde{\mu}_{ic} - (\Delta X_{it}(c)'\alpha_{ic} - \Delta X_{I_c,t}(c)'\alpha_{I_c,c}) - \delta_c(Y_{i,t-1}(c) - X_{it}(c)'\gamma_{ic} + X_{I_c,t}(c)'\gamma_{I_c,c}). \quad (4.22)$$

The vector of error terms is given by

$$\tilde{e}_t(c)(\tilde{\mu}_c, \delta_c, \alpha_c, \gamma_c) = (\tilde{e}_{1t}(c)(\tilde{\mu}_{1c}, \delta_c, \alpha_c, \gamma_c), \ldots, \tilde{e}_{I_c-1,t}(c)(\tilde{\mu}_{I_c,c}, \delta_c, \alpha_c, \gamma_c))', \quad (4.23)$$

and hence the likelihood of the model reads as

$$\prod_{c=1}^{C} \int_{\alpha_c,\gamma_c} \prod_{t=2}^{T_c} \phi\Big(\tilde{e}_t(c)(\tilde{\mu}_c, \delta_c, \alpha_c, \gamma_c); 0, \tilde{\Sigma}_c\Big) \prod_{i=1}^{I_c} \phi(\alpha_{ic}; \lambda_1' z_{ic}, \Sigma_\eta)\phi(\gamma_{ic}; \lambda_2' z_{ic}, \Sigma_\nu) d\alpha_c d\gamma_c, \quad (4.24)$$

where $\phi(x; \mu, \Sigma)$ is the density function of the multivariate normal distribution with mean μ and variance Σ evaluated at x.

To obtain posterior results, we use the Gibbs sampling technique of Geman and Geman (1984) with data augmentation, see Tanner and Wong (1987). An introduction into the Gibbs sampler can be found in Casella and George (1992), see also Smith and Roberts (1993) and Tierney (1994) Hence, the latent variables α_c and γ_c are sampled alongside the model parameters $\tilde{\mu}_c$, δ_c, $\tilde{\Sigma}_c$, λ_1, λ_2, Σ_η and Σ_ν. The Bayesian analysis is based on uninformative priors for the model parameters. To improve convergence of the MCMC sampler we impose inverted Wishart priors on the Σ_η and Σ_ν parameter with scale parameter $\kappa_1 \mathbf{I}_K$ and degrees of freedom κ_2. We set the value of κ_1 to $\frac{1}{1000}$ and κ_2 equal to 1 such that the influence of the prior on the posterior distribution is marginal, see Hobert and Casella (1996) for a discussion.

In the remainder of this appendix we derive the full conditional posterior distributions of the model parameters and α_c and γ_c. In deriving the sampling distributions we build on the results in Zellner (1971, Chapter VIII).

Sampling of $\tilde{\mu}_c$ and δ_c

To sample $\tilde{\mu}_c$ and δ_c we rewrite the model in (4.21) as

$$\Delta Y_{it}(c) - \Delta X_{it}(c)'\alpha_{ic} + \Delta X_{I_c,t}(c)'\alpha_{I_c,c} = \tilde{\mu}_{ic} + \delta_c(Y_{i,t-1}(c) - X_{it}(c)'\gamma_{ic} + X_{I_c,t}(c)'\gamma_{I_c,c}) + \tilde{\varepsilon}_{it}(c) \quad (4.25)$$

for $i = 1, \ldots, I_c - 1$. Stacking the equations in (4.25) we obtain the multivariate regression model

$$W_t(c) = V_t(c)\beta + \tilde{\varepsilon}_t(c), \quad (4.26)$$

where $W_t(c)$ is a $I_c - 1$ dimensional vector containing the left-hand side of equation (4.25), $V_t(c)$ an $(I_c - 1)$ dimensional identity matrix extended with a $(I_c - 1)$ dimensional column vector of error-correction terms $(Y_{i,t-1}(c) - X_{it}(c)'\gamma_{ic} + X_{I_c,t}(c)'\gamma_{I_c,c})$, and where $\beta = (\tilde{\mu}_{1c}, \ldots, \tilde{\mu}_{I_c-1,c}, \delta_c)'$. The error term is normal distributed with mean 0 and variance $\tilde{\Sigma}$. Hence, the full conditional posterior distribution of β is a matrix normal distribution with mean

$$\left(\sum_{t=2}^{T_c} V_t(c)'\tilde{\Sigma}^{-1}V_t(c)\right)^{-1}\left(\sum_{t=2}^{T_c} V_t(c)'\tilde{\Sigma}^{-1}W_t(c)\right), \quad (4.27)$$

and variance

$$\left(\sum_{t=2}^{T_c} V_t(c)'\tilde{\Sigma}^{-1}V_t(c)\right)^{-1}. \quad (4.28)$$

Sampling of $\tilde{\Sigma}_c$

To sample $\tilde{\Sigma}_c$ we again consider the multivariate regression model (4.26). The full conditional posterior distribution of $\tilde{\Sigma}_c$ is an inverted Wishart distribution with scale parameter $\sum_{t=2}^{T_c}(W_t(c) - V_t(c)\beta)(W_t(c) - V_t(c)\beta))'$ and degrees of freedom $T_c - 1$.

Sampling of λ_1 and λ_2

To sample λ_1, we note that we can write (4.18) as

$$\alpha_{ic}' = z_{ic}'\lambda_1 + \eta_{ic}'. \quad (4.29)$$

and hence it is a multivariate regression model with regression matrix λ_1. Hence, the full conditional posterior distribution of λ_1 is a matrix normal distribution with mean

$$\left(\sum_{c=1}^{C}\sum_{i=1}^{I_c} z_{ic}z_{ic}'\right)^{-1}\left(\sum_{c=1}^{C}\sum_{i=1}^{I_c} z_{ic}\alpha_{ic}\right), \quad (4.30)$$

and covariance matrix

$$\left(\Sigma_\eta \otimes \left(\sum_{c=1}^{C}\sum_{i=1}^{I_c} z_{ic}z_{ic}'\right)^{-1}\right). \tag{4.31}$$

The derivation of the sampling distribution of λ_2 proceeds in the same manner. The full conditional posterior distribution of λ_2 is a matrix normal distribution with mean

$$\left(\sum_{c=1}^{C}\sum_{i=1}^{I_c} z_{ic}z_{ic}'\right)^{-1}\left(\sum_{c=1}^{C}\sum_{i=1}^{I_c} z_{ic}\gamma_{ic}\right), \tag{4.32}$$

and covariance matrix

$$\left(\Sigma_\nu \otimes \left(\sum_{c=1}^{C}\sum_{i=1}^{I_c} z_{ic}z_{ic}'\right)^{-1}\right). \tag{4.33}$$

Sampling of Σ_η and Σ_ν

To sample Σ_η we note that (4.18) is a multivariate regression model. Hence the full conditional posterior distribution of Σ_η is an inverted Wishart distribution with scale parameter $\kappa_1\mathbf{I}_K+\sum_{c=1}^{C}\sum_{i=1}^{I_c}(\alpha_{ic}-\lambda_1'z_{ic})(\alpha_{ic}-\lambda_1'z_{ic})'$ and degrees of freedom $\kappa_2+\sum_{c=1}^{C} I_c$. The κ terms results from the inverted Wishart prior on Σ_η which is used to improve convergence of our Gibbs sampler, see Hobert and Casella (1996) for a discussion.

The sampling of Σ_ν can be done in exactly the same manner. The parameter Σ_ν is sampled from an inverted Wishart distribution with scale parameter $\kappa_1\mathbf{I}_K+\sum_{c=1}^{C}\sum_{i=1}^{I_c}(\gamma_{ic}-\lambda_2'z_{ic})(\gamma_{ic}-\lambda_2'z_{ic})'$ and degrees of freedom $\kappa_2+\sum_{c=1}^{C} I_c$.

Sampling of α_c

To sample $\alpha_c=(\alpha_{1c},\ldots,\alpha_{I_c,c})'$ we rewrite (4.17) as

$$\begin{aligned}\Delta Y_{it}(c)-\tilde{\mu}_i-\delta_c(Y_{i,t-1}(c)-X_{it}(c)'\gamma_{ic}+X_{I_c,t}(c)'\gamma_{I_c,c})&\\ =\Delta X_{it}(c)'\alpha_{ic}-\Delta X_{I_c,t}(c)'\alpha_{I_c,c}+\tilde{\varepsilon}_{it}(c),&\end{aligned} \tag{4.34}$$

for $i=1,\ldots,I_c-1$ which can be written in matrix notation

$$W_t(c)=V_t(c)\alpha_c+\tilde{\varepsilon}_t(c), \tag{4.35}$$

where $W_t(c)$ is a (I_c-1) dimensional vector containing $\Delta Y_{it}(c)-\tilde{\mu}_i-\delta_c(Y_{i,t-1}(c)-X_{it}(c)'\gamma_{ic}+X_{I_c,t}(c)'\gamma_{I_c,c})$ and

$$V_t(c)=\begin{pmatrix}\Delta X_{1t}(c)' & 0 & \ldots & 0 & -\Delta X_{I_c,t}(c)'\\ 0 & \Delta X_{2t}(c)' & \ldots & 0 & -\Delta X_{I_c,t}(c)'\\ \vdots & \vdots & \ddots & \vdots & -\Delta X_{I_c,t}(c)'\\ 0 & \ldots & 0 & \Delta X_{I_c-1,t}(c)' & -\Delta X_{I_c,t}(c)'\end{pmatrix}. \tag{4.36}$$

Furthermore, we write the I_c equations of (4.18) as

$$-U_c = -\mathbf{I}_{KI_c}\alpha_c + \eta_c, \tag{4.37}$$

where U_c is a (KI_c) dimensional vector containing the terms $\lambda_1' z_{i,c}$ and where $\mathbf{I}_{KI_c}$ is a (KI_c) dimensional identity matrix. The error term η_c is normal distributed with mean 0 and covariance matrix $(I_c \otimes \Sigma_\eta)$. To sample α_c, we combine (4.35) and (4.37)

$$\begin{aligned} \tilde{\Sigma}^{-1/2} W_t(c) &= \tilde{\Sigma}^{-1/2} V_t(c)\alpha_c + \tilde{\Sigma}^{-1/2}\tilde{\varepsilon}_t(c), \\ -(\mathbf{I}_c \otimes \Sigma_\eta^{-1/2})U_c &= -(\mathbf{I}_c \otimes \Sigma_\eta^{-1/2})\alpha_c + (\mathbf{I}_c \otimes \Sigma_\eta^{-1/2})\eta_c. \end{aligned} \tag{4.38}$$

Hence, the full conditional posterior distribution of α_c is normal with mean

$$\left((\mathbf{I}_c \otimes \Sigma_\eta^{-1}) + \sum_{t=2}^{T_c}(V_t(c)'\tilde{\Sigma}^{-1}V_t(c))\right)^{-1}\left((\mathbf{I}_c \otimes \Sigma_\eta^{-1})U_c + \sum_{t=2}^{T_c}(V_t(c)'\tilde{\Sigma}^{-1}W_t(c))\right), \tag{4.39}$$

and covariance matrix

$$\left((\mathbf{I}_c \otimes \Sigma_\eta^{-1}) + \sum_{t=2}^{T_c}(V_t(c)'\tilde{\Sigma}^{-1}V_t(c))\right)^{-1}. \tag{4.40}$$

Sampling of γ_c

To sample $\gamma_c = (\gamma_{1c}, \ldots, \gamma_{I_c,c})'$, we rewrite (4.17) as

$$\begin{aligned} \Delta Y_{it}(c) - \tilde{\mu}_i - \Delta X_{it}(c)'\alpha_{ic} + \Delta X_{I_c,t}(c)'\alpha_{I_c,c} - \delta_c Y_{i,t-1}(c) = \\ - \delta_c X_{it}(c)'\gamma_{ic} + \delta_c X_{I_c,t}(c)'\gamma_{I_c,c} + \tilde{\varepsilon}_{it}(c), \end{aligned} \tag{4.41}$$

for $i = 1, \ldots, I_c - 1$ which can be written in matrix notation

$$\tilde{\Sigma}^{-1/2} W_t(c) = \tilde{\Sigma}^{-1/2} V_t(c)\gamma_c + \tilde{\Sigma}^{-1/2}\tilde{\varepsilon}_t(c), \tag{4.42}$$

where now $W_t(c)$ is a $(I_c - 1)$ dimensional vector containing $\Delta Y_{it}(c) - \tilde{\mu}_i - \Delta X_{it}(c)'\alpha_{ic} + \Delta X_{I_c,t}(c)'\alpha_{I_c,c} - \delta_c Y_{i,t-1}(c)$ and

$$V_t(c) = \begin{pmatrix} -\delta_c X_{1,t-1}(c)' & 0 & \ldots & 0 & \delta_c X_{I_c,t}(c)' \\ 0 & -\delta_c X_{2t}(c)' & \ldots & 0 & \delta_c X_{I_c,t}(c)' \\ \vdots & \vdots & \ddots & \vdots & \delta_c X_{I_c,t}(c)' \\ 0 & \ldots & 0 & -\delta_c X_{I_c-1,t}(c)' & \delta_c X_{I_c,t}(c)' \end{pmatrix}. \tag{4.43}$$

Again, we write the I_c equations of (4.19) as

$$-(\mathbf{I}_c \otimes \Sigma_\nu^{-1/2})U_c = -(\mathbf{I}_c \otimes \Sigma_\nu^{-1/2})\gamma_c + (\mathbf{I}_c \otimes \Sigma_\nu^{-1/2})\omega_t, \tag{4.44}$$

where U_c is a (KI_c) dimensional vector containing the terms $\lambda_2' z_{ic}$. The distribution of the error term ω_t is normal with mean 0 and covariance matrix $(I_c \otimes \Sigma_\nu)$. If we combine (4.42) with (4.44) it is easy to see that the full conditional posterior distribution of γ_c is normal with mean

$$\left((\mathbf{I}_c \otimes \Sigma_\nu^{-1}) + \sum_{t=2}^{T_c}(V_t(c)'\tilde{\Sigma}^{-1}V_t(c))\right)^{-1}\left((\mathbf{I}_c \otimes \Sigma_\nu^{-1})U_c + \sum_{t=2}^{T_c}(V_t(c)'\tilde{\Sigma}^{-1}W_t(c))\right), \quad (4.45)$$

and covariance matrix

$$\left((\mathbf{I}_c \otimes \Sigma_\nu^{-1}) + \sum_{t=2}^{T_c}(V_t(c)'\tilde{\Sigma}^{-1}V_t(c))\right)^{-1}. \quad (4.46)$$

Part II

Household-level models

Chapter 5

Modeling category-level purchase timing with brand-level marketing variables

5.1 Introduction

In this chapter we focus on the modeling of purchase timing of households in frequently purchased product categories. To describe purchase timing several models are proposed in the literature, see for example Franses and Paap (2001, Chapter 8) and Seetharaman and Chintagunta (2003) for recent overviews. If one considers time to be discrete, one often uses a purchase incidence model. In practice this means that purchase incidences are recorded in weekly intervals. In this case it is assumed that the interpurchase times have a negative binomial distribution, see, among many others, Bucklin and Lattin (1991); Ailawadi and Neslin (1998) and Bell *et al.* (1999). In many cases, researchers consider time to be continuous. In theory purchases can then be made at every point in time. In this case, proportional hazard or accelerated lifetime models are used to described purchase timing, see, among many others, Gupta (1988); Jain and Vilcassim (1991); Vilcassim and Jain (1991) and Helsen and Schmittlein (1993). The type of model that is used usually depends on the frequency of the data available to the researcher.

When explaining purchase timing in an econometric model, one aims at describing the relation between interpurchase times and various explanatory variables. These explanatory variables can be divided into two groups. The first group corresponds to household-specific variables, like household size and family income, but also variables as the current stock of the product and the time since last purchase within the product category. These variables can be directly linked to the interpurchase times. The second group contains marketing-mix variables, like price and the presence of promotional activities. These vari-

ables cannot be directly linked to the interpurchase times, as marketing-mix variables are observed at the brand level and purchase timing is modeled at the category level.

In the ideal case, we would have knowledge of the preferred brand of each household at every moment in time. To explain purchase timing we could then use the marketing mix of the brand that is bought or would be bought at any moment in time. In practice this is of course infeasible. First of all the data collection would be practically impossible. Second, the household may not have a unique preferred brand at every point in time. It is therefore up to the researcher to somehow summarize the marketing efforts of all brands into category-level indices. This task is exactly the research question we study in this chapter. The key question of this chapter can be summarized as

— What to do with the marketing mix when modeling purchase timing? —

or in other words: how to construct category-level measures of marketing efforts from the marketing mix of individual brands that can be included in a category-level interpurchase time model?

One may think that the answer to this question is to use the marketing mix of the purchased brand. There are however two major problems with this approach. First of all, in the decision process of a household, the purchase timing decision precedes the brand choice decision. To the researcher, knowledge of the purchased brand therefore includes the information that a purchase is made in the category. Technically speaking, one therefore cannot use information on the purchased brand to construct explanatory variables for a purchase timing model. A second problem is that brand choice is not available at non-purchase moments. One may opt to use the marketing mix of the previously purchased brand, but this is likely to be sub-optimal as households may switch brands. In fact, a household may change preferences several times in between two purchases, especially if the marketing mix changes in this period.

We are of course not the first to notice these problems in modeling interpurchase time. In every purchase timing study the researcher will have to decide upon how to construct category-level marketing-mix variables. An often-used solution is to use a weighted average of brand-specific marketing-mix variables. The weights are usually household specific and obtained from choice shares of the particular household, see for example Gupta (1988, 1991). A disadvantage of weighting the marketing mix using choice shares is that household-specific information is required to obtain the weights. This approach is therefore less suitable for out-of-sample forecasting. Finally, as choice shares are by definition constant over (periods) of time the model does not take into account that preferences may change over time.

Another popular approach amounts to using the so-called inclusive value from a brand choice model as a summary statistic for the marketing efforts in a category, see, among others, Bucklin and Gupta (1992); Chintagunta and Prasad (1998) and Bell *et al.* (1999).

The inclusive value has the interpretation of the expected maximum utility over all brands in the category. The inclusive value naturally depends on the marketing mix of all brands. A large expected utility is expected to be positively correlated with the probability of a purchase in the category. Although theoretically appealing, this specification is rather restrictive. In the corresponding purchase timing model there is only one parameter that relates all marketing efforts of all brands to the purchase timing, that is, the coefficient corresponding to the inclusive value. Moreover the effects of marketing variables are restricted to be more or less the same on choice as on purchase timing. Another problem may be that the relation between the inclusive value and purchase incidence may only hold within households. Between households there may be substantial differences in inclusive value that are not related to differences in purchase timing. A household with a strong brand preference may have a larger inclusive value than a household with less pronounced preferences. Of course, one cannot conclude from this that the first household will on average have shorter interpurchase times. The between-household differences will be even more pronounced when the brand choice model allows for unobserved heterogeneity in brand preferences.

To meet the limitations of the above-mentioned approaches, we introduce in this chapter some alternative specifications. The idea behind these specifications is to use brand choice probabilities as indicators of brand preferences. One method to summarize the marketing efforts of all brands to the category level is again to use a weighted average of the marketing mix of each brand, but now using the current preferences of the household as weights. This approach is very similar to using choice shares as weights as in Gupta (1988, 1991). However, in our case the preference weights may change over time as they are captured by a brand choice model. Another method is to specify brand-specific purchase incidence probabilities, or, in a continuous model, brand-specific hazard functions. The category purchase probability is then obtained as the weighted average of these probabilities using preference probabilities.

Although these solutions meet the limitations of the standard approaches in the literature, they still consider brand choice and purchase timing as separate issues. However, the fact that a household does not make a purchase in a particular week, reveals information about the preferences of this household. For example, consider the situation where a household frequently purchases a certain brand that is also frequently promoted. Assume that this household never purchases other brands when they are promoted. If one only considers purchase occasions one may overestimate the effect of promotions on brand choice as the non-purchase promotional activities are completely ignored. The fact that the household does not purchase the other brands while they are promoted implies that it has a strong base preference for the frequently purchased brand. It would therefore be better to integrate the interpurchase time model with a brand choice model. In this model the brand choices of households are revealed at purchase occasions, while at non-purchase

occasions the preferred brand is treated as a latent (unobserved) variable. In this way, we also use information revealed by households at non-purchase occasions to model brand choices and interpurchase timing. We will call this specification the latent preferences purchase timing model.

To answer the question concerning the inclusion of the marketing mix in a purchase timing model, we consider the two standard approaches (based on choice shares and based on the inclusive value) and compare them with our two alternative approaches (based on brand choice probabilities) and the latent preference model. In Section 5.2 we discuss the statistical differences of the various model specifications and we discuss parameter estimation. The analysis is done for discrete time and continuous time interpurchase time models. In Section 5.3 we compare different specifications using data on purchases in three product categories. The comparison is based on in-sample fit and out-of-sample forecasting performance. Finally, in Section 5.4 we conclude. In this section we also discuss the practical implications for modeling purchase timing.

5.2 Modeling interpurchase timing

In this section we discuss the inclusion of marketing-mix variables in a category-level interpurchase time model. We consider the solutions put forward in the literature together with our alternative solutions. Sections 5.2.1 and 5.2.2 deal with discrete purchase timing models, while in Section 5.2.3 we extend the models to continuous time.

5.2.1 The discrete case

Let $\Pr[D_{it} = 1]$ denote the probability that a purchase is made by household i in week t, in a specific category. Furthermore, denote by d_{in} the (calendar) time of the n-th purchase of household i, where $i = 1, \ldots, I$ and $n = 1, \ldots, N_i$. Given a purchase at week $d_{i,n-1}$ the probability that household i's next purchase will be in week d_{in} is therefore

$$\Pr[D_{i,d_{in}} = 1] \times \prod_{t=d_{i,n-1}+1}^{d_{in}-1} \big(1 - \Pr[D_{it} = 1]\big). \tag{5.1}$$

This corresponds to a Negative Binomial distribution for interpurchase times. To relate the purchase incidence probability $\Pr[D_{it} = 1]$ to observable characteristics, denoted by w_{it}, one can use the standard logistic function, that is $\Pr[D_{it} = 1] = G(w_{it})$, where

$$G(w_{it}) = \frac{\exp(\gamma_0 + w_{it}'\gamma_1)}{1 + \exp(\gamma_0 + w_{it}'\gamma_1)}. \tag{5.2}$$

It is obvious how household-specific variables can be included in this logistic function. However, if one wants to include marketing instruments in the model, it is unclear which

brand's marketing-mix variables or which combination of brand-specific variables should be included in w_{it} as brand choice is only revealed at purchase occasions and not in between. Below we present several possibilities to solve this problem. For simplicity of notation we will assume that the model only includes marketing instruments. Other types of explanatory variables can be included in the usual way.

Choice share weighted average of marketing mix

One may include a weighted average of the marketing mix over the J brands. As weights one can use observed household-specific choice shares as in Gupta (1988). Hence, in this case the incidence probability is given by $\Pr[D_{it} = 1] = G(\sum_{j=1}^{J} c_{ij} x_{ijt})$, where c_{ij} denotes observed choice share of brand j for household i and x_{ijt} denotes the marketing mix of brand j as experienced by household i at time t.

The household-specific choice shares are usually estimated using the in-sample purchases. Out-of-sample forecasts will also have to be based on the in-sample choice shares. This approach is therefore not useful in case one wants to predict purchase timing of households for which no purchase history is available. If this type of forecasting is one of the aims of the analysis, one has to rely on one of the other solutions discussed below.

Inclusive value

Another frequently used approach is to include the inclusive value from a brand choice model as an explanatory variable in the purchase incidence model. To describe brand choice we consider a multinomial logit model

$$\Pr[Y_{it} = j] = \frac{\exp(\alpha_j + x'_{ijt}\beta)}{\sum_{s=1}^{J} \exp(\alpha_s + x'_{ist}\beta)}, \tag{5.3}$$

where Y_{it} denotes the brand choice of household i at time t, $\alpha_J = 0$ for identification, and where β measures the effect of the marketing mix on brand choice. For simplicity we again assume that there are no other explanatory variables besides the marketing mix. Note that the household only makes a purchase in some of the periods. Therefore, Y_{it} is only observed for $t \in \{d_{in}\}_{n=1}^{N_i}$. Although we only observe Y_{it} on purchase occasions, we assume that the choice probabilities do represent the household's preference for the brands at every point in time.

The inclusive or category value is defined by

$$I_{it} = \ln\Big(\sum_{j=1}^{J} \exp(\alpha_j + x'_{ijt}\beta)\Big). \tag{5.4}$$

This expression has the interpretation of the expected maximum utility over all brands in the category. Using the inclusive value as a summary statistic for the marketing efforts

in the category, we can define the purchase incidence probability as $\Pr[D_{it} = 1] = G(I_{it})$. In this case γ_1 in (5.2) will be only one-dimensional, that is, all marketing instruments are summarized in I_{it}.

The use of the inclusive value can also be justified as a nested logit model specification, see for example Ben-Akiva and Lerman (1985), Franses and Paap (2001) and Train (2003). In this specification, the inclusive value captures the correlation between the purchase timing and the brand choice decision. This approach is followed by for example Ailawadi and Neslin (1998) and Bell *et al.* (1999).

Preference weighted average of marketing mix

An alternative to the choice share approach is to use a weighted average of the marketing mix, where the weighting scheme follows from choice/preference probabilities $\Pr[Y_{it} = j]$ in week t. The purchase probabilities are now given by

$$\Pr[D_{it} = 1] = G\Big(\sum_{j=1}^{J} \Pr[Y_{it} = j] x_{ijt}\Big). \tag{5.5}$$

The advantage of this approach over using choice shares as weights, is that this method allows the weights to evolve over time. Changes in preferences over time, for example due to promotions, are therefore accounted for in this weighting scheme. Additionally, this approach can be used to construct out-of-sample weights for households with an unknown purchase history.

Weighted average of incidence probabilities

Instead of taking a weighted average of the marketing mix, one may also consider a weighted average of brand-specific purchase incidence probabilities. For the weighting scheme we can again use the choice probabilities resulting in

$$\Pr[D_{it} = 1] = \sum_{j=1}^{J} \Pr[Y_{it} = j] G(x_{ijt}). \tag{5.6}$$

This specification is very similar to the previous one where the weighting occurs inside the logit function $G(\cdot)$. However due to the nonlinearity of the logit function it will give different results.

Latent preference purchase timing model

A more sensible approach may be to integrate the choice model and the purchase incidence model. Hence, we jointly model brand choice and purchase incidence. The probability

that a household purchases brand j at time t is given by the probability that the household makes a purchase and prefers brand j at time t, that is,

$$\Pr[Y_{it} = j \wedge D_{it} = 1] = \Pr[Y_{it} = j] \Pr[D_{it} = 1 | Y_{it} = j], \tag{5.7}$$

where $\Pr[Y_{it} = j]$ is given in (5.3) and

$$\Pr[D_{it} = 1 | Y_{it} = j] = G(x_{ijt}). \tag{5.8}$$

Likewise, we can define the probability that a household prefers brand j at time t, but does not make a purchase in the category

$$\Pr[Y_{it} = j \wedge D_{it} = 0] = \Pr[Y_{it} = j](1 - \Pr[D_{it} = 1 | Y_{it} = j]) = \Pr[Y_{it} = j](1 - G(x_{ijt})). \tag{5.9}$$

On purchase occasions households reveal their brand preference through the brand choices. In case no purchase is made in a week, there is theoretically no brand choice variable Y_{it}. However, we assume that Y_{it} does represent the preferred brand of the household in week t, irrespective of a purchase being made. In non-purchase weeks, this variable is then of course latent. To determine the probability that household i does not purchase any brand in week t, we have to sum the non-purchase probabilities with respect to the preferred brand variable which results in

$$\begin{aligned} \Pr[D_{it} = 0] &= 1 - \sum_{j=1}^{J} \Pr[Y_{it} = j] \Pr[D_{it} = 1 | Y_{it} = j] \\ &= \sum_{j=1}^{J} (1 - \Pr[D_{it} = 1 | Y_{it} = j]) \Pr[Y_{it} = j], \end{aligned} \tag{5.10}$$

where again $\Pr[Y_{it} = j]$ is given in (5.3).

5.2.2 Estimation – The discrete case

All the models in this chapter can be estimated by Maximum Likelihood. For the general case the likelihood function reads

$$\mathcal{L} = \prod_{i=1}^{I} \prod_{n=1}^{N_i} \mathcal{L}_{in}, \tag{5.11}$$

where $\mathcal{L}_{in}$ denotes the likelihood contribution of the n-th purchase of household i. For the different specifications of the model this likelihood contribution will of course differ. For all specifications, but the one based on choice share weights, the likelihood contribution contains brand choice probabilities from a logit model. Even for the models for which the brand choice decision is defined conditional on purchase timing, we estimate the brand choice and the purchase timing model simultaneously.

Choice share weighted average of marketing mix

When choice shares are used as weights, one does not have to estimate a choice model. The likelihood contribution of the n-th purchase of household i equals

$$\mathcal{L}_{in} = G\Big(\sum_{j=1}^{J} c_{ij} x_{ij,d_{in}}\Big) \prod_{t=d_{i,n-1}+1}^{d_{in}-1} \left(1 - G\Big(\sum_{j=1}^{J} c_{ij} x_{ijt}\Big)\right). \tag{5.12}$$

Inclusive value

For the inclusive value specification, the likelihood contribution of the n-th purchase of household i equals

$$\mathcal{L}_{in} = \Pr[Y_{i,d_{in}} = y_{i,d_{in}}] \times G(I_{i,d_{in}}) \prod_{t=d_{i,n-1}+1}^{d_{in}-1} (1 - G(I_{it})), \tag{5.13}$$

where $y_{i,d_{in}}$ denotes the actual brand choice at the n-th purchase of household i. The structure of this likelihood contribution is very similar to (5.12). For the inclusive value the likelihood contribution however also includes the brand choice probability.

Preference weighted average of the marketing mix

When preference probabilities are used to aggregate the marketing mix, the contribution to the likelihood function of a purchase by household i is given by

$$\begin{aligned}\mathcal{L}_{in} = \Pr[Y_{i,d_{in}} = y_{i,d_{in}}] \times \\ G\Big(\sum_{j=1}^{J} \Pr[Y_{i,d_{in}} = j] x_{ij,d_{in}}\Big) \prod_{t=d_{i,n-1}+1}^{d_{in}-1} \left(1 - G\Big(\sum_{j=1}^{J} \Pr[Y_{it} = j] x_{ijt}\Big)\right).\end{aligned} \tag{5.14}$$

Weighted average of incidence probabilities

For the case where a weighted average of incidence probabilities is used the likelihood contribution reads

$$\begin{aligned}\mathcal{L}_{in} = \Pr[Y_{i,d_{in}} = y_{i,d_{in}}] \times \\ \sum_{j=1}^{J} \Pr[Y_{i,d_{in}} = j] G(x_{ij,d_{in}}) \prod_{t=d_{i,n-1}+1}^{d_{in}-1} \Big(1 - \sum_{j=1}^{J} \Pr[Y_{it} = j] G(x_{ijt})\Big).\end{aligned} \tag{5.15}$$

Latent preference purchase timing model

For the latent preference model the structure of the likelihood contribution is a bit more involved as the brand choice and the purchase incidence decisions are not considered separately. The contribution to the likelihood function of the n-th purchase of household i reflects the probability of no purchase in the weeks $d_{i,n-1}+1$ to $d_{in}-1$ and the purchase of brand $y_{i,d_{in}}$ in week d_{in}, in terms of (conditional) probabilities this contribution becomes

$$\mathcal{L}_{in} = \Pr[Y_{i,d_{in}} = y_{i,d_{in}}] \Pr[D_{i,d_{in}} = 1 | Y_{i,d_{in}} = y_{i,d_{in}}] \prod_{t=d_{i,n-1}+1}^{d_{in}-1} \left(\sum_{j=1}^{J} (1 - \Pr[D_{it} = 1 | Y_{it} = j]) \Pr[Y_{it} = j] \right), \quad (5.16)$$

or in terms of the logit function

$$\mathcal{L}_{in} = \Pr[Y_{i,d_{in}} = y_{i,d_{in}}] G(x_{ij,d_{in}}) \prod_{t=d_{i,n-1}+1}^{d_{in}-1} \left(\sum_{j=1}^{J} (1 - G(x_{ijt})) \Pr[Y_{it} = j] \right). \quad (5.17)$$

One can interpret this likelihood expression as follows. In the weeks where household i does not purchase the product we do not observe its preferred brand. The preferred brand of household i in these weeks is therefore a latent variable. To take care of this latent variable we sum over all possible realizations where we use the brand choice probabilities as weights. In periods where a purchase is made the brand preference is observed through the brand actually chosen.

When comparing (5.15) to (5.17) the only difference between the latent preference model and the specification where a weighted average of incidence probabilities is used is in the likelihood contribution of the week in which the purchase is made. Concerning the no-purchase weeks the two likelihood functions are identical. In the latent preference model the incidence probability takes into account the brand preference. In the other specification this is not the case as there the brand choice decision is modeled conditional on purchase incidence. In the next section we will see that the differences between the two approaches is more pronounced if one decides to model purchase timing in continuous time.

5.2.3 The continuous case

In the prev9ious section we discussed purchase timing models in discrete time. In this section we will extend these ideas to continuous time models. One of the most popular continuous time models for durations is the hazard model. The difficulty of using explanatory variables that are measured at the brand level to explain interpurchase timing at the category level also holds for models in continuous time. In fact it will turn out that

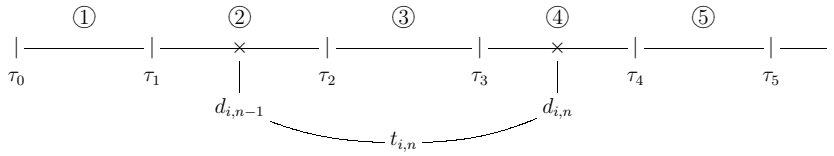

Figure 5.1: Graphical representation of purchase occasion $d_{i,n}$, interpurchase time $t_{i,n}$ and time indexes of changes in covariates τ_l

in this case there are some additional technicalities that make the analysis more involved. These difficulties especially concern the latent preference model.

Again denote by d_{in} the purchase timing of the n-th purchase of household i in calendar time, $n = 0, \ldots, N_i$. The N_i observed interpurchase times are therefore defined as $t_{in} = d_{in} - d_{i,n-1}$, with $n = 1, \ldots, N_i$. Note that for the continuous time models t refers to the time in a particular duration. For the models in discrete time we have used t as a time index in calendar time. For the continuous time models t is set to 0 at the beginning of each duration. We again assume that the marketing-mix variables are constant within one week, which leads to the natural assumption that the brand choice preferences of households are constant within a week. Denote by τ_l, $l = 1, \ldots, L$, the time indices of a change in the covariates. Using this notation, week 1 corresponds to the interval $[\tau_0, \tau_1]$. Furthermore, denote by $K_{in}(t)$ the week number corresponding to t time periods after the start of the n-th interpurchase spell. Note that after a purchase a new "week" will start. In Figure 5.1 we give a graphical representation of the purchase process. In this example we have purchases in weeks 2 and 4, and in this case we would have $K_{in}(0) = 2$ and $K_{in}(t_{in}) = 4$.

The hazard function for the n-th interpurchase time for household i is denoted by $\lambda_{in}(t)$, where $t = 0$ corresponds with the start of the interpurchase spell. As the basic building block of the model we now use a general hazard function $g(t; w_{in}(t))$ which plays the same role as the logistic function $G(w_{it})$ in the discrete case, $w_{in}(t)$ denotes the explanatory variables at duration t associated with the n-th interpurchase time. The specification of $g(\cdot)$ depends on the type of hazard model chosen, for example the proportional hazard may look like

$$g(t; w_{in}(t)) = \exp(w_{in}(t)'\gamma)\lambda_0(t), \tag{5.18}$$

where $\lambda_0(t)$ is a baseline hazard function, see Gupta (1991) for a similar approach. In this specification the sign of γ gives the direction of the effect of an increase in $w_{in}(t)$ on the hazard. That is, if $\gamma > 0$ an increase in w_{in} results in a decrease of the expected interpurchase time.

For brand choice/brand preference we use the same notation as before. Let $y_{in}(t) = y_{i,K_{in}(t)}$ denote the brand preference of household i at duration t associated with the n-th

interpurchase time. Note that we impose that the brand preferences are constant during weeks. Although one could consider smaller time intervals to allow for more frequent changes in preference, the (discrete) preference process, by definition, cannot develop in continuous time.

Again it is not clear which combination of brand-specific marketing-mix variables should be included in the hazard specification. All specifications discussed for the discrete time case have their continuous time equivalents.

Choice share weighted average of marketing mix

One may weigh the marketing mix over the J brands using observed market shares as in Gupta (1991). Hence, we have

$$\lambda_{in}(t) = g(t; \sum_{j=1}^{J} c_{ij} x_{inj}(t)), \tag{5.19}$$

where c_{ij} denotes observed choice share of brand j for household i, and where $x_{inj}(t)$ denotes the marketing mix of brand j at duration t of the n-th interpurchase spell of household i. This approach has of course the same drawbacks as for the discrete case.

The likelihood contribution of the n-th interpurchase time follows from standard duration theory, see for example Kiefer (1988), and is given by

$$\mathcal{L}_{in} = \lambda_{in}(t_{in}) S_{in}(t_{in}) = g(t_{in}; \sum_{j=1}^{J} c_{ij} x_{inj}(t_{in})) \exp(- \int_0^{t_{in}} g(s; \sum_{j=1}^{J} c_{ij} x_{inj}(s)) ds), \tag{5.20}$$

where $S_{in}(t) = \exp(- \int_0^t \lambda_{in}(s) ds)$ denotes the survivor function.

Inclusive value

Again one may use the inclusive value as explanatory variable in the hazard function as in Chintagunta and Prasad (1998). In continuous time, the inclusive value equals

$$I_{in}(t) = \ln \Big(\sum_{j=1}^{J} \exp(\alpha_j + x_{inj}(t)' \beta) \Big). \tag{5.21}$$

Note that the inclusive value will also be constant over periods of time. The hazard function is in this case given by $\lambda_{in}(t) = g(t; I_{in}(t))$. The likelihood contribution of the n-th interpurchase spell of length t_{in} resulting in a purchase of brand $y_{in}(t_{in})$ for this specification reads

$$\begin{aligned} \mathcal{L}_{in} &= \Pr[Y_{in}(t_{in}) = y_{in}(t_{in})] \times \lambda_{in}(t_{in}) S_{in}(t_{in}) \\ &= \Pr[Y_{in}(t_{in}) = y_{in}(t_{in})] \times g(t_{in}; I_{in}(t_{in})) \exp(- \int_0^{t_{in}} g(s; I_{in}(s)) ds). \end{aligned} \tag{5.22}$$

The brand choice probability enters the likelihood contribution as the inclusive value is obtained from a brand choice model.

Preference weighted average of the marketing mix/Weighted average of hazards

One may weigh the marketing mix with choice probabilities in the corresponding week resulting in

$$\lambda_{in}(t) = g\Big(t; \sum_{j=1}^{J} \Pr[Y_{in}(t) = j] x_{inj}(t)\Big), \tag{5.23}$$

or one may decide to weigh brand-specific hazards instead of the marketing mix, that is,

$$\lambda_{in}(t) = \sum_{j=1}^{J} \Pr[Y_{in}(t) = j] g(t; x_{inj}(t)). \tag{5.24}$$

For both specifications, the likelihood reads

$$\mathcal{L}_{in} = \Pr[Y_{in}(t_{in}) = y_{in}(t_{in})] \times \lambda_{in}(t_{in}) \exp(-\int_0^{t_{in}} \lambda_{in}(s) ds), \tag{5.25}$$

where either (5.23) or (5.24) is used for $\lambda_{in}(t)$.

Latent preference purchase timing model

Instead of using brand choice probabilities as convenient weights, it seems more sensible to integrate the duration model and the brand choice model. We assume again that the brand choice of a household in a certain week is known if a household makes a purchase in the product category. During non-purchase weeks, we do not observe brand choice but we assume that households do have a preferred brand. The preferred brand choice is then treated as a latent variable and takes the role of the brand choice.

To explain the model, consider the hypothetical situation where we know the preferred brands of household in all weeks, including those where no purchase is made. Assume that the preferred brand in week k is given by y_{ik} and that the hazard function in this week given Y_{ik} equals $\lambda(t|Y_{ik} = y_{ik})$. The brand choice probabilities are given by the logit probabilities $\Pr[Y_{ik} = y_{ik}]$. The joint density function of a duration from $d_{i,n-1}$ to d_{in} and preferred brands y_{ik} for weeks $k = K_{in}(0), \ldots, K_{in}(t)$ is given by

$$f(t, \{y_{ik}\}_{k=K_{in}(0)}^{K_{in}(t)}) = \lambda(t|Y_{i,K_{in}(t)} = y_{i,K_{in}(t)}) \times S(t|\{y_{ik}\}_{k=K_{in}(0)}^{K_{in}(t)}) \prod_{k=K_{in}(0)}^{K_{in}(t)} \Pr[Y_{ik} = y_{ik}]. \tag{5.26}$$

In practice, the marketing-mix variables are constant during a week. We assume that the brand preferences given the marketing mix are also constant during a week. Given these assumptions, we can expand the survivor function to obtain

$$
\begin{aligned}
f(t, \{y_{ik}\}_{k=K_{in}(0)}^{K_{in}(t)}) = \\
&\lambda(t|Y_{i,K_{in}(t)} = y_{i,K_{in}(t)}) \Pr[Y_{i,K_{in}(t)} = y_{i,K_{in}(t)}] \exp\left(- \int_{\tau_{K_{in}(t)-1}-d_{i,n-1}}^{t} \lambda(v|y_{i,K_{in}(t)})dv \right) \times \\
&\Pr[Y_{i,K_{in}(0)} = y_{i,K_{in}(0)}] \exp\Big(- \int_{0}^{\tau_{K_{in}(0)}-d_{i,n-1}} \lambda(v|y_{i,K_{in}(0)})dv \Big) \times \\
&\prod_{k=K_{in}(0)+1}^{K_{in}(t)-1} \Pr[Y_{ik} = y_{ik}] \exp\left(- \int_{\tau_{k-1}-d_{i,n-1}}^{\tau_k - d_{i,n-1}} \lambda(v|y_{ik})dv \right).
\end{aligned}
\tag{5.27}
$$

The first part of (5.27) refers to the week in which the purchase is made, the middle part concerns the period of the start of the duration to the first change in the marketing mix. The third part of (5.27) deals with all other periods of constant preferences and marketing mix.

So far we have assumed that we know the preferred brands, even at weeks where there is no purchase at all. Of course, we do not observe brand preferences at weeks without purchases. Hence, we have to sum over all possible realizations of the latent brand preferences in these weeks to obtain the joint density of the interpurchase time and the brand choice at the purchase occasion. Hence, we sum (5.27) over all possible values of y_{ik} in weeks $k = K_{in}(0), \ldots, K_{in}(t) - 1$, that is,

$$
\begin{aligned}
f(t,y_{i,K_{in}(t)}) &= \sum_{y_{i,K_{in}(0)}=1}^{J} \cdots \sum_{y_{i,K_{in}(t)-1}=1}^{J} f(t, \{y_{ik}\}_{k=K_{in}(0)}^{K_{in}(t)}) \\
&= \lambda(t|Y_{i,K_{in}(t)} = y_{i,K_{in}(t)}) \Pr[Y_{i,K_{in}(t)} = y_{i,K_{in}(t)}] \exp\Big(- \int_{\tau_{K_{in}(t)-1}-d_{i,n-1}}^{t} \lambda(v|y_{i,K_{in}(t)})dv \Big) \times \\
&\sum_{y_{i,K_{in}(0)}=1}^{J} \Pr[Y_{i,K_{in}(0)} = y_{i,K_{in}(0)}] \exp\Big(- \int_{0}^{\tau_{K_{in}(0)}-d_{i,n-1}} \lambda(v|y_{i,K_{in}(0)})dv \Big) \times \\
&\prod_{k=K_{in}(0)+1}^{K_{in}(t)-1} \left(\sum_{j=1}^{J} \Pr[Y_{ik} = j] \exp\Big(- \int_{\tau_{k-1}-d_{i,n-1}}^{\tau_k - d_{i,n-1}} \lambda(v|y_{ik})dv \Big) \right).
\end{aligned}
\tag{5.28}
$$

The likelihood contribution $\mathcal{L}_{in}$ of the n-th interpurchase time of household i resulting in a purchase of brand $y_{i,K_{in}(t_{in})}$ now equals $f(t_{in}, y_{i,K_{in}(t_{in})})$.

5.3 Empirical comparison

In this section we compare the various model specifications for the interpurchase timing model in continuous time as given in Section 5.2.3 using household panel scanner data. For three different categories of fast-moving consumer goods, we will estimate the five different specifications discussed in Section 5.2.3. The performance of the different specifications is measured using in-sample and out-of-sample criteria.

The data we use is part of the so-called ERIM database, which is collected by A.C. Nielsen. The data span the years 1986 to 1988, and the particular subset we use concerns purchases of catsup, laundry detergent and yogurt by households in Sioux Falls (South Dakota, USA). We split the data sets in two parts such that the number of households is roughly the same in both samples. The first part is used to estimate the parameters of the various models, while the second part is used for out-of-sample evaluation. Table 5.1 provides an overview of the number of brands, households and number of purchases in the samples. The catsup category contains the brands Del Monte, Heinz and Hunts, the yogurt category contains Dannon, Nordica, W-B-B, Yoplait and a rest brand, while for the detergent category we have Cheer, Oxydol, Surf, Tide, Wisk and a rest brand.

We use a standard multinomial logit model to describe brand choice and to describe interpurchase timing we use a proportional hazard model with a log-logistic baseline hazard, to be more precise the baseline hazard reads

$$\lambda_0(t) = \frac{\alpha\gamma t^{\alpha-1}}{1+\gamma t^{\alpha}}, \tag{5.29}$$

where $\alpha > 0$ and $\gamma > 0$. This specification allows the baseline hazard to be monotonically decreasing or inverted U-shaped. Chintagunta and Haldar (1998) show that for modeling purchase timing this baseline hazard outperforms commonly used alternatives as the Weibull or Erlang-2 specification. The multinomial logit model we use to model brand choice contains brand-specific intercepts, the marketing mix of all brands in the

Table 5.1: Data characteristics of three categories of fast-moving consumer goods

Category	No. brands	No. households		No. purchases	
		In-sample	Out-of-sample	In-sample	Out-of-sample
Catsup	3	363	356	3742	3610
Detergent	6	303	295	2318	2080
Yogurt	5	210	209	4337	3605

category (price, display and feature) and a lagged brand choice dummy capturing state-dependence. As explanatory variables in the hazard model we use household size and household income as these variables are known to influence interpurchase timing. To control for inventory effects, such as stockpiling, we use the volume previously bought in the category as an additional variable, see also Chintagunta and Prasad (1998) for a similar approach. These variables are household specific and therefore they are not subject to the difficulties presented in this chapter for brand-specific variables. Finally, we use the available marketing instruments, that is, price, display and feature. The five different ways to include the marketing mix in the hazard specification discussed in Section 5.2.3 lead to five alternative models.

In this chapter we are not so much interested in specific values of estimated parameters, the focus lies on the comparison of the various model specifications. To this end the analysis is split up in two parts. First, we analyze the differences in descriptive power, where we do not allow for unobserved heterogeneity in the brand choice model. Next, we consider the case where unobserved heterogeneity is modeled using a finite mixture distribution. In the homogeneous case the model specification using individual choice shares to obtain a category average of the marketing efforts of the different brands in a category (5.19) clearly has an advantage over the other specifications. It allows for an easy representation of between-household heterogeneity in brand preferences. Differences in brand preferences will have a large influence on the relative importance of the marketing mix of individual brands on the purchase incidence decision. We expect this specification to be superior in in-sample fit. For out-of-sample prediction, individual choice shares may not be available if we consider households outside the estimation sample. One may use the in-sample average choice share across households as a predictor for the out-of-sample individual choice shares. In this case the forecasting performance of the individual choice share specification will probably be lower.

Concerning in-sample performance we expect that explicit modeling of (unobserved) heterogeneity in brand preferences will lead to the same or an even better fit for the alternative models compared to the specification based on choice shares. This assertion is analyzed in the second part of this section.

No unobserved heterogeneity

First of all we compare the performance of the different specifications without controlling for unobserved heterogeneity. Table 5.2 shows some performance statistics of six models for the three categories under investigation. As in-sample measures we consider the maximum log likelihood value, the AIC and the BIC. The value of the log likelihood function for the out-of-sample observations evaluated at the estimate based on the in-sample observations is used to evaluate the forecasting performance of the various model

specifications. For the choice share specification we consider two values of the out-of-sample log likelihood. The first is based on household-specific choice shares estimated using out-of-sample observations, while for the second the choice shares are set to the in-sample average choice shares. This measure represents the case in which choice share information is not available when forecasting interpurchase times.

Table 5.2 displays an overview of the results. Several conclusions can be drawn from this table. First, if we consider in-sample measures, the specification that uses the household-specific choice shares as weights performs best for all criteria. Note that we did not count the choice shares as parameters in computing the information criteria, although strictly speaking these estimated shares are to be seen as parameters. If we would count the weights as parameters, the choice-share specification would be the lowest in rank on the AIC and BIC measures. Secondly, the inclusive value specification turns out to perform worst on all measures. Finally, if we ignore the choice share specification, the latent preference model performs best for all performance measures.

If we consider out-of-sample measures we notice the same pattern. The only difference is that for the catsup category both the weighted hazard and the weighted marketing mix specifications outperform the latent preference model. Furthermore, if we compute the out-of-sample log likelihood value using an average of in-sample household-specific choice shares, the forecasting performance of the "choice share model" is almost always worse than of the other specifications, with the exception of the inclusive value specification for the catsup category.

Unobserved heterogeneity

We have seen that the model based on household-specific choice shares performs best on in-sample and out-of-sample measures. As already discussed before, we expect this superiority to vanish if we explicitly model heterogeneity in brand preference among households. To validate this claim, we estimate the models while allowing for unobserved heterogeneity in brand preferences and the average purchase rate. That is, we allow the brand intercepts and the intercept of the hazard function to differ across households through the use of latent segments, see Wedel and Kamakura (1999).

To reduce the probability of ending up in a local maximum of the likelihood, we estimate the heterogeneous models with ten different starting values. The results below are based on the best of these ten starting values.

To summarize the results we only compare the performance of the individual choice share model with the latent preference model in detail. Note that the latent preference model turned out to be second best in the homogeneous case. Table 5.3 provides an overview of the results for the three product categories. If we correct for unobserved heterogeneity the relative performance of the latent preference model versus the model

Table 5.2: Performance measures of different interpurchase time models without correcting for unobserved heterogeneity[1]

	choice model	Duration/Choice models: choice shares[2]	inclusive value	weighted hazard	weighted mark. mix	latent preferences
Catsup category						
$\ln \mathcal{L}$	−1909.36	-13872.4	-13892.5	-13883.6	-13881.9	-13878.2
AIC	3830.72	27775.9	27810.9	27799.2	27795.8	27788.3
BIC	3854.08	27838.3	27861.6	27861.5	27858.1	27850.6
out-of-sample $\ln \mathcal{L}$	−1541.13	-13072.1	-13110.4	-13092.3	-13093.1	-13095.0
with in-sample shares		-13094.1				
Detergent category						
$\ln \mathcal{L}$	−2335.37	-8996.27	-9012.75	-9004.35	-9004.11	-8998.17
AIC	4688.73	18030.5	18057.5	18046.7	18046.2	18034.3
out-of-sample $\ln \mathcal{L}$	−2254.69	-8294.79	-8317.58	-8312.68	-8312.20	-8311.87
with in-sample shares		-8318.39				
Yogurt category						
$\ln \mathcal{L}$	−3869.15	-13277.2	-13414.7	-13399.0	-13402.0	-13387.3
AIC	7754.30	26590.4	26859.5	26834.0	26840.0	26810.6
BIC	7784.01	26657.2	26915.2	26900.8	26906.8	26877.5
out-of-sample $\ln \mathcal{L}$	−3707.60	-12372.7	-12408.4	-12398.6	-12397.1	-12387.7
with in-sample shares		-12425.1				

[1] Underlined entries indicate the best performing model, per performance measure. For the out-of-sample likelihood the best performing model based on in-sample shares is also underlined.

[2] The interpurchase timing model using choice shares can be estimated independently from the brand choice model. To allow for easy comparison, the performance statistics however show the results of the combination of the duration model and the brand choice model.

Table 5.3: Likelihood differences Latent preference model - Choice shares model, both with unobserved heterogeneity (average in-sample shares used for out-of-sample choice shares)

	Number of segments					
	1	2	3	4	5	6
In-sample log-likelihood difference						
Catsup	−5.84	−12.10	−13.50	−9.90	−13.50	1.30
Yogurt	−110.15	6.80	−5.00	−6.90	48.92	35.20
Detergent	−1.90	−3.27	13.08	28.34	35.90	13.33
Out-of-sample log-likelihood difference						
Catsup	−0.87	−3.70	3.40	6.20	6.70	53.40
Yogurt	37.40	−51.80	119.10	99.30	119.00	138.49
Detergent	6.52	12.33	47.14	2.83	44.65	9.35

based on choice shares indeed improves. As we only want to illustrate that the latent preference model with unobserved heterogeneity can outperform the choice share model, we stop adding segments when this goal is reached in-sample as well as out-of-sample. We see that with 6 segments the latent preference model outperforms the choice share model both in-sample as out-of-sample for all three categories. Unreported results show that similar results are found for all other suggested specifications, except for the inclusive value model for catsup. For this category the inclusive value model has the worst in-sample performance for all segments. For one to five segments the out-of-sample performance is also worst of all, when six segments are considered the inclusive value model performs better than the choice share specification.

To analyze the relative performance of all specifications in case of unobserved heterogeneity, we consider the detergent category in more detail. As can be seen from Table 5.3, for this category it holds that up to two segments the specification based on choice shares outperforms the latent preference model on the basis of in-sample log likelihood value. In case three or more segments are used the advantage of the household-specific choice shares is compensated by the heterogeneity captured by the part of the model that captures brand choice. In Table 5.4 we present the in-sample and out-of-sample log likelihood value for the detergent category for all non-choice share models. We conclude that the latent preference model performs relatively best for most number of segments. Only when

Table 5.4: Performance measures for interpurchase time models with unobserved heterogeneity for the detergent category (largest likelihood value per number of segments in boldface)

	Number of segments					
	1	2	3	4	5	6
In-sample $\ln \mathcal{L}$						
Choice model	-2335.37	-2213.76	-2133.46	-2085.64	-2046.62	-2016.14
Inclusive value	-9012.75	-8680.78	-8582.07	-8522.32	-8452.02	-8423.58
Weighting hazard	-9004.35	-8677.84	-8583.75	-8519.96	-8459.35	-8418.73
Weighting mark. mix	-9004.11	-8677.62	-8596.64	**-8513.10**	-8449.39	-8421.50
Latent preferences	**-8998.17**	**-8673.07**	**-8578.36**	-8516.65	**-8443.07**	**-8410.93**
Out-of-sample $\ln \mathcal{L}$						
choice model	-2254.69	-2195.01	-2132.77	-2115.76	-2077.77	-2071.70
Inclusive value	-8317.58	-8164.67	-8120.61	-8113.00	-8084.83	-8058.41
Weighting hazard	-8312.68	-8154.85	-8124.28	-8097.82	-8073.05	-8063.59
Weighting mark. mix	-8312.20	-8154.02	-8149.01	**-8086.90**	-8042.93	**-8042.36**
Latent preferences	**-8311.87**	**-8145.12**	**-8101.84**	-8088.08	**-8039.82**	-8062.23

four segments are used it is beaten by the specification based on a weighted marketing mix. This last specification however performs worst for three segments. The out-of-sample performance measures show a similar pattern.

To check whether these results also hold for the other two categories, we provide in Table 5.5 an overview of the performance of the non-choice share based models. To prevent that our results are influenced by the number of segments imposed, we report in the final column of the table the average rank of the model across the three product categories for the different segment sizes. If we consider the in-sample measures, the latent preference model specification performs best for all segment sizes. For out-of-sample measures the latent preference model is best or second best. The final column of the table shows the overall rank. We see that the overall rank of the latent preference model is best and that the inclusive value specification has the largest average rank value. This result holds for in-sample as well as out-of-sample performance.

Table 5.5: Average ranks of heterogeneous models over three categories (excluding choice shares specification)

	1	2	3	4	5	6	overall
In-sample average rank							
Inclusive value	4.00	3.33	2.33	3.00	2.67	4.00	3.22
Weighted hazard	2.67	3.33	3.33	3.33	3.33	2.67	3.11
Weighted mark. mix	2.33	2.33	3.00	2.00	2.67	2.33	2.44
Latent preferences	1.00	1.00	1.33	1.67	1.33	1.00	1.22
Out-of-sample average rank							
Inclusive value	4.00	3.67	3.33	4.00	4.00	3.33	3.72
Weighted hazard	2.33	2.33	2.67	2.67	2.33	2.67	2.50
Weighted mark. mix	2.00	2.00	3.00	1.33	2.33	1.67	2.06
Latent preferences	1.67	2.00	1.00	2.00	1.00	2.33	1.67

5.4 Conclusions

In this chapter we have considered the practical question of what to do with brand-specific marketing efforts when modeling interpurchase timing. As purchase timing is measured on the category level, one has to somehow aggregate the brand level information. In the literature there are two popular techniques. Category marketing efforts are often formed by calculating a weighted average of the marketing mix of individual brands. As weights household-specific choice shares are often used. Another approach is to summarize all marketing-mix variables of all brands into the so-called inclusive value.

We have proposed three alternative specifications. For the first alternative we create category level marketing instruments using household-specific weights that are obtained from a brand choice model. The second alternative uses the same weights to aggregate over brand-specific incidence probabilities (or hazards). Finally we suggested a specification that integrates a brand choice model with the purchase timing.

In an empirical comparison of the resulting five specifications for three categories of fast-moving consumer goods, we find that when unobserved heterogeneity is not accounted for the specification using choice shares performs best. However, this specification is less useful for out-of-sample forecasting as in this case household's choice shares are in general unknown. For out-of-sample forecasting the latent preference model tends to perform best.

If unobserved heterogeneity is accounted for the latent preference model also performs best in sample.

We conclude this chapter with a practical summary of the results. If one is only interested in describing purchase timing and not in brand choice, one obtains the best performance by weighting the marketing mix in the interpurchase time model using individual choice shares. However, if one wants to use the model for out-of-sample forecasting and out-of-sample individual choice shares are unknown, one has to use one of the other models. In that case, information from a brand choice model can be used to weight brand-specific marketing efforts for the interpurchase model. These models outperform the choice share based model if one explicitly models the unobserved heterogeneity in the brand choices. The overall performance of the latent preference model where one integrates brand choice and interpurchase timing is best, although the differences with the weighted marketing mix and weighted hazard specification are sometimes not substantial. However, the latent preference model outperforms the inclusive value specification.

Chapter 6

A dynamic duration model

6.1 Introduction

For marketing managers it is important to understand the dynamic effects of marketing-mix variables like promotion and advertising on marketing performance measures such as sales, market shares and profitability. Particularly, it is relevant to understand the long-run effects of marketing efforts, as this knowledge can for example lead to more efficient marketing strategies. Examples of recent studies that address this issue are Mela *et al.* (1997), Dekimpe *et al.* (1999), Jedidi *et al.* (1999) and Paap and Franses (2000), to mention just a few. The literature contains two different approaches. One approach tries to capture the (long-run) effects of marketing instruments on for example the price elasticity (Mela *et al.*, 1997; Jedidi *et al.*, 1999). Dynamics are then incorporated through the responses to marketing instruments. The second approach, which is considered in the present chapter, focuses on dynamic effects in behavior (Dekimpe *et al.*, 1999; Paap and Franses, 2000).

In this chapter we address the issue of measuring the long-run and short-run impact of marketing-mix variables on interpurchase times. The theoretical and empirical analysis of purchase-timing behavior of households has received considerable attention in the past and in recent years. The analysis of interpurchase timing can give interesting insights in household behavior. Purchase timing can be especially informative to learn about inventory management and consumption rate, in fact the purchase timing and the volume bought are the only two measures related to inventory that are usually available. Furthermore, we can study purchase acceleration and stock piling using interpurchase timing. Blattberg *et al.* (1981) show under stringent conditions that promotions lead to purchase acceleration. Empirical evidence for this behavior can be found in Gupta (1988) and Helsen and Schmittlein (1992) among others, although for example Neslin *et al.* (1985)

report that promotions were less likely to accelerate purchase times. A review of interpurchase time modeling before 1990 can be found in Jain and Vilcassim (1991, Table 1).

More recently, researchers focus on using hazard functions to analyze the effect of promotions on interpurchase times, see among others Helsen and Schmittlein (1992, 1993), Jain and Vilcassim (1991), Vilcassim and Jain (1991), Gönül and Srinivasan (1993a), Chintagunta and Prasad (1998) and Vakratsas and Bass (2002). The last four studies also incorporate unobserved household heterogeneity. An important extension to modeling interpurchase times for two related product categories is given in Chintagunta and Haldar (1998).

Dynamic models for interpurchase times are relatively scarce, although one might expect strong dynamic effects in practice. For example, a promotion may shorten the present interpurchase time, while it likely lengthens future interpurchase times due to stock piling. An example of a study that explicitly incorporates dynamic structures in purchase timing is Allenby *et al.* (1999). In that paper, dynamics in durations are modeled by lagged interpurchase times, but no explicit separation of long-run from short-run effects of marketing mix variables is pursued. As we believe that such differences might exist, we aim to contribute to the literature by putting forward a dynamic model for interpurchase times that does allow for different long-run and short-run effects. The model extends the familiar accelerated failure-time model by including lagged interpurchase times as well as lagged covariates. Rewriting this model as an Error Correction Model [ECM] allows us to distinguish the long-run from the short-run effects, see Hendry *et al.* (1984).

The values of covariates, like price and promotion, are likely to change during interpurchase spells. In most marketing applications of duration models, it is assumed that covariates remain constant during spells, which is perhaps imposed for convenience. In contrast, in this chapter we follow a similar approach as Gupta (1991), that is, we allow for time-varying covariates in the hazard specification. Additionally, many studies have emphasized the relevance of unobserved household heterogeneity, and that it should be taken into account when analyzing purchase behavior. Therefore, we accommodate for unobserved differences across households by a latent class approach. In many studies, unobserved heterogeneity is incorporated using a mixed proportional hazard model, where one introduces a stochastic multiplicative factor to the hazard specification, see Lancaster (1979) and see Gönül and Srinivasan (1993a) for an application in marketing. In this chapter we also allow for different effects of the covariates including the marketing mix on the interpurchase times, see also Vakratsas and Bass (2002) for a similar approach. Indeed, such heterogeneity may be especially relevant for modeling dynamics in purchase timing. For example, the population may contain households with very different dynamic purchase timing patterns.

In sum, we propose a dynamic model for interpurchase time, with possibly differing short-run and long-run effects of covariates, which incorporates unobserved heterogeneity and also takes care of time-varying covariates in between purchases.

The outline of this chapter is as follows. In Section 6.2, we discuss our dynamic duration model. We show how the accelerated failure-time model can be extended to allow for time-varying covariates and possibly differing long-run and short-run effects of marketing variables. We discuss in detail how one can interpret the parameters and estimate them using maximum likelihood. In Section 6.3, we apply our model to purchases in three distinct categories of frequently purchased consumption goods, that is liquid laundry detergent, catsup and yogurt. One of our main empirical findings is that, for some household segments, the short-run effects of marketing mix variables are significantly different from the long-run effects. In Section 6.4, we conclude this chapter with a discussion of the main results and with suggestions for further research topics.

6.2 A dynamic model for interpurchase times

In this section we put forward our dynamic model for interpurchase times, which enables a separate evaluation of long-run and short-run effects of covariates, such as promotion and other marketing-mix variables. In Section 6.2.1, we present the functional form of the hazard specification and discuss how we take care of time-varying covariates. In Section 6.2.2, we introduce autoregressive dynamics in our model. The interpretation of the dynamic structure is discussed in Section 6.2.3. Finally, in Section 6.2.4, we consider parameter estimation.

6.2.1 Hazard specification

Assume that a household $i = 1, \ldots, I$ purchases a certain product at time d_{in}, for $n = 0, \ldots, N_i$ over a certain period of time. The N_i interpurchase times of this household are therefore defined by $t_{in} = d_{in} - d_{i,n-1}$ with $n = 1, \ldots, N_i$. To model the interpurchase times, we consider a hazard specification. Denote the hazard corresponding to the n-th purchase decision of household i by

$$\lambda_{in}(t|x_{in}(t), \theta_i), \tag{6.1}$$

where $x_{in}(t)$ denotes a vector of covariates explaining the hazard of household i for the n-th purchase decision at time t and θ_i is a household-specific parameter vector. The explanatory variables are a function of time t. In this chapter, t denotes the interpurchase time. For the n-th interpurchase spell the calendar time is given by $d_{i,n-1}+t$. Hence $x_{in}(t)$ gives the value of the covariates at calendar time $d_{i,n-1} + t$.

When modeling interpurchase times, it is unrealistic to assume that the covariates are constant during the spell, see Gupta (1991). In other words, it is unrealistic to assume that households do not notice these variations at no-purchase store visits. The set of covariates will usually include marketing instruments such as price and display. These variables do not evolve smoothly over time, rather, they usually change on a weekly basis or perhaps every day. Denote by τ_l, for $l = 0, .., L$, the time indexes where there is a change in one of the covariates. For ease of exposition we assume that the covariates are constant within a week. Week 1 then corresponds to the time interval $[\tau_0, \tau_1]$. Denote by $K_{in}(t)$ the week number corresponding to t time periods after the start of the n-th spell of household i. This week then starts at $\tau_{K_{in}(t)-1}$ and ends at $\tau_{K_{in}(t)}$. This notation is also used in Chapter 5. In Figure 5.1 of that chapter we give a graphical representation of the purchase process. In this example we have purchases in weeks 2 and 4, and we would therefore have $K_{in}(0) = 2$ and $K_{in}(t_{in}) = 4$.

To derive the distribution of the interpurchase times we use the fact that the survivor function $S_{in}(t|x_{in}(t), \theta_i)$ equals $\exp(-\Lambda_{in}(t|x_{in}(t), \theta_i))$, where $\Lambda_{in}(t|x_{in}(t), \theta_i)$ is the integrated hazard function. This function is defined as

$$\Lambda_{in}(t|x_{in}(t), \theta_i) = \int_0^t \lambda_{in}(u|x_{in}(u), \theta_i)du. \tag{6.2}$$

Note that the integrated hazard function depends on the whole path of $x_{in}(u)$ for $u = 0$ to $u = t$, see Lancaster (1990) for a discussion. This integral can be decomposed by identifying intervals in which $x_{in}(t)$ is constant. We decompose the integral in three parts, that is, (i) from the start of the duration to the end of the corresponding week, (ii) the weeks completely contained in the duration, and (iii) from the start of the final week to the end of the duration. The integrated hazard can be decomposed as

$$\begin{aligned}\Lambda_{in}(t|x_{in}(t), \theta_i) = &\int_0^{\tau_{K_{in}(0)-d_{i,n-1}}} \lambda_{in}(u|x_{in}(u), \theta_i)du \\ &+ \sum_{k=K_{in}(0)}^{K_{in}(t)-2} \int_{\tau_k - d_{i,n-1}}^{\tau_{k+1}-d_{i,n-1}} \lambda_{in}(u|x_{in}(u), \theta_i)du + \int_{\tau_{K(t)-1}-d_{i,n-1}}^{t} \lambda_{in}(u|x_{in}(u), \theta_i)du,\end{aligned} \tag{6.3}$$

see Gupta (1991) for a similar approach.

As the computation of the integrated hazard function is computationally intensive, it is convenient to have a closed-form expression for the individual elements of (6.3). Therefore, in this chapter we use as starting point the hazard structure of an accelerated failure-time model with a log-logistic baseline hazard. This hazard specification leads to an analytical expression of the integrated baseline hazard, while it allows for a non-monotonic hazard function. Another advantage of using an accelerated failure-time specification is that it corresponds with a linear representation for the case of constant covariates during spells,

see Kalbfleisch and Prentice (1980), Kiefer (1988) and Ridder (1990). This facilitates the inclusion and interpretation of an autoregressive dynamic structure in the model, see also the Autoregressive Conditional Duration [ACD] model of Engle and Russell (1998) for modeling financial transaction data. The hazard we consider thus reads as

$$\lambda_{in}(t|x_{in}(t),\theta_i) = \exp(-x_{in}(t)'\beta_i)\lambda_0(t\exp(-x_{in}(t)'\beta_i)|\delta_i), \tag{6.4}$$

where $\theta_i = (\beta_i, \delta_i)$ and where $\lambda_0(u|\delta)$ denotes the baseline hazard. In case of the log-logistic distribution, this baseline hazard is defined as

$$\lambda_0(u|\delta) = \frac{\delta u^{\delta-1}}{1+u^{\delta}}. \tag{6.5}$$

The hazard function then becomes

$$\begin{aligned}\lambda_{in}(t|x_{in}(t),\theta_i) &= \exp(-x_{in}(t)'\beta_i)\lambda_0(t\exp(-x_{in}(t)'\beta_i)) \\ &= \frac{\delta_i t^{\delta_i-1}\exp(-x_{in}(t)'\beta_i)^{\delta_i}}{1+t^{\delta_i}\exp(-x_{in}(t)'\beta_i)^{\delta_i}}.\end{aligned} \tag{6.6}$$

When the covariates for the n-th spell for household i are constant over the interval $(a, b]$, the integrated baseline hazard over this interval equals

$$\begin{aligned}\int_a^b \lambda_{in}(u|x_{in}(b),\theta_i)du = \\ \log[1+b^{\delta_i}\exp(-x_{in}(b)'\beta_i)^{\delta_i}] - \log[1+a^{\delta_i}\exp(-x_{in}(b)'\beta_i)^{\delta_i}].\end{aligned} \tag{6.7}$$

This result can be used to compute (6.3). The density function for observation t_{in} can be expressed in terms of the hazard function and the integrated hazard function, that is,

$$f_{in}(t_{in}|x_{in}(t),\theta_i) = \lambda_{in}(t_{in}|x_{in}(t),\theta_i)S_{in}(t_{in}|x_{in}(t),\theta_i), \tag{6.8}$$

where $S_{in}(t_{in}|x_{in}(t),\theta_i) = \exp(-\Lambda_{in}(t_{in}|x_{in}(t),\theta_i))$ denotes the survivor function. Note that the density function depends on the integrated hazard function and therefore takes into account the complete history of the marketing instruments.

In Figure 6.1 we present an illustrative example of the accelerated failure-time hazard model with time-varying explanatory variables. In this figure the left-hand vertical axis gives the hazard rate, while the right-hand vertical axis gives the "score" $(x_{in}(t)'\beta)$ which changes at calendar times τ_l (lower line). The smooth curve shows the baseline hazard, that is, the hazard without correcting for the explanatory variables. Finally the kinked top line shows how the hazard is scaled and stretched as a result of the time-varying explanatory variables. A low "score" $x_{in}(t)'\beta$ yields a higher hazard rate and therefore lowers the expected interpurchase time.

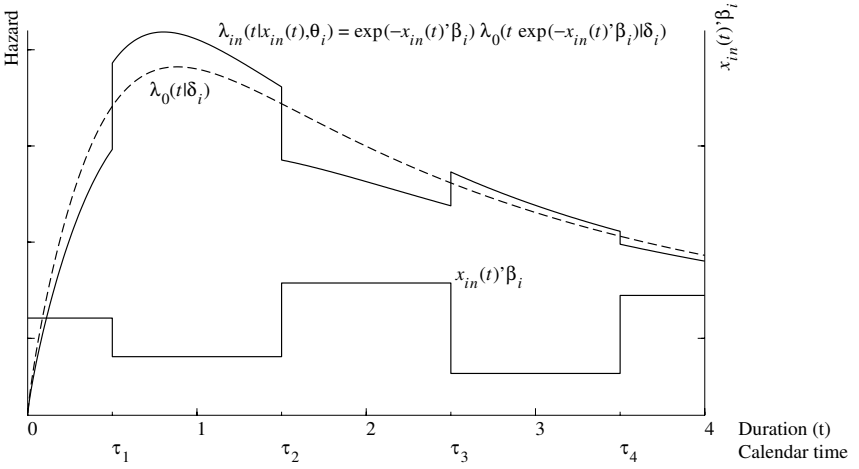

Figure 6.1: Illustrative example of the hazard rate for an accelerate failure-time model with time-varying explanatory variables.

6.2.2 Dynamics

The model discussed in the previous section is static, in the sense that interpurchase times are only explained by current explanatory variables. It is however likely that the interpurchase times of households are correlated over time. For example, promotional activities may not only have an effect on current but also on future interpurchase times.

A flexible specification of these dynamic patterns is obtained by adding lagged interpurchase times and by incorporating the value of the marketing instruments at the last purchase. Technically speaking, we add $\ln t_{i,n-1}$ and $x_{i,n-1}(t_{i,n-1})$ to explain the n-th interpurchase time of household i. Defining $w_{in}(t) = (x_{in}(t), \ln t_{i,n-1}, x_{i,n-1}(t_{i,n-1}))$ and $\gamma_i = (\alpha_i, \rho_i, \omega_i)$, we can easily obtain the hazard specification for the dynamic case from (6.6) by replacing $x_{in}(t)'\beta_i$ with $w_{in}(t)'\gamma_i$.

The hazard corresponding with this dynamic duration model has to be defined as a conditional hazard given the previous interpurchase time. This conditional hazard function for t_{in} given $t_{i,n-1}$ reads as

$$\lambda_{in}(t|w_{in}(t),\theta_i) = \frac{\delta_i t^{\delta_i-1}\exp(-w_{in}(t)'\gamma_i)^{\delta_i}}{1+t^{\delta_i}\exp(-w_{in}(t)'\gamma_i)^{\delta_i}}, \tag{6.9}$$

with $\theta_i = (\gamma_i, \delta_i)$ and note that $w_{in}(t)$ contains $\ln t_{i,n-1}$. The density function of the timing of the n-th purchase occasion of household i given $t_{i,n-1}$ is therefore

$$f_{in}(t|w_{in}(t),\theta_i) = \lambda_{in}(t|w_{in}(t),\theta_i)\exp(-\Lambda_{in}(t|w_{in}(t),\theta_i)), \tag{6.10}$$

where $\Lambda_{in}(t|w_{in}(t),\theta_i)$ is the integrated hazard function.

6.2.3 Interpretation of dynamics

For a straightforward interpretation of the parameters in the dynamic specification of the duration model, it is useful to consider the effects of a promotion during a spell. More specifically, in this section we study the dynamic effects of a promotion starting directly following a purchase made by a focal household and the promotion will end directly after the next purchase of this household. Under this strategy the marketing instruments are constant during spells. The fact that marketing instruments do not change between purchases allows for an intuitive interpretation of the parameters.

In case the covariates are constant during spells we can denote $w_{in}(t) = w_{in}$, furthermore the survivor function now simplifies to

$$S_{in}(t|w_{in}, \theta_i) = \frac{1}{1 + [\exp(-w'_{in}\gamma_i)t]^{\delta_i}}. \tag{6.11}$$

For the distribution of $t^*_{in} = \delta_i(\ln t_{in} - w'_{in}\gamma_i)$, we derive that

$$\begin{aligned} \Pr[t^*_{in} < E] &= \Pr[\delta_i(\ln t_{in} - w'_{in}\gamma_i) < E] = \Pr[t_{in} < \exp(x'_{in}\beta_i + 1/\delta_i E)] \\ &= 1 - S(\exp(w'_{in}\gamma_i + 1/\delta_i E)) = 1 - \frac{1}{1 + \exp(1/\delta_i E)^{\delta_i}} \\ &= 1 - \frac{1}{1 + \exp(E)}. \end{aligned} \tag{6.12}$$

Hence, in the case of constant regressors, t^*_{in} has a logistic density. As its density does not depend on covariates and model parameters, we can linearize the duration model as

$$\begin{aligned} \ln t_{in} &= w'_{in}\gamma_i + \sigma_i\eta_{in} \\ &= \rho_i \ln t_{i,n-1} + x'_{in}\alpha_i + x'_{i,n-1}\omega_i + \sigma_i\eta_{in}, \end{aligned} \tag{6.13}$$

where $\sigma_i = 1/\delta_i$, and where η_{in} is logistic distributed such that $\mathrm{E}[\eta_{in}] = 0$. Thus, under the restriction of constant covariates during interpurchase spells, (6.13) is an exact alternative representation of the hazard model in (6.9).

Following ideas from the area of time-series analysis, we further rewrite (6.13) into the so-called error-correction format, that is,

$$\Delta \ln t_{in} = \Delta x'_{in}\alpha_i + (\rho_i - 1)(\ln t_{i,n-1} - x'_{i,n-1}\beta_i) + \sigma_i\eta_{in}, \tag{6.14}$$

where $\beta_i = (\alpha_i + \omega_i)/(1 - \rho_i)$ and where Δ is the first difference operator defined as $\Delta z_{in} = z_{in} - z_{i,n-1}$, where z_{in} can be $\ln t_{in}$ or x_{in}. To exclude the implausible explosive behavior of the interpurchase times, we impose that $|\rho_i| < 1$, for all i. The term $\Delta x'_{in}\alpha_i$ in (6.14) concerns the short-run effects of a change in x_{in} on the interpurchase time, while the term $x'_{i,n-1}\beta_i$ in the so-called error correction part concerns the long-run effects. Notice that we cannot estimate different short- and long-run effects of variables that do

not change during the time period considered, like for example household size, as then Δx_{in} is zero and α_i is not identified. We label the model in (6.14) the error-correction model [ECM] and it will be the main specification in the rest of the chapter. The long-run effects in this specification may differ in size or even in sign from short-run effects.

Following the usual time series terminology, see Hendry *et al.* (1984), there may be a restricted specification that is relevant. First of all, we can restrict the short-run and the long-run parameters to be equal ($\alpha_i = \beta_i$). The resulting specification is known as the common-factor specification. Under this specification, (6.13) transforms into

$$(\ln t_{in} - x'_{in}\beta_i) = \rho_i(\ln t_{i,n-1} - x'_{i,n-1}\beta_i) + \sigma_i\eta_{in}. \tag{6.15}$$

The autoregressive parameter ρ_i is in this case equal to the correlation between $(\ln t_{in} - x'_{in}\beta_i)$ and $(\ln t_{i,n-1} - x'_{i,n-1}\beta_i)$. This specification is equivalent to a model where only contemporaneous explanatory variables are included and where the error term follows an AR(1) model, that is (6.15) is equivalent to

$$\begin{aligned} \ln t_{in} &= w'_{in}\gamma_i + \sigma_i\eta_{in} \\ \eta_{in} &= \rho_i\eta_{i,n-1} + u_{in}, \end{aligned} \tag{6.16}$$

where u_{in} is again an unobserved error term for which we take the same distributional assumptions as for η_{in}. If we additionally impose ρ_i to be zero we obtain a static specification in which there are no dynamic effects.

Effects of a promotion during spells

We next analyze the dynamic effects of the explanatory variables on interpurchase times. The short-run effect of a marketing instrument is defined as the instantaneous effect of its (permanent) change on the interpurchase time. The long-run effect measures the effect of a permanent change of a marketing instrument at time t' on the interpurchase times at t as $t \to \infty$. We focus on the error-correction duration model (6.14) as this model nests the common factor representation (6.15) ($\alpha_i = \beta_i$) and the static model ($\alpha_i = \beta_i$ and $\rho_i = 0$).

First, we consider the derivative of $\ln t_{in}$ with respect to x_{in}, that is,

$$\frac{\partial \ln t_{in}}{\partial x_{in}} = \alpha_i. \tag{6.17}$$

Hence, an ε change in x_{in}, for example due to a price reduction or a promotional activity, leads to $\alpha_i\varepsilon$ change in the log current interpurchase time. Note that if x_{in} is for example the natural log of a variable, we can interpret α_i as an elasticity.

To analyze the effects of changes in the explanatory variables on future log interpurchase times, we can follow a similar procedure. The partial derivative of $\ln t_{i,n+1}$ with

respect to x_{in} is given by

$$\frac{\partial \ln t_{i,n+1}}{\partial x_{in}} = -\alpha_i - (\rho_i - 1)\beta_i + \rho_i \frac{\partial \ln t_{in}}{\partial x_{in}} = (\rho_i - 1)(\alpha_i - \beta_i). \tag{6.18}$$

An ε change in x_{in} leads to a change of $\varepsilon(\rho_i - 1)(\alpha_i - \beta_i)$ in $\ln t_{i,n+1}$. The derivative is zero if $\alpha_i = \beta_i$. Note that the case with $\rho_i = 1$ is ruled out to avoid explosive interpurchase times. Hence, the common factor specification (6.15) (and of course the static model) imposes that changes in x_{in} have no effect on the *next* interpurchase time. If $\beta_i < \alpha_i$, some of the effect of the change in x_{in} on the current interpurchase time is compensated by an opposite effect on the next interpurchase time.

For the subsequent interpurchase time it holds that

$$\frac{\partial \ln t_{i,n+2}}{\partial x_{in}} = \rho_i \frac{\partial \ln t_{i,n+1}}{\partial x_{in}} = \rho_i(\rho_i - 1)(\alpha_i - \beta_i). \tag{6.19}$$

To derive the partial derivative of $\ln t_{i,n+k}$ with respect to x_{in}, we note that for $r > 2$ $\partial \ln t_{i,n+r}/\partial x_{i,n} = \rho_i \partial \ln t_{i,n+r-1}/\partial x_{in}$ and hence that

$$\frac{\partial \ln t_{i,n+k}}{\partial x_{in}} = \rho_i^{(k-1)}(\rho_i - 1)(\alpha_i - \beta_i). \tag{6.20}$$

If $|\rho_i| < 1$ the effect of a change in x_{in} on future interpurchase times will decline exponentially, and eventually it becomes zero.

From the above exercise it can already be understood that permanent changes in interpurchase times can only be obtained when x_{in} changes permanently. For example, our model implies that only a permanent lower price can generate a permanent reduction in interpurchase times. To derive the long-run effects of a permanent change in x_{in}, we apply repeated backward substitution to (6.14) and obtain

$$\begin{aligned}
\ln t_{in} &= \rho_i \ln t_{i,n-1} + \Delta x'_{in}\alpha_i - (\rho_i - 1)x'_{i,n-1}\beta_i + \sigma_i \eta_{in} \\
&= \rho_i^2 \ln t_{i,n-2} + \Delta x'_{in}\alpha_i + \rho_i \Delta x'_{i,n-1}\alpha_i \\
&\qquad - (\rho_i - 1)x'_{i,n-1}\beta_i - \rho_i(\rho_i - 1)x'_{i,n-2}\beta_i + \sigma_i \eta_{in} + \rho_i \sigma_i \eta_{i,n-1} \\
&= \rho_i^n \ln t_{i0} + \sum_{j=0}^{n-1} \rho_i^j (\Delta x'_{i,n-j}\alpha_i - (\rho_i - 1)x'_{i,n-j-1}\beta_i + \sigma_i \eta_{i,n-j}),
\end{aligned} \tag{6.21}$$

where t_{i0} denotes the pre-sample starting value of t_{in}. As $|\rho_i| < 1$, $\rho_i^n \to 0$ for large n and the influence of $\ln t_{i0}$ can be neglected. If we further impose that x_{in} is fixed over the purchase occasions, that is $x_i = x_{in} = x_{i,n-j}$, $j = 1, \ldots, \infty$, then for $n \to \infty$, (6.21) becomes

$$\ln t_{in} = \sum_{j=0}^{\infty} \rho_i^j (-(\rho_i - 1)x'_i\beta_i + \sigma_i \eta_{i,n-j}) = x'_i\beta_i + \sum_{j=0}^{\infty} \rho_i^j \sigma_i \eta_{i,n-j}. \tag{6.22}$$

Hence, as $\mathrm{E}[\eta_{i,n-j}] = 0$ for all j, the long-run expectation of $\ln t_{in}$ given x_i is

$$\mathrm{E}[\ln t_{in}|x_i] = x_i'\beta_i. \tag{6.23}$$

If follows from (6.23) that the long-run effect of a permanent change in x_i on the log interpurchase time is β_i. In sum, our error-correction model for interpurchase times has short-run effects α_i and long-run effects β_i. Note again that in the common factor model these effects both equal β_i and in the static model there are no dynamic effects at all.

Finally, to describe the dynamics in the interpurchase time model, we obtain the effects of an unexplained shock during the n-th interpurchase time on the subsequent purchase timings. The effect of a shock on future interpurchase times can easily be obtained from (6.21). The effect of an ε shock during the n-th interpurchase spell on the $n+k$-th spell is given by $\rho_i^k\varepsilon$.

Table 6.1 gives an overview of all the dynamic effects for several relevant versions of the model, that is the static version ($\rho_i = 0, \alpha_i = \beta_i$), the common factor model ($\rho_i \neq 0, \alpha_i = \beta_i$), the error-correction model with $\rho_i = 0$ and the error-correction model with unrestricted ρ_i. For ease of exposition we suppress the index i in this table. The table clearly shows the differences across the models. The static model does not allow for dynamic effects. The common factor model only captures the dynamics through unexplained shocks. The error-correction model with $\rho_i = 0$ does not capture dynamics in shocks but allows the effect of a marketing effort to carry over to the next interpurchase spell. Depending on the estimated difference between α_i and β_i, the effect on this next spell may be positive or negative. Finally, the unrestricted error-correction model allows for dynamics through shocks, for multi-period carry-over effects of marketing instruments and it allows that the effect of a permanent change in a marketing instrument smoothly evolves over time.

6.2.4 Parameter Estimation

Differences in interpurchase times across households may only be partly captured by including household-specific explanatory variables in the model. Furthermore, it is also not unlikely that households may react differently to promotional activities. Therefore, we allow for household-specific parameters. Using similar arguments as for brand choice, neglecting this household heterogeneity may lead to an overestimate of the persistence (in our case ρ_i) in interpurchase times. See for example Keane (1997) for a discussion of the effects of neglecting household heterogeneity on state dependence in brand choice.

Estimation of these household-specific parameters may however be difficult if we do not have enough observations for each household. To circumvent this problem, one usually assumes that the parameters are draws from a certain population distribution. This

Table 6.1: Dynamic effects of temporary and permanent changes in marketing instruments and of unexplained shocks in different model versions

	Static	Common Factor	ECM ($\rho = 0$)	ECM ($0 < \lvert\rho\rvert < 1$)
		Temporary change in x at time n		
$\partial \ln t_n / \partial x_n$	β	β	α	α
$\partial \ln t_{n+1} / \partial x_n$	0	0	$-(\alpha - \beta)$	$(\rho - 1)(\alpha - \beta)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\partial \ln t_{n+k} / \partial x_n$	0	0	0	$\rho^{k-1}(\rho - 1)(\alpha - \beta)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\partial \ln t_\infty / \partial x_n$	0	0	0	0
		Permanent change in x starting at time n		
$\partial \ln t_n / \partial x$	β	β	α	α
$\partial \ln t_{n+1} / \partial x$	β	β	β	$\rho\alpha + (1 - \rho)\beta$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\partial \ln t_{n+k} / \partial x$	β	β	β	$\rho^k\alpha + (1 - \rho)\beta \sum_{i=0}^{k-1} \rho^i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\partial \ln t_\infty / \partial x$	β	β	β	β
		Effect of an ε shock at time n		
$\ln t_n$	ε	ε	ε	ε
$\ln t_{n+1}$	0	$\rho\varepsilon$	0	$\rho\varepsilon$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\ln t_{n+k}$	0	$\rho^k\varepsilon$	0	$\rho^k\varepsilon$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\ln t_\infty$	0	0	0	0

approach is followed in the brand choice models in for example Kamakura and Russell (1989), Chintagunta *et al.* (1991) and Gönül and Srinivasan (1993b) among others.

A convenient choice is to assume that the parameters are draws from a finite mixture distribution which approximates the household heterogeneity distribution, see for example Jain *et al.* (1994) and Allenby and Rossi (1999), and see Vakratsas and Bass (2002) for an application in the context of interpurchase time modeling. The density function for household i then becomes

$$g_i(t_{i1}, \ldots, t_{i,N_i}|\theta) = \sum_{m=1}^{M} p_m h_i(t_{i1}, \ldots, t_{i,N_i}|\theta_m), \tag{6.24}$$

where M denotes the number of mixture components with $0 < p_m < 1$, $m = 1, \ldots, M$ and $\sum_{m=1}^{M} p_m = 1$, and where θ collects the parameters and $h(t_{i1}, \ldots, t_{i,N_i}|\theta_m)$ is the density function conditioned on segment m, defined as

$$h_i(t_{i1}, \ldots, t_{i,N_i}|\theta_m) = f_{i1}(t_{i1}|x_{i1}, \theta_m) S_{i,N_i}(t_{i,N_i}|w_{i,N_i}(t), \theta_m) \prod_{n=2}^{N_i - 1} f_{in}(t_{in}|w_{in}(t), \theta_m), \tag{6.25}$$

where the density function $f_{in}(t_{in}|w_{in}(t), \theta_m)$ is given in (6.10). The second term, involving the survivor function, is included for the last observation of household i when it is censored from the right, see for example Kiefer (1988) for a discussion. If there is no censoring, one can simply remove this term and replace the upper limit of the product by N_i.

The density for the first observation is denoted by $f_{i1}(t_{i1}|x_{i1}, \theta_m)$. For the first interpurchase time we do not observe the lagged interpurchase time. Instead of fixing its value, we choose to describe the initial observation by the long-run relation between interpurchase times and the marketing instruments in (6.23). To be more specific, we take the initial observation as

$$\ln t_{i1} = \mu_{0i} + x'_{i1}\beta_i + \tilde{\sigma}_i \eta_{i1}, \tag{6.26}$$

where η_{i1} has a logistic distribution. Note that we allow for a different intercept and scale parameter for the initial observation for flexibility reasons.

The parameters of duration models generally can be estimated using maximum likelihood [ML]. The log likelihood function is given by

$$\ell(\theta) = \sum_{i=1}^{I} \ln(g_i(t_{i1}, \ldots, t_{i,N_i}|\theta)), \tag{6.27}$$

where $g_i(t_{i1}, \ldots, t_{i,N_i}|\theta)$ is defined in (6.24). This log likelihood function can be maximized using standard numerical optimization algorithms. In case of household heterogeneity one may opt for the EM-algorithm of Dempster *et al.* (1977). The resulting maximum likelihood estimator denoted by $\hat{\theta}$ is normally distributed with mean θ and the information

matrix as covariance matrix. To compute this covariance matrix, we take the outer product of gradients.

Parameter estimates for the static duration model and the common factor duration model (6.15) can be obtained in a similar way. As both models are nested in the error-correction model (6.14), we can use standard likelihood ratio tests to compare the three models. For instance, under the parameter restriction $\alpha_m = \beta_m$ for $m = 1, \ldots, M$ the error-correction duration model (6.14) simplifies to the common factor model (6.15). To compare both models, we can perform a likelihood ratio test for the hypothesis $\alpha_m = \beta_m$. The corresponding likelihood ratio test statistic, is asymptotically $\chi^2(J)$ distributed under the null hypothesis, where J denotes the number of parameter restrictions.

It should be stressed that the likelihood ratio test procedure to compare two model specifications is only valid if the two models under consideration are nested. They should then have the same number of mixture components M to describe household heterogeneity. If the number of mixture components is different in the two model specifications, the test includes a test for the number of mixture components M. Likelihood ratio tests for the number of mixture components M are not asymptotically χ^2-distributed. To illustrate this, consider a common factor model with two mixture components ($M = 2$). Under the restriction $\beta_1 = \beta_2$ the mixing proportion p_1 is not identified and the likelihood ratio test statistic for $\beta_1 = \beta_2$ is not asymptotically χ^2-distributed under the null hypothesis. This phenomenon is known as the Davies (1977) problem. We will abstain from a further analysis of this issue here, and in our empirical work we will use the out-of-sample log-likelihood to determine the value of M.

6.3 Application

In this section we illustrate the dynamic duration models on scanner panel data on purchases of fast-moving consumer goods in three different categories. In Section 6.3.1, we discuss the data. In Section 6.3.2, we consider the maximum likelihood estimates of various duration models and we examine the presence of dynamic effects in interpurchase times. In Section 6.3.3, we use the estimation results to analyze the short-run and long-run effects of promotions on interpurchase times.

6.3.1 The data

The data we use are A.C. Nielsen household scanner panel data on purchases in three different categories from 1985 to 1988 in Sioux Falls, South Dakota. The three categories are liquid laundry detergent, catsup and yogurt. These three categories differ substantially in their average purchase rate, consumption patterns and storability. The dynamic patterns in purchase timing are likely to differ across these categories. A subset of these data

Table 6.2: Data characteristics three scanner panel data sets (interpurchase time measured in weeks)

	# brands	# stores	T [a]	I [b]	Interpurchase time count	Interpurchase time mean
Catsup	3	15	139	1435	14489	10.00
Detergent	13	13	97	624	7290	6.94
Yogurt	6	13	91	585	7019	4.98

[a] Number of interpurchase spells

[b] Number of households

are analyzed in Chintagunta and Prasad (1998) using a Dynamic McFadden Model. We aggregate marketing efforts to a weekly level as, in fast moving consumer goods markets, these efforts tend to be constant during a week, where the week is defined from Wednesday to Tuesday. We do allow households to have multiple purchase occasions during one week.

For each category we select households buying only of the top brands. The top brands are defined as those brands that are sold frequently enough to build up the entire marketing effort history. This selection will delete more households in some categories than in others. Furthermore, we select households which are observed to buy at least four times in the observational period. Table 6.2 shows an overview of the data. This table shows that, compared to the catsup category, there are less households for the detergent and the yogurt categories. The main reason seems to be that there are just a few (main) brands in the catsup category. For the other two categories there are more main brands but also many smaller brands. Households buying such small brands have to be completely removed from the data. The selected brands account for almost 90% of the market in all three categories. However, the selected households, that is, those never buying another brand, only account for 38, 35 and 59% for yogurt, detergent and catsup, respectively. Furthermore, we trim the number of weeks to yield a period in which all brands are available. For example, for the detergent category the data contain a brand introduction and for this category we start our analysis after this event.

For each purchase occasion, we know the timing and the volume purchased. Furthermore, for each week we know the shelf price (dollars/32oz.) of all brands and which brands are featured or displayed. As the interpurchase time is defined at the category level we need to aggregate the marketing information over stores and brands. To keep as much information as possible, we use household-specific weights in this aggregation.

Table 6.3: Explanatory variables

	Brand characteristics			Household char.		
	Price ($/32 oz.)	Display (%)	Feature (%)	Income	Size	Volume (32 oz.)
Catsup	1.20	1.52	2.96	6.40	3.44	1.11
Detergent	1.60	0.60	0.47	6.00	3.01	2.77
Yogurt	2.32	0.88	2.59	6.16	2.86	0.76

Following Gupta (1991), we use household-specific volume brand shares to aggregate over brands. In Chapter 5 we have shown that, when the focus is not so much on out-of-sample forecasting, using choice shares is good way to summarize brand specific marketing-mix instruments. Aggregation over stores is done using household-specific store weights. Note that, by using this weighting scheme, we use for each household only data on the relevant store and brand options. The easier method, that is, using the marketing instruments of the store actually visited and the brand actually chosen, is not an option here. This is because in the weeks between purchases, we have not yet observed the brand choice and therefore this information cannot be used. Due to this aggregation, the display and feature variables represent the percentage of stores featuring a brand, or having the brand on display in the category. Next to information on marketing activities and purchase timing, we also have access to some household characteristics. In our purchase timing models we use the household size, household income and the volume purchased at the previous purchase occasion.

Tables 6.2 and 6.3 give some summary statistics on the three categories and the explanatory variables. After the above-mentioned selections, we are left with at least 585 households in each category and over 7000 observed interpurchase spells. We have the most data for the catsup category, that is 1435 households with 14489 interpurchase spells. The mean interpurchase time ranges from 5 to 10 weeks. Note that marketing instruments in Table 6.3 are averaged over weeks, stores and different UPCs. The display and feature variables therefore take values between 0 and 1. The average levels of these variables seem quite small, but there are many weeks in which some UPCs were on display or featured.

6.3.2 Estimation results

To analyze interpurchase times, we consider three versions of our model, that is, the commonly considered static duration model, the common factor duration model (6.15),

and our error-correction duration model (6.14), where this last model is the most flexible. Note that we do not intend to model brand choice or purchase quantity, at least, not in this chapter. As explanatory variables we use household size, household income and the volume purchased at the previous purchase occasion (divided by 32 oz.). The latter variable is used as a proxy for "regular" and "fill-in" trips and it also accounts for the effects of household inventory behavior on purchase timing, see also Chintagunta and Prasad (1998). Furthermore, we use the actual price in dollars per 32 oz. Finally, two variables are used to indicate whether brands were on display or were featured.

To select the optimal model for each category, we use the following model selection strategy. First of all, we select the optimal number of segments to use to capture the heterogeneity. To this end, we estimate the error-correction specification (6.14) of the duration model for different numbers of segments. The performance of each model is measured using the log likelihood on an out-of-sample selection of households. We use 75% of the households for parameter estimation and the remaining 25% are the out-of-sample households. The number of segments yielding the highest out-of-sample log likelihood is then selected. Note that by using an out-of-sample measure to select the number of segments, we reduce the probability of overfitting the data. For the selected number of segments, we test whether we can restrict the dynamic structure of the error-correction model to a common factor (6.15) or to a static representation. As these last two models are nested within the error-correction specification, we can use Likelihood Ratio [LR] tests to test for restricted dynamic structures.

For the catsup and the yogurt category, it turns out that four segments are sufficient to capture the heterogeneity, while for the detergent category we need six segments. Upon using LR-tests, the static and the common-factor specification are rejected against the error-correction model for all categories. This shows that there are indeed significant dynamic effects in interpurchase timing. Furthermore, as the common-factor model is rejected against the error-correction model, short-run effects apparently differ from the corresponding long-run effects.

In Table 6.4 and Table 6.5, we present the estimation results for the three categories for the final models. Table 6.4 shows the mean parameters over the sample, that is $\sum_{m=1}^{M} \hat{p}_m \hat{\theta}_m$. Parameters in boldface are significant at 5%. In parentheses, the table gives the segment numbers for which the (segment-specific) parameters are significantly different from zero. Note that the segments are ordered such that segment 1 is the largest. Next, Table 6.5 shows the average parameter estimates over the segments for which there is a significant effect. This table also gives the corresponding fraction of the sample. This table will be especially useful to compare the three categories.

Tables 6.4 and 6.5 display the parameter estimates of the error-correction duration model (6.14) for the yogurt category in the second column. As household size and income are constant over the time period considered, we cannot estimate a different short-run and

Table 6.4: Sample averaged ML parameter estimates for the yogurt, catsup and detergent categories for the error-correction duration model, with household heterogeneity. Significant estimates (at 5-% level) are given in boldface, and in parentheses we indicate the segments for which the segment-level estimate is significant.

	yogurt	detergent	catsup
	short-run parameters (α)		
price (32 oz.)	0.629 (12)	**0.795** (2)	**1.115** (12)
display	−1.265 (1)	**−3.010** (12 4)	**−2.347** (1234)
feature	−0.619 ( )	−0.920 (2 4)	**−1.652** (1234)
volume prev. (32 oz.)	−0.020 (3)	**0.135** (12 456)	**0.090** (234)
	long-run parameters (β)		
price (32 oz.)	**1.044** (1 3)	0.234 (2)	**−0.437** (1)
display	−0.396 ( )	**−2.090** (1)	**−1.615** (12)
feature	−0.045 ( )	−0.963 (2)	**−2.009** (1234)
household income	−0.004 ( )	**−0.019** (5)	0.006 (2)
household size	**−0.040** (1)	**−0.176** (123456)	**−0.188** (1234)
volume prev. (32 oz.)	−0.085 ( )	**0.133** (1234 6)	**0.065** (1234)
μ_1	**1.609** (123)	**2.148** (123456)	**2.892** (1234)
$\mu_0 - \mu_1$	0.067 (3)	0.037 (3)	−0.018 ( )
δ_0	**1.637** (123)	**2.341** (123456)	**2.277** (1234)
δ_1	**1.732** (123)	**2.613** (123456)	**2.008** (1234)
ρ	0.199 (123)	0.003 (56)	0.010 ( )
p_1	**0.703**	**0.432**	**0.487**
p_2	**0.182**	**0.236**	**0.258**
p_3	**0.082**	**0.139**	**0.209**
p_4	**0.034**	**0.099**	**0.046**
p_5	–	**0.048**	–
p_6	–	**0.047**	–
In-sample $\ell(\hat{\theta})$	-12927.33	-14968.93	-35693.80
LR-tests (p-values)			
static vs ECM	0.000	0.000	0.000
common factor vs ECM	0.000	0.000	0.000

Table 6.5: Average parameter estimates over segments for which the segment-specific estimates are significant at 5%, with the corresponding fraction of the sample in parentheses.

	yogurt		detergent		catsup	
			short-run effects (α)			
price (32 oz.)	0.648	(88.5)	1.538	(43.2)	1.356	(74.5)
display	−1.836	(70.3)	−3.174	(76.7)	−2.347	(100)
feature	–	(0.0)	−2.267	(33.5)	−1.652	(100)
volume prev. (32 oz.)	0.414	(8.2)	0.155	(86.1)	0.232	(51.3)
			long-run effects (β)			
price (32 oz.)	1.369	(78.5)	0.937	(23.6)	−1.014	(48.7)
display	–	(0.0)	−4.569	(43.2)	−1.933	(74.5)
feature	–	(0.0)	−3.583	(23.6)	−2.009	(100)
household income	–	(0.0)	−0.177	(4.8)	0.025	(25.8)
household size	−0.056	(70.3)	−0.176	(100)	−0.188	(100)
volume prev. (32 oz.)	–	(0.0)	0.138	(92.5)	0.065	(100)
μ_1	1.664	(96.6)	2.148	(100)	2.892	(100)
$\mu_0 - \mu_1$	0.783	(8.2)	0.298	(13.9)	–	(0.0)
δ_0	1.562	(96.6)	2.341	(100)	2.277	(100)
δ_1	1.651	(96.6)	2.613	(100)	2.008	(100)
ρ	0.205	(96.6)	0.231	(9.5)	–	(0.0)

long-run effect of these variables. Price has a significant short-run effect for the first two segments (88.5%) of the sample and a significant long-run effect for the first and the third segment (78.5%). Display only has a significant short-run effect for segment 1. Finally, the ρ parameter is significant for all but the smallest segment of the sample, its mean equals 0.199. Therefore, on average, 20% of a shock to the interpurchase time is carried over to the next spell.

The third column shows the estimation results for the detergent category. Display has significant short-run and long-run effects for 76.7% and 43.2% of the sample, respectively, where now the short-run effect is larger. In comparison to the yogurt category, the pur-

chased volume has a much larger effect on the purchase timing of detergent. If a household buys a large quantity of detergent, it is likely that the next observed interpurchase time will be longer than on average. An obvious explanation is that this product can be easily held on stock, in contrast to yogurt. For yogurt, one may expect the usage rate to increase with the purchased quantity. In the detergent category, the autocorrelation parameter ρ is only significant for the smallest categories. Averaged over these two categories the estimate of ρ equals 0.231. Although on average there is no significant correlation, for 9.5% of the sample we do find a significant autoregressive relation.

The third category we study concerns catsup. For this category we do not find a significant ρ-parameter for any segment. The speed of convergence to the steady state for this category is therefore very large. Not surprisingly, this category has the largest mean interpurchase time, see Table 6.2. Purchase occasions tend to be relatively far apart in (calendar) time, leading to a smaller correlation between interpurchase times. We however do find significantly different long-run and short-run effects. Surprisingly, for the largest segment of households a price increase decreases interpurchase times on the long-run. A possible explanation for this finding is that households may buy smaller quantities/sizes of catsup when the price is high. A permanent increase in price will in that case lead to smaller interpurchase times. Finally, we see that previous purchases only have a short-run effect, thereby again illustrating that an error-correction model can yield useful inference.

When we compare the results for the three categories in Table 6.5, we find that (i) categories with large interpurchase times have smaller autocorrelations, that (ii) short-run effects of marketing instruments differ significantly from the long-run effects, and that (iii) price has most long-run effect on the perishable product.

6.3.3 Short-run effects of promotions

In this section we examine the short and long-run effects of specific promotion scenarios on interpurchase timing. We discuss two different scenarios, one that is not very realistic and one that is realistic but more difficult to evaluate. It will turn out that the first scenario does provide very useful insights in the dynamics, and hence has important managerial implications.

First of all we analyze the effects of a promotion targeted at one specific household, the promotion starts right after an observed purchase and ends at the time of the next purchase. Under this marketing plan, the marketing efforts are constant during the interpurchase spells of this household. We can therefore use the results of Section 6.2.3 to derive the effects of such a promotion on the future interpurchase times. Using the same results we can evaluate the effects of a permanent promotion on the purchase timing of this household.

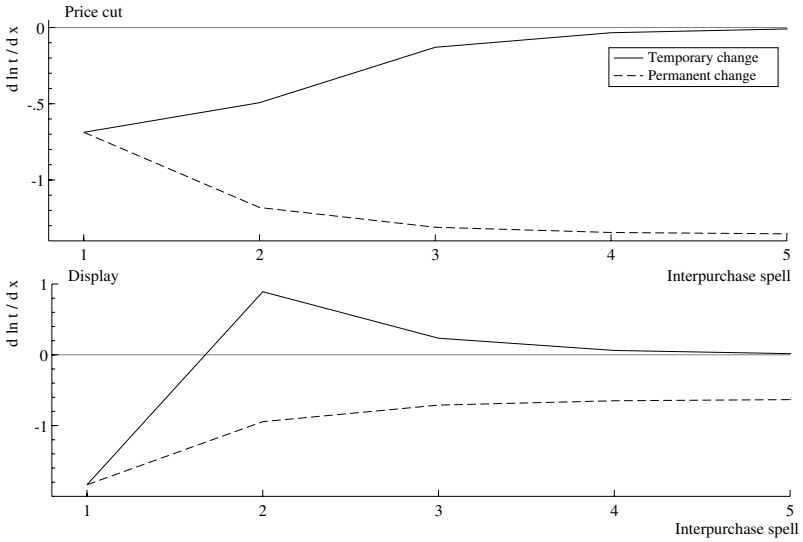

Figure 6.2: Graphical presentation of the effect of a temporary/permanent price cut or display on interpurchase times. Results apply to the largest segment of the four-segment error-correction specification for the yogurt category.

Figure 6.2 shows the effects of a price promotion and the effects of a display on the interpurchase times in the yogurt category. The results are based on the parameter estimates for the largest segment in the error-correction specification (6.14), see the second column of Table 6.4. The top graph in Figure 6.2 shows the effects of a price cut on five consecutive interpurchase spells, where the effect is represented by the partial derivative of the log interpurchase time to the marketing instrument. We see that a temporary price cut (solid line) has a strong negative effect on the first interpurchase time. The effect on the next interpurchase times is negative as well, but the size of the effect quickly converges to zero. Only for the first two interpurchase times does the price cut have a substantive impact. As the carry-over effects have the same sign as the direct effect, the long-run effect of a permanent price cut (dashed line) is larger than the direct effect.

For display, we obtain a different pattern. Although display shortens the first interpurchase time the next ones are expected to be larger than normal. For display, it therefore holds that the long-run effect of a permanent display is smaller than the direct effect. Actually the long-run effect for display is not significantly different from zero for any segment, see Table 6.4. Display therefore mainly has short-run effects, while price has short as well as long-run effects.

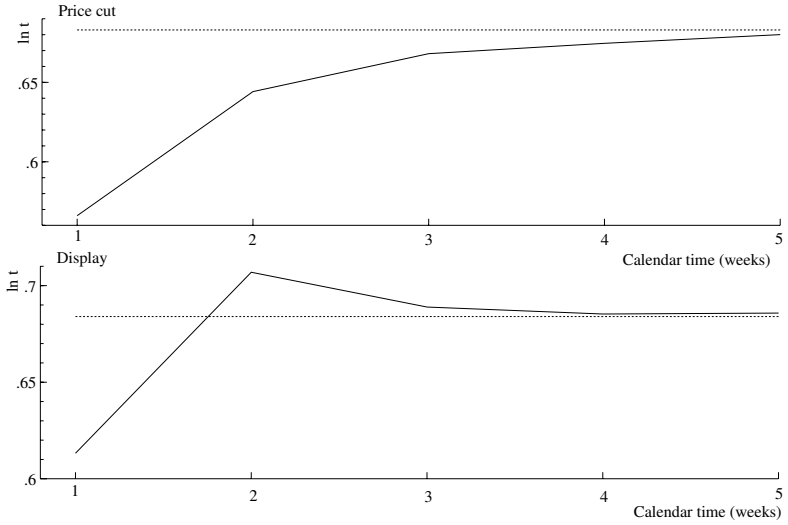

Figure 6.3: Graphical presentation of the effect of a promotion during one week on the average (log) interpurchase time.

The above analysis is directly obtained from the parameter estimates. However, it only gives the effects for a very specific situation, that is, the effects of a promotion targeted at a single household. In practice one is interested in the effects on a more aggregate level and the promotion will then not start at the beginning of the interpurchase spells of all households. A more realistic setting is a promotion during one week. To assess the impact of such a promotion, we have to rely on simulation. We use the estimated model to simulate purchases for a number of households starting at calendar time t_0, where we have a promotion from t_1 to t_2, and for $t < t_1$ and $t > t_2$, we set the marketing instruments to their sample mean, see Table 6.3. The size of the promotion is set to one standard deviation. For every week, we calculate the percentage of households that would have made a purchase. The average interpurchase time for this week is the reciprocal of this percentage. Note that we now measure the absolute effects of promotion instead of marginal effects as in the previous exercise.

Figure 6.3 shows the effects of a promotion during one week, again for the largest segment of households in the yogurt category. Contrary to the previous analysis the current scenario analysis is done in calendar time. Interestingly, this graph shows the same general pattern as the analysis on the household level, see Figure 6.2, thereby demonstrating how one can use the modeling results to evaluate the effects of managerial decisions.

6.4 Conclusion

In this chapter we proposed a dynamic model for interpurchase times, in which we can disentangle short-run from long-run effects of marketing variables. We discussed representation, interpretation and estimation issues. We illustrated our model for purchases on three different categories and we found that the short-run effects of marketing-mix variables can be significantly different from the long-run effects. Also, we found that the effects of marketing instruments, both in the short-run and the long-run, can vary substantially across categories. Additionally, we showed that our model can be used to evaluate marketing strategies.

A topic of further research amounts to building on the work of Gupta (1988); Chintagunta (1993); Ailawadi and Neslin (1998) and Bucklin *et al.* (1998), where interpurchase times or purchase incidence decisions are combined with brand choice and purchase quantity. Indeed, one could construct models for long-run and short-run effects of marketing mix variables for all marketing performance measures jointly. In this case, we could use the approach of Paap and Franses (2000) to capture dynamics in brand choice and the ideas in Böckenholt (1999) to model dynamics in purchased quantity.

Chapter 7

Responsiveness to marketing efforts

7.1 Introduction

The use of brand choice models has become standard practice in marketing research (Guadagni and Little, 1983; Chintagunta *et al.*, 1991; Jain *et al.*, 1994; Keane, 1997). In many applications of these choice models, the random utility theory framework (McFadden, 1973, 1981) is used to represent the choice process. An often made assumption in these models concerns homogeneity of households. That is, it is often assumed that all households have similar tastes, and that they only differ in their (observed) characteristics. In the relevant literature there is however ample evidence that households do differ. They may differ in their preferences or in the way they make their decisions, or in both respects. Differences in base preferences are usually referred to as preference heterogeneity. Differences with respect to the decision process are labeled structural heterogeneity.

Preference heterogeneity can partly be explained by observable characteristics. This corresponds with so-called observed preference heterogeneity. Taste is in this case usually explicitly modeled, for example by including demographic variables (see *e.g.* Maddala, 1983). However, it may be that not all heterogeneity can be attributed to observed characteristics, and hence there might be so-called unobserved preference heterogeneity. There are two popular techniques to deal with unobserved preference heterogeneity, see Allenby and Rossi (1999) and Wedel *et al.* (1999) for a discussion. These techniques are both based on the notion that when there is unobserved heterogeneity in taste, there is a corresponding preference distribution in the population. One approach imposes a continuous distribution of a known form to capture the heterogeneity. The other approach tries to approximate the unknown distribution by a discrete distribution with a fixed number of probability masses. A choice model using the latter approach is an example of a finite mixture model, see for example Wedel and Kamakura (1999).

Differences across households may not be fully attributable to preference heterogeneity. Households may also differ in the actual decision process they use to make their choices, that is, there might be structural heterogeneity. For example, Kamakura *et al.* (1996) examine brand choice within a product category where the brands carry different product forms, like volume. A household might first choose a brand and then choose the specific form to purchase. Another household might first choose a specific product form and only then consider the different brands. A third household might completely ignore all this and chooses directly from all available brand and product form combinations. Kamakura *et al.* (1996) develop a model which combines preference and structural heterogeneity and they show that the inclusion of both types of heterogeneity leads to useful managerial insights for brand competition.

Structural heterogeneity is not only relevant to sequential choice processes as choice mechanisms of households can differ in many respects. Yang and Allenby (2000), for example, present a model in which households are allowed to differ in the reference point to which options are compared. These authors use a hierarchical Bayes model to model credit card adoption, where households are allowed to differ in their decision rule and where behavior can change over time. Yang and Allenby (2000) show that there is heterogeneity in decision rules and that it can be modeled using a mixture of sub-models.

In the present chapter we extend their idea to a brand choice setting. Households, who choose amongst brands within a specific product category, may differ in their decision rules. For example, some households will spend more time and effort while making their choice than others do. If little time and effort is invested in the decision process, it is perhaps less likely that the household will respond to marketing instruments. For example, to be able to respond to price changes, one of course needs to recall the previous prices of all brands. To be able to respond to advertising, one has to read the newspaper in which the advertisement is printed. It may be unrealistic to assume that all households show such a strong involvement with the product category at all purchase occasions. Hence, it is likely that households will differ in the extent to which they are responsive to marketing efforts. Furthermore, within a household there may be differences in the responsiveness across purchase occasions. These differences correspond to differences in the decision rules being used, and therefore they can be seen as an example of structural heterogeneity.

Our model is somewhat related to the work of Bucklin and Lattin (1991). They consider a two-state model of purchase incidence and brand choice, where they distinguish between households that plan their purchases and households which act opportunistic. Furthermore Bucklin and Lattin (1991) assume homogeneous preferences, while our model also incorporates preference heterogeneity.

One reason why some households are unresponsive to marketing efforts could be just a lack of interest in marketing efforts issued by brand managers. On the other hand, economic motivations may also explain varying responsiveness across households and across

time. For example, search costs play an important role in the decision process of a household or an individual. As mentioned before, to be responsive to price changes one needs to remember the prices of every option on every purchase occasion. Additionally, people usually face time constraints. It takes time for a household to compare all prices of the options in a specific market at the time of purchase. Consider a household planning to buy many different items during the same shopping trip. There is obviously a limited amount of time available for this and therefore it may be unrealistic to assume that the household will allocate much time to each item. Following this reasoning, the more items a household purchases at a shopping trip, the less responsive this household might be to marketing efforts. Hence, the monetary value of all products purchased at a shopping trip may be inversely related to the responsiveness to marketing efforts.

Taking the above arguments along, as the decision process differs across households and across purchase occasions, the observed choice of different households cannot be explained by the same variables. Choice behavior of responsive households can be explained by their base preferences, by marketing efforts, and by their purchase history. Brand choice by unresponsive households may only be described by base preferences and purchase history. Moreover, household characteristics are rarely seen to significantly contribute to explaining brand choice, but these might be informative for the type of decision process used by the household. As such, household characteristics might indirectly influence brand choice.

In this chapter we put forward a brand choice model which incorporates responsiveness to marketing efforts as a form of structural heterogeneity. We assume that all brands are equal in appearance so that the source of structural heterogeneity studied by Kamakura *et al.* (1996) is not present, although our model can be extended to such a setting. We introduce two latent segments. In the first segment the households are assumed to respond to marketing efforts, while in the second segment households are assumed not to do so. Whether a specific household is a member of the first or the second segment at a specific purchase occasion is described by household-specific characteristics and characteristics concerning buying behavior. Additionally, to capture differences in responsiveness over time, households are allowed to switch between the two segments over time.

The remainder of this chapter contains the following. In Section 7.2, we present our model. First of all, we discuss the modeling of responsiveness to marketing efforts as a form of structural heterogeneity. Next, we extend the model to also capture preference heterogeneity. In Section 7.3, we consider parameter estimation. In this section we extensively discuss the estimation method we propose, that is, importance sampling within simulated maximum likelihood. We show that the use of importance sampling leads to substantially more accurate likelihood approximations compared to direct sampling. Section 7.4 discusses the application of this model to panel data concerning purchases of

detergent, where we also compare the performance of our model to various related choice models. In Section 7.5, we conclude this chapter with some remarks.

7.2 Model

In this section we discuss our responsiveness model. Section 7.2.1 deals with the basic model with only structural heterogeneity. In Section 7.2.2 we consider preference heterogeneity.

7.2.1 Preliminaries

To keep the presentation of the basic model simple, we will first ignore possible preference heterogeneity and only concentrate on structural heterogeneity. We assume that household $i = 1, \ldots, I$ chooses from J brands at each purchase occasion $t = 1, \ldots, T_i$. Note that different households can have a different number of observed purchase occasions. Also note that purchase occasion t of household i not necessarily corresponds to the same period in time as purchase occasion t of household $l \neq i$. The variable y_{ijt} denotes the chosen alternative, that is,

$$y_{ijt} = \begin{cases} 1 & \text{if household } i \text{ purchases brand } j \text{ at occasion } t \\ 0 & \text{otherwise.} \end{cases} \tag{7.1}$$

Furthermore we will use $y_{it} \in \{1, \ldots, J\}$ to denote the index of the chosen brand at time t.

Each household is, at any point in time, either responsive or unresponsive to marketing efforts. In case a household is unresponsive to marketing efforts, the choice can only be attributed to base preference, habit, state dependence and random influences. We introduce a latent indicator variable Z_{it} to denote the state a household is in at a specific point in time, that is,

$$Z_{it} = \begin{cases} 1 & \text{if household } i \text{ is responsive} \\ & \text{to marketing efforts at purchase occasion } t \\ 0 & \text{otherwise.} \end{cases} \tag{7.2}$$

Over time households may switch between responsiveness states. For example a household may be responsive on regular shopping trips, but unresponsive on so-called filler trips where just a few products are bought. Note that we do not observe the responsiveness state of a household over time, and hence that these have to be inferred from the data. To model the responsiveness, we consider a binary logit model (Maddala, 1983) which relates Z_{it} to household characteristics collected in W_{it}. These characteristics may also

include variables concerning the shopping trip. Examples of such variables are recency of the last purchase and the monetary amount spent on the shopping trip. The specification of the model for the responsiveness state becomes

$$\begin{aligned} Z_{it}^{*} &= \mu^{(z)} + W_{it}\gamma^{(z)} + \varepsilon_{it}^{(z)} \\ Z_{it} &= \begin{cases} 1 & \text{if } Z_{it}^{*} \geq 0 \\ 0 & \text{if } Z_{it}^{*} < 0. \end{cases} \end{aligned} \tag{7.3}$$

The disturbances $\varepsilon_{it}^{(z)}$ are assumed to follow a logistic distribution, such that,

$$\Pr[Z_{it} = 1] = \frac{\exp(\mu^{(z)} + W_{it}\gamma^{(z)})}{1 + \exp(\mu^{(z)} + W_{it}\gamma^{(z)})}. \tag{7.4}$$

In case a household is responsive to marketing efforts, marketing instruments, such as price and promotion, have an effect on the choice made by this household. We denote the marketing instruments for brand $j = 1, \ldots, J$, as experienced by household i at purchase occasion t, as X_{ijt}. To model the choice process of a marketing-responsive household we consider the Multinomial Logit [MNL] model of McFadden (1973). Conditional on responsiveness, the utility of brand j for household i at purchase occasion t is modeled as

$$U_{ijt}^{(r)} = \mu_{j}^{(r)} + X_{ijt}\beta^{(r)} + \alpha^{(r)} y_{ij,t-1} + \varepsilon_{ijt}^{(r)}, \tag{7.5}$$

where $\varepsilon_{ijt}^{(r)}$ follows a type-I extreme-value distribution and where $y_{ij,t-1} = 1$ if person i purchased brand j at purchase occasion $t-1$. This last term is included to model state dependence. State dependence refers to a dynamic property of the choice process, as it incorporates if the household's tendency to buy the same brand as purchased at the previous occasion. The degree of state dependence is measured by $\alpha^{(r)}$.

Of course, in case a household is unresponsive to marketing activities, the marketing instruments will not have an effect on its choice behavior. On these purchase occasions the brand choice will be mainly determined by base preferences, recent behavior (state dependence), and random effects. This type of behavior can be modeled by a second MNL model, that is,

$$U_{ijt}^{(u)} = \mu_{j}^{(u)} + \alpha^{(u)} y_{ij,t-1} + \varepsilon_{ijt}^{(u)}, \tag{7.6}$$

where, obviously, the X_{ijt} are excluded and where we allow the brand intercepts and the state dependence parameter to be different from the responsive case. Strictly speaking this specification does not correspond to a proper utility maximization problem. Under standard utility maximization, prices must enter the (reduced-form) utility model as prices are obviously part of a household's budget restriction[1]. Our implicit assumption in (7.6) is that households which are unresponsive to marketing efforts maximize

[1]We thank an anonymous reviewer of the Journal of Business & Economic Statistics for raising this point.

utility without considering the *actual* price differences among the brands. Instead, they aim at an approximate utility maximization that costs less effort. In this case the average, or baseline, price for each brand is used instead of the actual price. This implies that although "unresponsive" households do not take into account price promotions, they will react to permanent changes in price. The utility specification in (7.6) actually reads $U_{ijt}^{(u)} = \mu_j^{(u)} + \delta\bar{p}_j + \alpha^{(u)} y_{ij,t-1} + \varepsilon_{ijt}^{(u)}$, where $\bar{p}_j$ denotes the average long-run price of brand j. In practically available data the long run price does not vary over time, therefore we cannot separately identify δ and $\mu_j^{(u)}$. The utility specification we use for the unresponsive case therefore does not include prices, the brand intercepts in (7.6) give a combination of base preferences and price effects.

In sum, household i purchases brand j at purchase occasion t when, conditional on responsiveness, $U_{ijt}^{(r)}$ is the maximum utility among $U_{ikt}^{(r)}$, $k = 1, \ldots, J$ or, when, conditional on unresponsiveness, $U_{ijt}^{(u)}$ is the maximum utility among $U_{ikt}^{(u)}$, $k = 1, \ldots, J$. In shorthand, brand j is purchased when

$$\begin{aligned} &\left(U_{ijt}^{(r)} = \max_{k=1,\ldots,J} U_{ikt}^{(r)}\right) \Big| Z_{it} = 1 \text{ or} \\ &\left(U_{ijt}^{(u)} = \max_{k=1,\ldots,J} U_{ikt}^{(u)}\right) \Big| Z_{it} = 0. \end{aligned} \tag{7.7}$$

As the random parts of the utilities are assumed to be independently extreme-value distributed, the probability of purchasing brand j for household i, that is responsive at purchase occasion t, is

$$\Pr\left[U_{ijt}^{(r)} = \max_{k=1,\ldots,J} U_{ikt}^{(r)} \,\middle|\, Z_{it} = 1\right] = \frac{\exp(\mu_j^{(r)} + X_{ijt}\beta^{(r)} + \alpha^{(r)} y_{ij,t-1})}{\sum_{k=1}^{J} \exp(\mu_k^{(r)} + X_{ikt}\beta^{(r)} + \alpha^{(r)} y_{ik,t-1})}, \tag{7.8}$$

where $\mu_J^{(r)}$ is restricted to 0 for identification, see McFadden (1973). If the household is unresponsive at t, the probability of purchasing brand j is

$$\Pr\left[U_{ijt}^{(u)} = \max_{k=1,\ldots,J} U_{ikt}^{(u)} \,\middle|\, Z_{it} = 0\right] = \frac{\exp(\mu_j^{(u)} + \alpha^{(u)} y_{ij,t-1})}{\sum_{k=1}^{J} \exp(\mu_k^{(u)} + \alpha^{(u)} y_{ik,t-1})}, \tag{7.9}$$

with $\mu_J^{(u)} = 0$ for identification. Finally, as we do not observe whether a household at purchase occasion t belongs to the responsive segment or not, the probability that it purchases brand j at purchase occasion t is obtained by summing the conditional probabilities over the segments, that is,

$$\begin{aligned} \Pr[y_{ijt} = 1] = \Pr[U_{ijt}^{(r)} = \max_{k=1,\ldots,J} U_{ikt}^{(r)} | Z_{it} = 1] \Pr[Z_{it} = 1] \\ + \Pr[U_{ijt}^{(u)} = \max_{k=1,\ldots,J} U_{ikt}^{(u)} | Z_{it} = 0] \Pr[Z_{it} = 0], \end{aligned} \tag{7.10}$$

where $\Pr[Z_{it} = 1]$ is given in (7.4).

An interesting by-product of our model concerns the possibility to calculate the conditional probability of responsiveness at purchase occasion t, that is

$$\begin{aligned}\Pr[Z_{i,t} =1|y_{ijt} = 1] &= \frac{\Pr[Z_{it} = 1, y_{ijt} = 1]}{\Pr[y_{ijt} = 1]} \\ &= \frac{\Pr[y_{ijt} = 1|Z_{it} = 1]\Pr[Z_{it} = 1]}{\Pr[y_{ijt} = 1|Z_{it} = 1]\Pr[Z_{it} = 1] + \Pr[y_{ijt} = 1|Z_{it} = 0]\Pr[Z_{it} = 0]}.\end{aligned} \tag{7.11}$$

This expression gives the probability that household i is responsive to marketing efforts at purchase occasion t, given the fact that brand j is purchased. In the application we will show a histogram of these conditional probabilities to give an impression of the average value and the dispersion of the responsiveness in the population.

7.2.2 Preference heterogeneity

The proposed model does not include unobserved preference heterogeneity. Such preference heterogeneity can be assigned to differences in base preference. For example, some households prefer one brand over the other, while other households may have the opposite preference. In general, by allowing for different base preferences, one gets a more realistic description of consumer behavior.

There is also another reason why it is important to account for heterogeneity in preferences. It is known that when base preferences are not correctly taken into account in a model with state dependence, the state dependence of households can be overestimated, see Allenby and Lenk (1994) and Keane (1997), among others. State dependence and differences in base preferences both describe observed persistence in brand choice, but they refer to different patterns of behavior. State dependence refers to a causal link between brand choice at period t and brand choice at period $t + 1$. The fact that brand j is purchased at time t increases the probability that brand j will be purchased again at $t + 1$. Differences in base preferences also capture persistence in brand choice. In this case however, there is no causal link between brand choice at time t and brand choice at $t + 1$. Stated differently, state dependence refers to a property of the dynamics of the choice process whereas base preferences are related to exogenous factors.

If base preferences are ignored, the model component designed to capture state dependence will also capture the persistence induced by base preferences. Households with a strong base preference for a certain brand do not switch often. Such a strong base preference may therefore easily be confused for state dependence. In general, it is therefore important to model state dependence as well as heterogeneity in base preferences. In our brand choice model we capture possible differences in base preferences by allowing the intercepts in the brand choice model ($\mu^{(r)}$ and $\mu^{(u)}$) to be different across households. As

mentioned in the introduction, there are two different approaches for modeling preference heterogeneity in choice models, see Wedel *et al.* (1999). In this chapter we choose to approximate the population distribution of the utility constants by the normal distribution as in, for example, Chintagunta *et al.* (1991) and Jain *et al.* (1994).

The continuous nature of this heterogeneity distribution improves the practical identification of our model. With a mixture approach it may be difficult to separate the (discrete) preference heterogeneity from the discrete structural heterogeneity. The utility specifications in (7.5) and (7.6) allow brand intercepts to be different across the responsiveness states. The parameters of a model with a mixture distribution to capture preference heterogeneity are however identified. Identification comes from the fact that the base preferences of a household are constant, while the household may switch between responsiveness states over time. In the empirical section we will compare the relative performance of the model with discrete heterogeneity with the continuous case.

As mentioned above the base preferences of households are usually assumed to be constant over long periods of time. In the standard choice models the base preferences are constant over the entire data range. In our model we therefore have to make sure that the base preference for a given household does not depend on the responsiveness state of the household on a particular purchase occasion. We cannot simply restrict the utility intercepts of the two conditional logit models (7.8) and (7.9) to be equal as the intercepts also correct for the means of the explanatory variables. Furthermore, for the unresponsive model the brand intercepts also capture differences in baseline prices across brands. The two conditional logit models contain different explanatory variables, and we therefore cannot simply restrict the utility intercepts to be equal for the responsive and the unresponsive households. Instead, we have to use another strategy. Denote the deviation of the base preference of household i from the population mean by the J-dimensional vector ω_i. For the model with continuous heterogeneity, we model the population distribution of these deviations as $N(0, \Sigma_\omega)$. The MNL model combined with this heterogeneity specification is known as as the mixed MNL model, see McFadden and Train (2000) for an extensive discussion of this model. As intercepts for the choice model conditional on responsiveness (7.8) we now use $\mu^{(r)} + \omega_i$ and for the model conditional on unresponsiveness (7.9) we use $\mu^{(u)} + \omega_i$. Hence, the household-specific vector ω_i measures the deviation of household i's preferences from the population mean for both responsiveness states. In case a discrete distribution would be used to capture the heterogeneity, ω_i would be modeled by a discrete distribution with a fixed number of mass points. For one of the segments, ω has to be restricted to zero for identification.

We only observe the final choice of a household. The choice process, nor the base preferences of the household, are observed. The probability that household i purchases brand j at purchase occasion t now has to be marginalized on both the base preferences

as well as on the responsiveness segments, that is,

$$\Pr[y_{ijt}=1]=\int_{-\infty}^{\infty}\Pr[y_{ijt}=1|\ \omega_i]\phi(\omega_i;0,\Sigma_\omega)d\omega_i, \tag{7.12}$$

where $\phi(x;0,\Sigma)$ denotes the normal density with mean 0 and variance Σ evaluated at x. The conditional choice probability is given by

$$\begin{aligned}\Pr[y_{ijt}=1|\ \omega_i]=\Pr[y_{ijt}=1|\ \omega_i,Z_{it}=1]\Pr[Z_{it}=1]\\ +\Pr[y_{ijt}=1|\ \omega_i,Z_{it}=0]\Pr[Z_{it}=0],\end{aligned} \tag{7.13}$$

where

$$\begin{aligned}\Pr[y_{ijt}=1|\ \omega_i,Z_{it}=1]&=\frac{\exp(\mu_j^{(r)}+\omega_{ij}+X_{ijt}\beta^{(r)}+\alpha^{(r)}y_{ij,t-1})}{\sum_{k=1}^{J}\exp(\mu_k^{(r)}+\omega_{ik}+X_{ikt}\beta^{(r)}+\alpha^{(r)}y_{ik,t-1})},\\ \Pr[y_{ijt}=1|\ \omega_i,Z_{i,t}=0]&=\frac{\exp(\mu_j^{(u)}+\omega_{ij}+\alpha^{(u)}y_{ij,t-1})}{\sum_{k=1}^{J}\exp(\mu_k^{(u)}+\omega_{ik}+\alpha^{(u)}y_{ik,t-1})},\end{aligned} \tag{7.14}$$

and $\Pr[Z_{it}=1]$ and $\Pr[Z_{it}=0]$ as in (7.4). Hence, the choice probability (7.12) equals

$$\begin{aligned}\Pr[y_{ijt}=1]=\int_{-\infty}^{\infty}\Big(\Pr[y_{ijt}=1|\ \omega_i,Z_{it}=1]\Pr[Z_{it}=1]\\ +\Pr[y_{ijt}=1|\ \omega_i,Z_{it}=0]\Pr[Z_{it}=0]\Big)\phi(\omega_i;0,\Sigma_\omega)d\omega_i,\end{aligned} \tag{7.15}$$

For the case where differences in base preferences are captured by a discrete distribution similar equations hold, where the integral in the equations above is replaced by a sum over the segments. Furthermore, the normal density function is replaced by segment probabilities, see Fok *et al.* (2001) for a complete presentation of the responsiveness model with discrete preference heterogeneity.

7.3 Inference

In this section we discuss inference in the responsiveness model with continuous preference heterogeneity. Section 7.3.1 deals with parameter estimation. In Section 7.3.2 we consider conditional inference on the latent variables in our model.

7.3.1 Parameter estimation

To estimate the model parameters, we use the Maximum Likelihood method. Hence, to obtain parameter estimates we numerically maximize the log likelihood over the parameter

space. The log likelihood function of our model (7.4), (7.8) and (7.9) is given by $\ln \mathcal{L} = \sum_{i=1}^{I} \ln \mathcal{L}_i$, where $\mathcal{L}_i$ equals the likelihood contribution of household i, that is,

$$\mathcal{L}_i = \int_{\omega_i} \mathcal{L}_i(\omega_i)\phi(\omega_i; 0, \Sigma_\omega)d\omega_i, \tag{7.16}$$

and $\mathcal{L}_i(\omega_i)$ denotes the likelihood contribution conditional on the household's base preference ω_i, that is,

$$\mathcal{L}_i(\omega_i) = \prod_{t=1}^{T_i}\prod_{j=1}^{J} \Pr[y_{ijt} = 1|\ \omega_i]^{y_{ijt}}. \tag{7.17}$$

where $\Pr[y_{ijt} = 1|\ \omega_i]$ is given in (7.13). The I integrals in the log likelihood function, given in (7.16), cannot be evaluated analytically. Instead we will rely on simulation to evaluate the log likelihood function. The resulting estimation procedure is called Simulated Maximum Likelihood [SML], see Gourieroux and Montfort (1993); Lee (1995) and Hajivassiliou and Ruud (1994) for a discussion of SML. A practical discussion of SML in the context of choice models can be found in Train (2003).

Likelihood evaluation

The obvious way to approximate the integral in (7.16) is to compute the expectation of $\mathcal{L}_i(\omega_i)$ by sampling w_i from the distribution of base preferences, that is,

$$\tilde{\mathcal{L}}_i = \frac{1}{L}\sum_{l=1}^{L} \mathcal{L}_i(\omega_i^{(l)}), \tag{7.18}$$

where L denotes the number of simulation draws used and $\omega_i^{(l)},\ l = 1, \ldots, L$ is a draw from $N(0, \Sigma_\omega)$. However the variance matrix Σ_ω is unknown and also has to be estimated. Hence the simulation method in (7.18) would require new draws for every value of Σ_ω. This complicates the convergence of the numerical optimization algorithm which is needed to find the maximum of the log likelihood function. Therefore, one uses transformations of draws from the standard normal distribution, that is,

$$\tilde{\mathcal{L}}_i = \frac{1}{L}\sum_{l=1}^{L} \mathcal{L}_i(\Sigma_\omega^{1/2}\eta_i^{(l)}), \tag{7.19}$$

where $\eta_i^{(l)} \sim N(0, \mathbf{I})$ and where $\Sigma_\omega^{1/2}$ denotes the Choleski decomposition of Σ_ω such that $\Sigma_\omega^{1/2}\eta_i^{(l)} \sim N(0, \Sigma_\omega)$. With this simulation scheme we can rely on one set of draws to calculate the likelihood function for all parameter values.

Variance reduction by Importance Sampling

A disadvantage of using (7.19) or (7.18), is that often many draws are necessary to obtain a precise evaluation of the likelihood contribution. In other words the likelihood contribution estimates often have a large variance for moderate L. In these cases many draws are necessary, and the numerical optimization of the simulated log likelihood may become very computer intensive. To reduce the variance of the sampler we propose to use Importance Sampling, see Kloek and van Dijk (1978) and Geweke (1989a). To this end we introduce the so-called importance function $g(\eta_i; m_i, S_i)$. For our model we choose a normal density function with mean m_i and variance S_i as the importance function. We rewrite the likelihood contribution (7.16) as

$$\mathcal{L}_i = \int_{\eta_i} \frac{\mathcal{L}_i(\Sigma_\omega^{1/2}\eta_i)\phi(\eta_i; 0, \mathbf{I})}{g(\eta_i; m_i, S_i)} g(\eta_i; m_i, S_i) d\eta_i. \tag{7.20}$$

To approximate the likelihood we use

$$\tilde{\mathcal{L}}_i = \frac{1}{L}\sum_{l=1}^{L} \frac{\mathcal{L}_i(\Sigma_\omega^{1/2}\eta_i^{(l)})\phi(\eta_i^{(l)}; 0, \mathbf{I})}{g(\eta_i^{(l)}; m_i, S_i)}, \tag{7.21}$$

where $\eta_i^{(l)}$ is a draw from $g(\eta_i; m_i, S_i)$. Note that the same strategy is used to compute marginal likelihood functions by Geweke (1989b), see also Neal (1994) in the discussion of Newton and Raftery (1994).

To reduce the variance of the simulator we choose the importance function such that it has most of its probability mass in the range of values of η_i where the likelihood conditional on η_i is large, that is, where $\mathcal{L}_i(\Sigma_\omega^{1/2}\eta_i)\phi(\eta_i; 0, \mathbf{I})$ is relatively large. If the importance function is chosen to be a density, the optimal case would be to have the importance function proportional to $\mathcal{L}_i(\Sigma_\omega^{1/2}\eta_i)\phi(\eta_i; 0, \mathbf{I})$. For this importance function the quotient in (7.21) would not depend on the specific draws and it would give the likelihood contribution without simulation error. In practical settings such an importance function is impossible to find. The task therefore is to set the values of m_i and S_i such that the importance function resembles this likelihood component. To find these values we use an iterative scheme. First we set $m_i = 0$ and $S_i = \mathbf{I}$, and simulate $\eta_i^{(l)}$ from $g(\eta_i; m_i, S_i)$. Next, we calculate the importance weights

$$w_i^{(l)} = \frac{\mathcal{L}_i(\Sigma_\omega^{1/2}\eta_i^{(l)})\phi(\eta_i^{(l)}; 0, \mathbf{I})}{g(\eta_i^{(l)}; m_i, S_i)}. \tag{7.22}$$

Using the importance weights $w_i^{(l)}$ we can now update m_i and S_i to find a closer match between the importance function and the likelihood. The match is improved by computing

the weighted mean and variance of $\eta_i^{(l)}$

$$\begin{aligned} m_i &= \frac{\sum_{l=1}^{L} \eta_i^{(l)} w_i^{(l)}}{\sum_{l=1}^{L} w_i^{(l)}} \\ S_i &= \frac{\sum_{l=1}^{L} (\eta_i^{(l)} - m_i)(\eta_i^{(l)} - m_i)' w_i^{(l)}}{\sum_{l=1}^{L} w_i^{(l)}}. \end{aligned} \tag{7.23}$$

This strategy can be iterated to find even better values of m_i and S_i. The resulting location and scale for the importance function can then be used to calculate the log likelihood function, which in turn is optimized over the model parameters. Note that the optimal m_i and S_i will depend on the model parameters. Therefore for the best performance the location and scale parameters of the importance function will have to be updated a few times during the optimization of the likelihood. Geweke (1989a) shows that $\sqrt{L}(\tilde{\mathcal{L}}_i - \mathcal{L}_i) \Rightarrow N(0, \nu)$, where ν can be estimated by the sample variance of the importance weights. This result allows us to develop an easy rule to decide when the location and scale of the importance function have to be updated. For example, one could update the location parameters m_i and S_i if the sampling variance exceeds a specified threshold value.

One of the requirements for successful use of Importance Sampling is that the importance function most be wide enough to sample all important regions of the conditional likelihood. The location of the important areas of the likelihood function will depend on the parameter values, during the optimization the parameter values will obviously change. The method to calculate the location and scale parameters discussed above however depends on a single choice of the model parameters. To prevent undersampling of important regions when the parameter values have changed we increase the variance of the importance function using a multiplication factor. In our setting we choose to increase the variance by a factor $\kappa > 1$.

Asymptotic distribution of parameter estimator

Under the usual regularity conditions, the SML estimator is consistent for $I \to \infty$ and $L \to \infty$, see Hajivassiliou and Ruud (1994). Furthermore, the SML estimator is asymptotically efficient. The SML estimator is asymptotically normal distributed, with mean the true value of the parameters and covariance matrix the inverse of the information matrix. However, for finite L the information matrix of the simulated log likelihood will underestimate the true covariance matrix due to simulation noise. Therefore, McFadden and Train (2000) recommend to use the sandwich estimator (Newey and McFadden, 1994) to compute standard errors. This estimator produces more reliable estimates of the

standard error and is given by

$$\widehat{\mathrm{Var}}(\hat{\theta}) = \left(\frac{\partial^2 \ln \tilde{\mathcal{L}}}{\partial\theta\partial\theta'}\right)^{-1'} \left[\sum_{i=1}^{I} \left(\frac{\partial \ln \tilde{\mathcal{L}}_i}{\partial\theta}\right)' \left(\frac{\partial \ln \tilde{\mathcal{L}}_i}{\partial\theta}\right)\right] \left(\frac{\partial^2 \ln \tilde{\mathcal{L}}}{\partial\theta\partial\theta'}\right)^{-1}, \tag{7.24}$$

where θ denotes a vector containing all parameters and where all the derivatives are evaluated at the parameter estimates $\hat{\theta}$.

Performance comparison

In this subsection we illustrate the advantage of using importance sampling (7.21) over direct sampling (7.19) to calculate the log likelihood function. The advantage of Importance Sampling is that the variance of the likelihood approximation is reduced to a large extent, thereby reducing the need for a large number of simulation draws. That is, for a specific set of draws the likelihood evaluation using importance sampling will be closer to the true value than when using direct sampling.

To show how the simulator variance depends on the number of draws and the simulation method, we consider the mixed MNL model for purchases of detergent, see Section 7.4 for more details on the data. We evaluate the log likelihood function at the parameter estimates using both sampling methods and different number of simulation draws. The location and scale parameters used in the importance sampler are set using 1000 draws (the same as used to set these parameters in the parameter estimation in Section 7.4). Figure 7.1 shows the distribution of the sampler for 50, 100, 250, 500 and 1000 draws based on 100 log likelihood evaluations. The top graph shows the distributions for direct sampling, the middle graph shows the distributions for importance sampling on the same scale as for direct sampling. The lower graph in Figure 7.1 shows a close-up of the distributions for importance sampling. Figure 7.1 clearly shows the strong reduction in sampling variance. Even with as few as 50 draws the importance sampler has a smaller variance than direct sampling with 250 draws. Furthermore note that for direct sampling the simulation bias in the log likelihood evaluation is quite large. As we sample the likelihood contribution to calculate the log likelihood function there will be a simulation bias for finite L for all simulation methods. Clearly the simulation bias depends on the variance of sampler. Therefore the Importance Sampling approach does not only yield log likelihood evaluations with a smaller variance, the bias also tends to be smaller compared to direct sampling.

7.3.2 Conditional inference

The responsiveness model contains several latent variables, that is, the latent base preferences and the responsiveness states. To analyze the in-sample and out-of-sample behavior

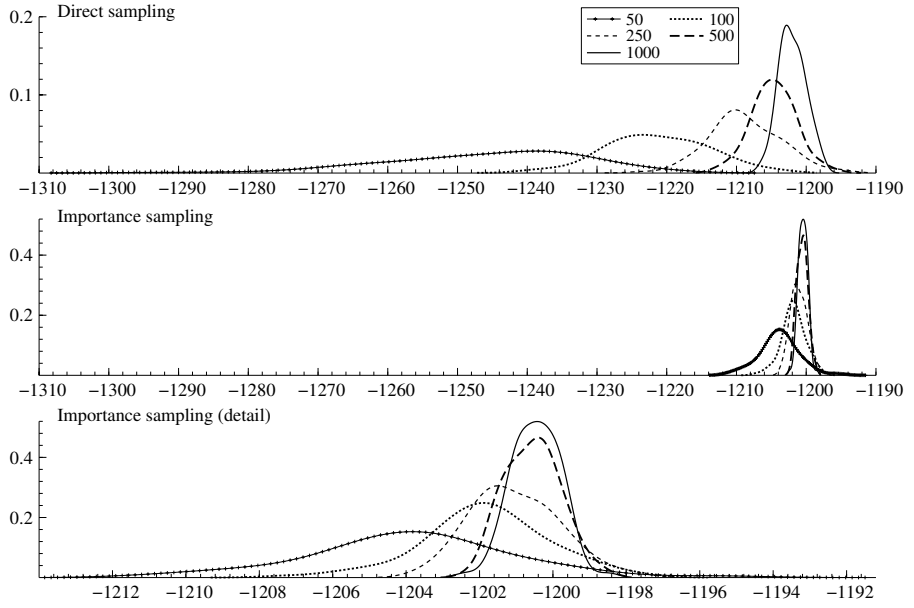

Figure 7.1: Distribution of log likelihood evaluation

of households, it is often necessary to estimate (functions of) these latent variables. For example, if we want to know whether a household was responsive to marketing instruments at a particular purchase occasion, we also need an estimate of the latent base preference of this household. Likewise, if we want to forecast out-of-sample purchase behavior we also need to estimate the latent base preferences of households.

To estimate functions of these latent base preferences, we condition on the actual purchases of the households. In general the above-mentioned problems can be translated to the calculation of the conditional mean of a function $h(\omega_i)$, that is,

$$\begin{aligned}
\mathrm{E}[h(\omega_i)|y_{i1},\ldots,y_{iT_i}] &= \int h(\omega_i) f(\omega_i|y_{i1},\ldots,y_{iT_i}) d\omega_i \\
&= \int h(\omega_i) \frac{\mathcal{L}_i(\omega_i)\phi(\omega_i;0,\Sigma_\omega)}{\int \mathcal{L}_i(\omega_i)\phi(\omega_i;0,\Sigma_\omega)d\omega_i} d\omega_i \\
&= \frac{\int h(\omega_i)\mathcal{L}_i(\omega_i)\phi(\omega_i;0,\Sigma_\omega)d\omega_i}{\int \mathcal{L}_i(\omega_i)\phi(\omega_i;0,\Sigma_\omega)d\omega_i} \\
&\approx \frac{\sum_{l=1}^{L} h(\Sigma_\omega^{1/2}\eta_i^{(l)})u_i^{(l)}}{\sum_{l=1}^{L} u_i^{(l)}},
\end{aligned} \tag{7.25}$$

where $f(\omega_i|y_{i1},\ldots,y_{iT_i})$ denotes the conditional density function of ω_i given the purchase decisions $\{y_{i1},\ldots,y_{iT_i}\}$, where $\eta_i^{(l)}$ are draws from $g(\eta;m_i,S_i)$, and where $u_i^{(l)} = \mathcal{L}_i(\Sigma_\omega^{1/2}\eta_i^{(l)})\phi(\eta_i^{(l)};0,\mathbf{I})/g(\eta_i^{(l)};m_i,S_i)$ with the location and scale (m_i and S_i) are as defined earlier.

Given (7.25) we can compute several quantities of interest. To obtain an estimate of the base preferences of household i we set $h(\omega_i)=\omega_i$. The posterior probability of being in the responsive state at purchase occasion t equals

$$\mathrm{E}[Z_{it}|y_{i1},\ldots,y_{iT_i}] = \int_{\omega_i} \Pr[Z_{it}=1|y_{i1},\ldots,y_{iT_i},\omega_i] f(\omega_i|y_{i1},\ldots,y_{iT_i})d\omega_i. \tag{7.26}$$

To calculate this expectation we therefore set

$$\begin{aligned} h(\omega_i) &= \Pr[Z_{it}=1|y_{i1},\ldots,y_{iT_i},\omega_i] \\ &= \Pr[Z_{it}=1|y_{it},\omega_i] \\ &= \frac{\Pr[y_{it}|Z_{it}=1,\omega_i]\Pr[Z_{it}=1]}{\sum_{z=0}^{1}\Pr[y_{it}|Z_{it}=z,\omega_i]\Pr[Z_{it}=z]}. \end{aligned} \tag{7.27}$$

Finally, if we are interested in out-of-sample forecasts of brand choice conditional on the observed in-sample choices made by a household, that is, $\Pr[y_{ij,T_i+1}|\ y_{i1},\ldots,y_{iT_i}]$, we have to set $h(\omega_i)=\sum_{z=0}^{1}\Pr[y_{ij,T_i+1}|\ \omega_i, Z_{i,T_i+1}=z]\Pr[Z_{i,T_i+1}=z]$, where these last two probabilities are given in (7.14) and (7.4), respectively.

7.4 Illustration

We apply our model to a data base containing liquid detergent purchases in Sioux Falls, South Dakota, during the period July 1986 – July 1988. The sample contains 400 households making 2,657 purchases. The last observed purchase of each household is used as a hold-out sample for model comparison and evaluation. All other recorded purchases are used for estimation. The same data are analyzed in Chintagunta and Prasad (1998) for other purposes. The households in the panel are selected to only purchase the top six national brands, Tide, Eraplus, Solo, Wisk, All and Surf. For each purchase occasion we know the time since the last liquid detergent purchase, the volume last purchased (in multiples of 32 oz.), the shelf prices of the six alternatives and which brands are featured or on display. Table 7.1 gives a brief overview of the number of purchases and the use of marketing instruments in this market. In our sample the most popular brands are Tide and Wisk, and these two brands are also most often featured and on display. The two smallest brands in choice share are Solo and All, which are rarely featured. Next to these variables, we know the chosen brand and the expenditures on non-detergent products made on the same shopping trip. On average, households shop for detergent every 80

Table 7.1: Data properties

Brand	Brand characteristics			
	Number of purchases	Avg. price per 32 oz.	Feature[1]	Display[1]
Tide	701	1.926	9.82%	9.17%
Wisk	703	1.510	14.57%	10.11%
EraPlus	507	1.952	3.34%	2.62%
Surf	406	1.702	4.65%	3.73%
Solo	253	1.901	1.90%	1.44%
All	87	1.261	0.07%	0.07%

[1] Percentage of times over all observed purchase occasions that the product was featured or on display.

days, purchase 77 oz. of detergent and spend almost \$40 per shopping trip. The average household size in our sample is 2.8. A preliminary analysis of the data using simple versions of our model, as well as using the MNL model, indicated that display does not have a relevant effect on explaining brand choice, and hence this variable will be discarded.

7.4.1 Preference heterogeneity

First, we compare the performance of the responsiveness model with and without modeling unobserved preference heterogeneity. In Table 7.2 we present the empirical fit of the responsiveness model for four cases. The first column of Table 7.2 gives the fit for the responsiveness model without unobserved heterogeneity. The second column gives the performance of the model where preference heterogeneity is captured by a normal distribution. The final two columns give the results for the case where a discrete distribution is used to model the base preferences. The model with the continuous preference distribution has by far the highest in-sample log likelihood. If out-of-sample forecasts are made conditional on the observed in-sample choices this model also performs much better than the other models. However, for unconditional forecasting this model does not perform that well, as can be seen by the in-sample hit rate and the unconditional out-of-sample hit rate. The responsiveness model without unobserved preference heterogeneity also performs rather well. Although it has the lowest in-sample likelihood, it has the best unconditional out-of-sample performance. The performance of the models with discrete

preference heterogeneity is somewhere in between the basic responsiveness model and the model with continuous heterogeneity.

Table 7.2: Goodness of fit measures for the responsiveness model for different preference heterogeneity specifications

	Homogeneous preferences	Heterogeneous preferences		
		continuous	2 segments	3 segments
Number of parameters	19	34	25	31
$\log \mathcal{L}$	-1403	-1189	-1338	-1308
Predicted $\log \mathcal{L}$[1]	-192.6	-267.6	-202.5	-207.3
Hit rate in-sample	77.61	70.95	76.68	75.34
Hit rate out-of-sample				
conditional[2]	77.31	80.00	75.39	75.39
unconditional	77.31	67.31	76.15	74.62

[1] Log likelihood value for out-of-sample data

[2] Out-of-sample forecasts made conditional on the observed in-sample brand choices.

The choice for the optimal model is difficult to make. If one is interested in making conditional forecasts one will prefer the model specification with continuous preference heterogeneity. For unconditional forecasts the model with homogeneous preferences is to be preferred. This specification also uses substantially less parameters. For completeness we will present below the estimation results for the homogeneous and the continuous preference heterogeneity specifications.

7.4.2 Estimation results

Table 7.3 shows the estimation results for the model equation concerning responsiveness, see (7.4), for the model without heterogeneity and for the model with the continuous form of preference heterogeneity. Note that the household characteristics and the marketing efforts are all normalized to have mean 0 and variance equal to 1. The main difference between the two specifications is in the intercepts, measuring the baseline responsiveness. For the model with preference heterogeneity we find a much larger intercept, indicating that under this specification households tend to be more responsive to marketing efforts. Another difference is in the estimated effect of the household size. Under the homogeneous specification there is no significant effect of this household characteristic. For the

Table 7.3: Parameter estimates in the responsiveness model (equation (7.4)), standard errors in parentheses

	Base preferences			
	Homogeneous		Heterogeneous	
Intercept ($\mu^{(z)}$)	0.60	(0.18)	2.70	(0.42)
Non-Detergent expenditures	−0.21	(0.09)	−0.47	(0.22)
Household size	−0.07	(0.07)	−1.03	(0.48)
Time since last purchase	0.65	(0.18)	0.99	(0.55)
Volume previously purchased	−0.47	(0.09)	−0.42	(0.36)

alternative specification the household size is negatively correlated with responsiveness. The variables related to the shopping trip also influence the responsiveness. For both specifications we find similar effects. The time since the last purchase is positively correlated with the responsiveness. If a household has not purchased liquid detergent for a long period relative to a more frequent shopper, it might spend more time thinking about the next purchase and this may increase the responsiveness to marketing efforts. The reverse holds for the volume of liquid detergent purchased previously. Households purchasing large quantities may have a strong preference for one of the brands, and therefore could be less responsive to marketing efforts of other brands. As expected, the total expenditures on the shopping trip spent on products other than detergents negatively correlates with responsiveness. If a household plans to purchase many items on a single shopping trip, it cannot spend much time making a selection in every single category. Therefore, the household will tend to base its decisions more on habit and base preference.

Next we elaborate upon the inferred responsiveness for a household at a specific purchase occasion conditional on the observed brand choices, that is $\mathrm{E}[Z_{it}|\{y_{il}\}_{l=1}^{T_i}]$. Figure 7.2 gives the distribution over all observed choices of this conditional expectation. From this figure we conclude that at a large proportion of the purchase occasions the household was responsive to marketing efforts. For the heterogeneous preferences specification this proportion is even larger. Especially for the homogeneous case there is quite a large fraction of purchase occasions where the household was not so much responsive to marketing efforts. In fact, the average expected responsiveness equals 0.61, the empirical standard deviation of the expected responsiveness across all purchase occasions equals 0.26. For the heterogenous specification the average responsiveness equals 0.86 with a standard deviation of 0.19.

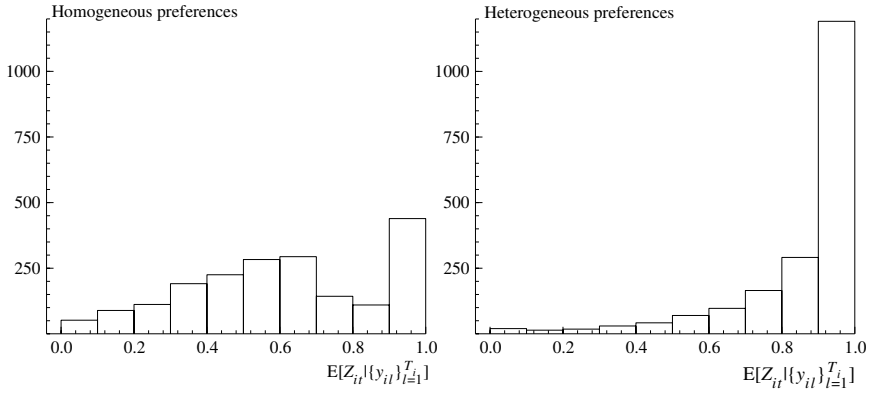

Figure 7.2: Empirical distribution of conditional expectation of responsiveness

To evaluate whether households switch responsiveness segments, we use the conditional expected responsiveness to assign households to one of the two segments for every purchase occasion. If the conditional responsiveness probability exceeds 0.5, the household is, at that purchase occasion, assigned to the responsive segment. An analysis of the resulting responsiveness assignments indicates that there are substantial differences in responsiveness across purchase occasions within the same household. For the specification with homogeneous preferences there is the most switching between responsiveness states. Based on this specification, households switch responsiveness states on average in 25.8% of the cases (with a standard deviation of 26.2%). If the segment membership of a household is tracked over time, we see that of the 210 households that purchased detergent more than once, 131 households switched from unresponsive to responsive at least once. For the heterogeneous specification there is switching in 7.6% of the cases (standard deviation 18.6%).

Table 7.4 presents the parameter estimates of the multinomial logit models representing the brand choice decisions made by households being either responsive or unresponsive to marketing instruments. Again we present the results for both specifications of preference heterogeneity.

All estimated effects of marketing efforts on purchase probabilities have the expected sign, with price having a negative effect on brand choice and feature a positive. The effect of price is clearly the largest. An interesting observation is that, as expected, households which are unresponsive to marketing efforts act more state dependent than responsive households. Choices made in the unresponsive state are more driven by habit and state dependence than in the responsive case. Concerning these parameter estimates there are a few differences across the two heterogeneity specifications. As expected, we find less state

Table 7.4: Parameter estimates for models (7.8) and (7.9) under homogeneous and heterogeneous preferences (standard errors in parentheses)

	Homogeneous preferences				Heterogeneous preferences			
	Responsive		Unresponsive		Responsive		Unresponsive	
				Marketing efforts/state dependence				
Price	−1.35	(0.09)	–	–	−1.65	(0.14)	–	–
Feature	0.45	(0.05)	–	–	0.50	(0.09)	–	–
Lagged choice	1.94	(0.13)	4.84	(0.51)	1.15	(0.22)	2.11	(0.89)
				Brand intercepts				
Tide	0.34	(0.13)	0.01	(0.52)	1.22	(0.38)	0.87	(1.41)
Wisk	0.37	(0.15)	−0.49	(0.76)	1.04	(0.42)	−0.15	(1.65)
Surf	−0.04	(0.15)	−2.58	(0.73)	0.38	(0.43)	0.93	(1.63)
Solo	−0.69	(0.19)	−0.67	(0.75)	−3.23	(1.14)	−7.98	(2.92)
All	−2.27	(0.37)	−1.14	(0.87)	−6.61	(2.64)	−7.68	(5.17)
EraPlus	0*	–	0*	–	0*	–	0*	–

* Restricted for identification

dependence for the specification with heterogeneous preferences. As discussed in Section 7.2.2, ignored unobserved preference heterogeneity in general leads to overestimated state dependence, and this is what we see here too.

Although the differences across the two specifications in the estimates of the brand intercepts in Table 7.4 may seem very large, there are some striking similarities. Households in the unresponsive state do not consider the brands to be very different. This is reflected in the fact that there are almost no significant brand intercepts in the MNL models for the unresponsive decision process. For the responsive state we do find significant brand intercepts, indicating that households in this state do consider the brands to be different. For the heterogeneous specification we see that a household with the average preferences in the unresponsive state has a very low preference for the brands "Solo" and "All". In fact the corresponding probability of choosing one of these brands is almost zero for such a household. However, the estimated variance across the households of the base preference for these two brands is rather large. Although the average unresponsive household will not choose one of these brands, there are households with different base preferences who do consider them.

Summarizing, we find for this data set that many households are responsive to marketing efforts, that households purchasing large amounts are less responsive to marketing efforts, that households who do not regularly purchase detergent are more responsive, and that unresponsive households behave more state dependent than responsive households do. Finally, we find that modeling unobserved preference heterogeneity only leads to a substantial improvement in the forecasting performance when the forecasts are made conditional on the observed in-sample brand choices.

7.4.3 Competing models

The results above show that the segmentation of purchase occasions in cases where households are responsive or unresponsive clearly separates the purchase occasions where the household acts state dependent from the cases where the household responds to marketing efforts. The other parameter estimates also seem to have a high face validity. All this together leads us to believe that the concept of responsiveness seems to be empirically useful. Of course, the ultimate test for a newly proposed model is whether it fits the data better and generates better forecasts compared to other models. In choice modeling, the MNL model seems to be the standard. In this section we will therefore compare our model to various forms of the MNL model. In the MNL models, all variables used in the responsiveness model are included, including those that are used to model the probability of being responsive and the brand choice on the previous purchase occasion. We again consider various forms of heterogeneity. We use an MNL model without incorporating unobserved heterogeneity, one where the utility intercepts follow a discrete distribution, and finally one where these intercepts are captured by a normal distribution.

Table 7.5 shows the model performance criteria for the various MNL models as well as for the responsiveness model. The in-sample forecasting results suggest that the homogeneous responsiveness model outperforms all MNL models. On out-of-sample forecasting, only the MNL model where the preference heterogeneity is modeled with a continuous distribution beats this model when the forecasts are made conditional on the in-sample brand choices. However the heterogeneous responsiveness model provides the best conditional forecasts. The homogeneous responsiveness model performs best on unconditional forecasting. Overall our model seems to perform much better than the MNL models, notably with less parameters.

Additionally, we also compare the responsiveness model to an MNL model where, next to the brand intercepts, other parameters are also allowed to differ in the population. We consider three additional MNL models. All these models are based on the mixture approach of modeling heterogeneity. In the first model we allow the price and feature parameter to differ across the population. In another model we model the state dependence parameter with a mixture distribution. Finally, we have estimated the MNL

Table 7.5: Comparison of the new model versus various forms of the MNL model

	Homogeneous preferences		Heterogeneous preferences			
	Responsiveness	MNL	Responsiveness continuous	MNL continuous	MNL mixture 2 segments	MNL mixture 3 segments
No. parameters	19	28	34	43	34	40
$\log \mathcal{L}$	-1403	-1429	-1189	-1199	-1349	-1306
predicted $\log \mathcal{L}$	-192.6	-213.1	-267.6	-269.6	-215.8	-217.6
Hit rate						
in-sample	77.61	76.73	70.95	72.60	75.08	73.27
out-of-sample						
conditional*	77.31	74.23	80.00	78.08	74.62	74.23
unconditional	77.31	74.23	67.31	62.31	73.47	71.92

* Out-of-sample forecasts made conditional on the observed in-sample brand choices.

model for the case where the brand intercepts and the price and feature parameters differ across the households. To save space we do not report the performance measures of these models, they can be found in Fok *et al.* (2001). Not surprisingly, these alternative models outperform our homogeneous responsiveness model on in-sample log likelihood, as they use many more parameters. None of these alternatives gives a higher in-sample likelihood than the heterogeneous responsiveness model. With one exception, the out-of-sample likelihood, and the in-sample and out-of-sample hit rate are not better than the homogeneous responsiveness model. The conditional hit rate for the MNL model where both the brand intercepts and the marketing response parameters are modeled by a mixture distribution with three segments is slightly better than that for the homogeneous responsiveness model (77.69% versus 77.31%). Again, when the out-of-sample forecasts are not made conditional on the observed brand choices, the alternative model does not forecast better (hit rate 73.85%). Altogether, these results seem to indicate that the concept of responsiveness apparently captures an important part of the heterogeneity.

It is also interesting to compare the parameter estimates of some of the models. In Table 7.6, we compare parameter estimates for the responsiveness model with and without preference heterogeneity and the MNL model with two segments where the intercepts, the marketing sensitivity and the state dependence parameter are allowed to differ. An interesting observation for the 2-segment MNL model is that the second segment partly captures the unresponsive households, as sensitivity to marketing efforts is quite low

Table 7.6: Comparison of effectiveness of marketing instruments responsiveness versus heterogeneous MNL(standard errors in parentheses)

	Responsiveness model				2-segment MNL	
	Homogeneous		Heterogeneous			
	Resp.*	Unresp.	Resp.	Unresp.	Segment 1	Segment 2
Price	-1.35	-	-1.65	-	-1.25	-0.67
	(0.09)	-	(0.13)	-	(0.07)	(0.10)
Feature	0.45	-	0.50	-	0.43	0.21
	(0.05)	-	(0.09)	-	(0.05)	(0.06)
Lagged choice	1.94	4.84	1.15	2.11	1.37	3.80
	(0.13)	(0.51)	(0.22)	(0.89)	(0.12)	(0.15)

* Resp.: Responsive to marketing efforts
Unresp.: Unresponsive to marketing efforts

compared to the first segment whereas there is a high degree of state dependence. The main difference between this model and our model is of course that in the responsiveness model the segment membership is correlated to household characteristics and also that households are allowed to switch between the segments over time.

7.5 Concluding remarks

Households might not respond to marketing-mix instruments at each purchase occasion. To be able to respond to these efforts, one needs to invest time and effort in, for example, remembering price changes and reading newspapers and leaflets to notice advertisements. Households differ in the amount of effort they wish to invest in a particular purchase, and therefore they will most likely also differ in their responsiveness to marketing efforts.

The choice model we developed in this chapter incorporates the responsiveness of a household at a specific purchase occasion as a form of structural heterogeneity. Households differ in their decision-making process. In essence, we assume there are two decision processes. Households either take marketing efforts into account or they base their choice on base preference and their past experience. The specific decision process used can differ across households and across purchase occasions. To explain and forecast the decision process, used by a specific household at a specific purchase occasion, household character-

istics can be used together with information on buying behavior. To take into account this form of structural heterogeneity, we extended the MNL model. Basically, we introduce two segments of households, one segment is unresponsive to marketing efforts whereas the other segment does respond to these efforts. The segment membership is separately modeled using a binary logit model. This "responsiveness model" can be further extended to also include preference heterogeneity. For the estimation of the models with preference heterogeneity we have developed a rather efficient sampling method to approximate the likelihood. This sampling method, which is based on importance sampling, gives more accurate likelihood evaluations compared to direct sampling.

The main behavioral conclusions from the application of our model to a six-brand market of liquid detergent are that most households are responsive to marketing efforts, and that large basket shoppers and large volume shoppers tend to be less responsive. Infrequent shoppers however are more responsive. Further, an unresponsive household seems to act more state dependent.

Finally, we compared the in- and out-of-sample performance of our model to various forms of the MNL model where we also corrected for heterogeneity. From this comparison we conclude that, while using the smallest amount of parameters, our model outperforms all MNL variants on forecasting. This, together with the face validity of our parameter results, leads us to believe that incorporating responsiveness seems to be a worthwhile exercise.

Chapter 8

Summary and conclusion

In this thesis we have discussed a wide variety of marketing models. In the first part we focused on aggregate market response models. The central model in this part was the market share attraction model. The second part of this thesis dealt with household level models. We considered modeling interpurchase timing and brand choice.

In the literature, there are more or less standard models available for each marketing measure we have considered. However, as implicit in the definition of a model, they are never perfect. In this thesis we worked on improving the performance of some of the existing models. Some chapters in this thesis dealt with improvements with respect to the econometric properties of the model and the corresponding estimation and forecasting methods. In other chapters we focused on incorporating relevant knowledge from the marketing literature in the models to make them more realistic. For proper inference in the suggested models, it turned out that advanced econometric techniques are necessary. Each chapter therefore had quite a strong econometric orientation.

Below we provide a summary of this thesis, where the focus is on the main contributions, the empirical results and topics for further research.

8.1 Summary

In Chapter 2 we have presented an extensive econometric analysis of the well-known market share attraction model. One of the main findings in this chapter is that routinely made forecasts are biased. Remarkably, in the literature there is no research available in which the forecasting method is explicitly discussed. We showed that unbiased forecasts can be obtained using simulation techniques. Another contribution of this chapter is a formal model selection strategy to guide the choice of an appropriate attraction specification. In practice one often selects an attraction specification without formal testing.

We found that the use of a model selection strategy leads to an empirical model with improved in-sample fit and out-of-sample performance.

A third issue we have discussed is parameter estimation. We have shown that the commonly used reduced-form specification of the attraction model is more complicated than necessary. In this method, which is known as the log-centering approach, the reduced-form model is specified in terms of the log of market shares relative to the geometrically averaged market share. Due to the logical consistency feature of the market share attraction model, the reduced-form model equations are linearly dependent. This, and the fact that the covariance matrix of the reduced-form error terms is singular leads to complications in the estimation procedure and parameter inference. An alternative reduced-form specification is specified in terms of log market shares relative to a base brand. We have formally shown that both reduced-form specifications lead to equivalent maximum likelihood estimates. However, parameter estimates are much easier to obtain if the base brand approach is used.

Chapter 3 discussed the effects of the introduction of a new brand on an existing market. Many studies in the literature present a normative view on this issue. The focus is then on deriving the optimal reaction to entry. However, these optimal reactions are usually not confronted with empirical data. Although changes in (aggregate) consumer behavior due to the entry may have a substantial influence on the optimal strategy to use, they are usually not considered in these studies.

We focused on developing tools to test for changes in empirical data. To this end, we have proposed statistical tests to assess whether the competitive structure has changed in an empirical market, where the competitive structure is summarized by the brand preferences of the consumers and their reactions to marketing instruments. Furthermore, we have proposed statistical tests to analyze whether the incumbent brands have changed the use of their marketing mix as a reaction to the entry. For this analysis we have again chosen for market shares as the focal marketing measure.

Changes in the competitive structure are reflected in observable changes in the market shares. Some of the changes in the market shares may however be due to the marketing instruments of the new brand or due to (efficient) marketing strategies of the incumbent brands. Therefore, just testing for changes in the market shares is not sufficient. Instead, one has to correct for the marketing instruments of the entrant and the existing brands. To this end, we have extended the market share attraction model to deal with a changing number of brands.

The testing methodology was illustrated using data on market shares of detergents. For this market we have found fewer competitive reactions than predicted by the normative studies. We did find changes in the consumer behavior. After the introduction, households react significantly different to price and coupon promotions. The sensitivity

to display and feature promotions remained unchanged. Furthermore, we have tested for changes in the relative positioning of the incumbent brands. Both the relative strength of the brands (measured by the brand intercepts) and the brand similarities (measured by the covariances of the unobserved attractions) changed. These changes in competitive structure may explain why the normative predictions do not hold in this example. Another explanation may be that brand managers did not respond optimally to the entry.

There are at least two promising paths for further research in this area. First of all, the normative studies may be extended to allow for changes in consumer behavior. The main challenge will be in the formalization of these reactions. Obviously the consumer's reaction will depend on the entrant's strategy and the following competitive reactions by the incumbent brands. Another fruitful avenue of research may focus on empirical generalizations of the effects of introductions. This will however depend on the availability of sufficient data.

In Chapter 4 we have studied dynamic effects in market shares. In the context of the market share attraction model, dynamic effects can easily be captured by adding lagged marketing instruments and lagged market shares as explanatory variables to the attraction specification. However, such a model does not easily provide insight into relevant dynamic properties of the market at hand. We presented an alternative specification of the dynamic market share model in which the parameters directly represent interesting dynamic features.

The direct and the permanent effect of a temporary promotion are often considered as interesting dynamic features. Recent literature has shown that most marketing series, including market shares, are stationary, which implies that a temporary promotion will not have a permanent effect on market shares. For most market share series the permanent effect will therefore equal zero. As an alternative measure of the long-run effects of marketing instruments we suggested the cumulative effect of a temporary promotion. Summarizing, the dynamic features we consider to be relevant are the direct effect and the cumulative effect of a promotion. We labeled the direct effect as the short-run effect and the cumulative effect as the long-run effect. To allow for direct identification of the short-run and the long-run effects of promotions, we have rewritten the market share model in so-called error-correction format.

Interesting practical questions concern the difference between the short-run and the long-run effect and the sign of the long-run effect itself. In the weeks after a promotion one often observes a decrease in market shares. In case of such a post-promotional dip, the long-run effect of a promotion will be smaller than the short-run effect. The long-run effect can even be negative if the increase in market share at the promotional week is more than compensated by a decrease in the weeks after.

Next to deriving a useful econometric model to analyze the dynamic features of market shares, we studied the extent to which differences in the dynamics across brands and product categories can be explained by observable characteristics. The Hierarchical Bayes framework turned out to be very useful for this purpose. The first layer of the model corresponds to separate attraction models, one for each market. In the second layer of the hierarchical model, the parameters representing relevant dynamic properties are related to brand and market characteristics.

We have applied our model to a database consisting of seven different product categories in two distinct geographical regions. Across these 14 markets we found that the long-run effects of feature and display promotions tend to be larger than the corresponding short-run effects. In other words, display and feature promotions in general have positive carry-over effects. For price we found the opposite. We found substantive interaction between brand characteristics and the price effects. In general, higher priced brands and brands that more frequently issue coupons have stronger price effects. This holds for the short-run as well as for the long-run effect. The degree of market concentration affects the short-run price effect, a higher market concentration leads to stronger price effects. The short-run effect of coupons is correlated to the relative price and the market concentration. Coupons of higher priced brands or of brands in markets with low concentration are less effective.

Future research in this area may involve our model as input for a game-theoretic study. In our model we have estimated the marketing effectiveness while correcting for possible competitive reactions. The resulting parameter estimates can be used to assess the optimal dynamic reaction to a competitive marketing action. The derivation of the optimal response may be obtained analytically or by evaluating various scenarios.

Our analysis may be extended by analyzing the markets at a lower level of aggregation. In our model we analyzed differences across categories and across different brands within a category. Additionally, one could study differences across stores within a region or differences across the product items that make up a brand. As the typical market may contain a very large number of items, the advantages of the Bayesian approach become more pronounced at this level of aggregation. At this level of aggregation, the dynamic features may be explained by item characteristics such as the weight, packaging and flavor of an item. In van Nierop *et al.* (2002) we have made a first attempt to such an analysis. However, in this paper we did not consider disentangling the short-run and the long-run effects.

Chapters 5 and 6 both discussed models for the purchase timing decision of households. In Chapter 5 we took on quite a technical perspective, and studied the optimal way to include explanatory variables. As the purchase timing process is defined on the category level, explanatory variables in an econometric model should also be defined on this level.

Marketing-mix variables should therefore somehow be aggregated. In the literature a number of ad-hoc methods are available. In Chapter 5, we proposed three alternative methods. We furthermore have studied the relative merits of all these methods.

The most popular method in the literature seems to be to use a weighted average of the marketing mix across the available brands, where the weights are household-specific choice shares. The main disadvantage of this approach is that it is not very suitable for out-of-sample forecasting. The advantage of this method is that differences in brand preferences across households are captured without having to explicitly model them. Changes in the preferences of households, for example induced by promotional activities, are not accounted for. Another popular approach is to summarize the marketing efforts by the so-called inclusive value. Our results indicated that this method is quite restrictive and performs poorly on in-sample and out-of-sample fit. The three alternative methods we have introduced are based on using weights obtained from a choice model to weigh the marketing instruments. The most elaborate model we have discussed actually integrates the brand choice model with the purchase timing model by considering the preferences of households to be latent variables in the weeks where no purchase is made.

We have compared the empirical performance of the different methods using data on three product categories. The main finding from our comparisons is that when unobserved heterogeneity in brand preferences and in purchase timing is not explicitly accounted for the popular method based on choice shares has the best in-sample performance. Our "latent preference" model tends to perform best for out-of-sample forecasting. This model also performs best in-sample when unobserved heterogeneity is captured by the model. The practical conclusion from this chapter is that in case one is not so much interested in forecasting, the approach based on choice shares is sufficient. However, if the object of the study is forecasting, one of the alternatives should be used. Although the differences between the various models are sometimes small, the latent preference model performs best.

Future research in this area may involve extending the model for the purchase planning process of a household to include the choice of the store in which the purchases are made. The difficulties involved with the store choice decision in a purchase timing model closely resemble those concerning brand choice. The researcher somehow has to aggregate the marketing efforts in each store to obtain, for example, an overall index of price. By integrating a model for store choice and a purchase timing model one may improve upon the popular method where marketing instruments are averaged over stores with household-specific weights.

In Chapter 6 we studied dynamic features in interpurchase times. A review of the literature showed that consecutive interpurchase spells are usually assumed to be independent. In practice, however, we would expect a correlation over time. Correlation

between spells could, for example, be induced by purchase acceleration effects or alternating periods of heavy and low use. Another feature that could lead to a positive correlation is unaccounted unobserved heterogeneity. Such correlation has no relation to behavioral characteristics and is usually referred to as spurious correlation. We have explicitly modeled possible unobserved differences in interpurchase times by allowing the intercepts of the model to differ across households.

Identifying dynamic effects in interpurchase timing is also relevant for practical purposes. We have shown that the effects of marketing instruments are not confined to the interpurchase spell in which the efforts are made. To assess the total effect of a marketing action a manager should also take into account the "carry-over" effects. The model we have proposed exactly fits this need. Using similar time-series techniques as in Chapter 4, we shape our model in such a way that allows for easy identification of the direct effect and the cumulative, or long-run, effect of specific marketing efforts.

Our results showed that there are significant dynamic effects in interpurchase times. Furthermore, the short-run effects of marketing instruments turned out to be significantly different from the long-run effects. We have analyzed three different categories of fast-moving consumer goods. Across these categories we found large differences in the short-run and long-run effects of marketing instruments. For the yogurt category we find that the short-run effect of price is smaller than the long-run effect, where both effects are positive. However, for the detergent category we found the opposite and for catsup we found a negative long-run price effect for a large fraction of households. The interpretation of this last finding is that if price would be permanently increased, the average interpurchase time will decrease for these households. Note that this does not mean that the sales of catsup would increase. Households could also choose to keep the consumption rate constant and purchase smaller package sizes. In general, feature and display promotions shorten interpurchase times. For the yogurt category we do not find any effect of feature. For this category there is also no long-run effect of display. That is, for this category the direct purchase accelerating effect of a display on is compensated by a decelerating effect on future interpurchase spells.

Again, there are a number of topics for future research. Further research into possible practical uses of the model for marketing managers seems useful. In the end marketing managers will be interested in long-run sales or market share. An interesting extension of the model is to capture purchase timing, brand choice and purchase quantity in a single dynamic model. In the literature there are a number of integrated models of these three marketing measures available. However, these models do not allow for a straightforward identification of relevant dynamic features. Future research could integrate our model with dynamic models for choice and purchase quantity to obtain a single dynamic framework in which all marketing measure on the household level are included. For the choice model we could use the model in Paap and Franses (2000), for purchase quantity one could

build on Böckenholt (1999). An alternative model that captures dynamics in these three measures simultaneously is Erdem *et al.* (2003). This paper focuses on forward looking behavior by households as the main source of dynamics.

In Chapter 7, we have studied the brand choice decision at the household level. In this chapter we have focused on two different decision processes that could be used by households at different purchase occasions. Under the first decision process, which we named "responsive to marketing efforts", the household actively compares the different available brands while taking into account promotional pricing and other marketing stimuli. Under the alternative decision process, households do not take into account the marketing mix. Brand choice decisions are in this case solely based on preferences, habit formation and baseline prices. Promotional price cuts do not influence the decisions made by these households.

The actual decision process used by a household at a specific purchase occasion is in general not observed. We infer the decision process used at a specific purchase occasions from the typically available choice data. The probability that a household is responsive at a specific purchase occasion depends on household characteristics, such as family size and household income, and characteristics of the shopping trip, for example the amount of money spent.

Next to modeling structural heterogeneity, that is, the fact that decision processes may differ, it is important to capture differences in base preferences that cannot be attributed to observable characteristics. In our model we have captured this preference heterogeneity by imposing a normal distribution on the brand intercepts. As a consequence of this, the likelihood function no longer evaluates to a closed-form expression. Consequently, the estimation of the resulting model requires advanced econometric techniques. An appealing approach is based on Simulated Maximum Likelihood [SML], where an approximation of the likelihood function is optimized over the parameter space. With SML the likelihood function is approximated using simulation. Usually, many simulation draws are necessary to obtain an accurate approximation. We have used importance sampling to improve the efficiency of the simulator. With importance sampling the number of draws needed to obtain an accurate approximation can be drastically reduced.

Our application of the model to the detergent category showed that the responsiveness model performs very well. Compared to various variants of the multinomial logit model, our model provides better forecasts. Concerning the responsiveness to marketing efforts, we found that most households tend to be responsive. On purchase occasions where many different items are bought, households are less responsive. The interpurchase time is positively correlated with the probability to respond to marketing efforts.

Models that incorporate different unobserved decision processes by households are also useful in other areas. For example, in an on-line environment, our approach could be used

to assess the likelihood of consumers to use the advice obtained from a so-called agent to make their decisions. For this analysis only data on the actual choices would be necessary. As in the responsiveness model, under both decision rules (use the agent or do not use the agent) different explanatory variables will play a role. This distinction will allow us to assess the likelihood of the agent being used.

The estimation method we have proposed seems very promising. Future research can be devoted to analyze the general properties of performing simulated maximum likelihood using importance sampling. This approach seems particularly useful in non-linear models with unobserved heterogeneity for which it is not possible to calculate the likelihood by analytic integration. The responsiveness model is an example of such a model. Another example, is a heterogeneous panel version of the diffusion model which we consider in Fok and Franses (2002).

Nederlandse samenvatting (Summary in Dutch)

Voor marketingmanagers is het belangrijk het effect te weten van marketinginstrumenten, zoals bijvoorbeeld prijs en promotie, op het koopgedrag van consumenten. Wat is bijvoorbeeld het effect van een 10% prijskorting op de totale verkochte hoeveelheid van een product? Men is niet alleen geïnteresseerd in het effect van de eigen instrumenten, maar ook in die van de concurrenten. Een belangrijke vraag is bijvoorbeeld of een toename in het marktaandeel van een concurrent ten koste gaat van het eigen marktaandeel of vooral van het marktaandeel van andere concurrenten.

Om de effecten van marketinginstrumenten te meten zijn verschillende technieken beschikbaar. Binnen de kwantitatieve marketing worden gegevens van promoties en gerealiseerde aankopen uit het verleden gebruikt om het effect van toekomstige marketingacties in te schatten. In bijna alle supermarkten worden de gemaakte aankopen geregistreerd via het scannen van streepjescodes. Van elke aankoop wordt opgeslagen welk merk gekocht is, hoeveel er gekocht is en wanneer de aankoop is gemaakt. Eventueel wordt er ook geregistreerd wie de aankoop heeft gedaan. De identificatie van huishoudens gebeurt in dat geval via persoonlijke klantenkaarten, of via speciale onderzoeksprogramma's. De huishoudens die aan zo'n programma deelnemen, scannen bij thuiskomst zelf nogmaals hun aankopen. Naast deze aankoopgegevens worden de prijzen en eventuele promoties van alle merken bijgehouden. De combinatie van deze twee databases bevat enorm veel informatie over het koopgedrag van consumenten. Het is echter niet eenvoudig om nuttige informatie uit de vaak zeer grote databases te halen. Econometrische modellen en technieken vormen handige gereedschappen hiertoe.

In dit proefschrift behandelen wij een aantal econometrische marketingmodellen. In het eerste deel van dit proefschrift staan modellen voor data op een geaggregeerd niveau centraal. In het bijzonder betreft dit modellen voor de marktaandelen van alle merken binnen een productcategorie. Het tweede deel behandelt modellen op het niveau van individuele huishoudens. Om precies te zijn, deze modellen beschrijven de aankoopplanning van huishoudens en hun merkvoorkeur.

Modellen op marktniveau

In hoofdstuk 2 presenteren wij een uitgebreide econometrische analyse van het bekende attractiemodel. Dit model beschrijft de relatie tussen de marktaandelen van alle merken binnen een productcategorie en de marketinginspanningen van deze merken.

Een veel gestelde eis aan een marktaandeelmodel is dat het model logisch consistente voorspellingen dient te genereren. Voorspelde marktaandelen moeten tussen 0 en 100% liggen en de som van de voorspellingen over alle merken moet 100% bedragen. Hoewel het attractiemodel één van de weinige modellen is dat aan deze voorwaarde voldoet, is er geen formeel econometrische analyse van dit model beschikbaar. In hoofdstuk 2 bespreken wij een aantal aspecten van dit model. Eén van de bevindingen is dat de meest populaire voorspellingsmethode geen zuivere marktaandeelvoorspellingen geeft, met andere woorden, gemiddeld gesproken zijn de voorspellingen ongelijk aan de werkelijke marktaandelen. Wij presenteren een alternatieve, op simulatie gebaseerde, voorspellingsmethode die wel zuivere voorspellingen genereert.

Een andere bijdrage van hoofdstuk 2 is de ontwikkeling van een modelselectiestrategie. In de praktijk wordt meestal uit de vele varianten van het attractiemodel slechts één specificatie gekozen. Onze analyse toont echter aan dat het gebruik van een modelselectiestrategie over het algemeen een model oplevert met betere voorspelkracht in vergelijking tot een vaste, vooraf gekozen, modelspecificatie.

In hoofdstuk 3 wordt het attractiemodel gebruikt om het effect van een merkintroductie op de concurrentie tussen bestaande merken te evalueren. De marketingliteratuur bevat vele studies naar de optimale competitieve reactie op een merkintroductie. Studies naar de werkelijke reacties in een markt zijn echter relatief schaars. Verschillen tussen de optimale en de werkelijke reacties kunnen erop wijzen dat managers niet optimaal hebben gereageerd. Een andere verklaring voor eventuele verschillen kan zijn dat de voorgeschreven strategieën in de praktijk niet optimaal zijn. Een gebruikelijke aanname bij het afleiden van de optimale reactie is namelijk dat de introductie geen effect heeft op de voorkeuren en het gedrag van consumenten. In de praktijk is het echter niet onwaarschijnlijk dat de introductie, of de daaropvolgende competitieve reactie, wel invloed heeft op het consumentengedrag. Een zeer competitieve prijsstelling van het nieuwe merk kan de consument bijvoorbeeld prijsgevoeliger maken.

In ons onderzoek concentreren wij ons op het ontwikkelen van methoden om te toetsen of er veranderingen hebben plaatsgevonden ten gevolge van een merkintroductie. Hierbij beschouwen wij zowel veranderingen in consumentengedrag als competitieve reacties van bestaande merken. Het gedrag van consumenten kan samengevat worden door de parameters van een marktaandeelmodel; veranderingen in consumentengedrag vertalen zich in veranderingen van de modelparameters. Via het marktaandeelmodel kunnen we

veranderingen in consumentengedrag onafhankelijk van mogelijke competitieve reacties bestuderen. Voor het implementeren van deze toestingsstrategie moet het standaard attractiemodel aangepast worden zodat het toepasbaar is voor markten waarbij het aantal beschikbare merken verandert, hetgeen immers gebeurt bij een introductie. In hoofdstuk 3 bespreken wij deze uitbreiding in detail.

Competitieve reacties van merken zijn eenvoudiger te identificeren. Hiervoor beschouwen wij redelijk eenvoudige modellen voor de geobserveerde tijdreeksen van marketinginstrumenten. Een verandering in het gebruik van een marketinginstrument correspondeert met een verandering in de parameters van zo'n model.

De bovenstaande technieken passen wij toe op een database met de marktaandelen van 12 merken wasmiddelen. Hoewel de bestaande literatuur over optimale reacties een prijsdaling voor de bestaande merken voorspelt, vinden wij dit niet in deze markt. Over het algemeen vinden wij weinig competitieve reacties. Er zijn wel significante veranderingen in het gedrag van consumenten. De prijs- en coupongevoeligheid en de basispreferenties voor de merken veranderen sterk als gevolg van de introductie. Deze veranderingen zijn een mogelijke reden waarom de gevonden competitieve reacties niet overeenkomen met de voorspellingen op basis van speltheoretische studies.

In hoofdstuk 4 besteden wij aandacht aan dynamische effecten in marktaandelen. Er is veel bewijs voor de stelling dat de effecten van marketinginstrumenten niet beperkt zijn tot de periode waarin ze gebruikt worden. Een bekend voorbeeld hiervan is de zogenaamde post-promotie dip. In weken na een promotie vinden we vaak een lager marktaandeel dan gewoonlijk. Een gedeelte van de aankopen die in deze weken gemaakt zouden zijn, zijn namelijk reeds in de week van de promotie gemaakt. Het is dus onverstandig om alleen het directe effect van een promotie te beschouwen, immers, alleen als het cumulatieve effect positief is is de promotie waardevol.

In het attractiemodel kunnen dynamische effecten eenvoudig gemodelleerd worden door vertraagde marktaandelen en marketinginstrumenten uit vorige perioden als verklarende variabelen op te nemen. De interpretatie van de dynamische effecten is in dit model minder eenvoudig. Uit de parameters van het model is bijvoorbeeld niet direct af te lezen wat het totale effect van een promotie is. Wij stellen een alternatieve formulering van het model voor, waarbij de parameters praktisch relevante dynamische eigenschappen representeren. Deze formulering staat bekend als het foutencorrectiemodel. De parameters in dit model representeren het directe effect en het cumulatieve effect van een promotie. Deze twee parameters geven, samen met de snelheid waarmee het effect van een promotie uitdooft, alle relevante informatie over de dynamische effecten van een marketinginstrument.

Naast het meten van de effecten van marketinginstrumenten zijn we ook geïnteresseerd in het verklaren van mogelijke verschillen in deze effecten tussen merken. Hiertoe breiden

wij ons model uit met een extra laag. In deze tweede laag worden de dynamische effecten van marketinginstrumenten gerelateerd aan markt- en merkkarakteristieken. Belangrijke karakteristieken zijn bijvoorbeeld de concentratiegraad van de markt en de relatieve prijs van een merk.

Wij passen de besproken methode toe op een database van zeven productcategorieën in twee Amerikaanse steden. Totaal hebben wij de marktaandelen en marketinginstrumenten van 50 merken. De resultaten laten zien dat het lange termijn (cumulatieve) effect van prijs over het algemeen kleiner is dan het korte termijn (directe) effect. Voor feature en display promoties geldt het omgekeerde. Voor coupons geldt dat het directe effect ongeveer gelijk is aan het cumulatieve effect. Verder vinden wij dat een intensiever gebruik van marketinginstrumenten en een hogere marktconcentratie leidt tot sterkere prijseffecten, zowel op de korte als op de lange termijn.

Modellen op huishoudniveau

In het tweede deel van dit proefschrift staan modellen op huishoudniveau centraal. Via speciale onderzoeksprogramma's, of met behulp van klantenkaarten, wordt het aankoopgedrag van een groot aantal huishoudens tot in detail in kaart gebracht. Van elk huishouden in het programma is bekend wanneer en waar een aankoop is gemaakt, welk merk gekozen is en hoeveel er van het product is gekocht. Uit de combinatie van deze gegevens en het gebruik van marketinginstrumenten kan veel informatie over individueel consumentengedrag afgeleid worden. In het algemeen zijn er drie belangrijke vragen. Ten eerste is men geïnteresseerd in de aankoopplanning van huishoudens. Belangrijke aspecten hierbij zijn de consumptiesnelheid en de invloed van marketinginstrumenten op de aankoopbeslissing. Een tweede belangrijk punt is de merkkeuzebeslissing van consumenten. Tot slot is de aangekochte hoeveelheid van een product een interessante variabele. In dit proefschrift besteden wij aandacht aan de aankoopplanning en de merkkeuze.

In hoofdstuk 5 bestuderen wij de aankoopplanning van huishoudens. Om de tussenaankooptijden van huishoudens in een bepaalde productcategorie te verklaren en te beschrijven zijn verschillende modellen beschikbaar. Een gemeenschappelijk probleem van alle modellen is echter dat marketinginstrumenten niet eenvoudig als verklarende variabelen opgenomen kunnen worden. De aankoopbeslissing wordt namelijk op het niveau van de productcategorie gemodelleerd, terwijl de marketinginstrumenten per merk worden gemeten. Een onderzoeker zal dus de marketinginspanningen van de verschillende merken moeten aggregeren. Er bestaat nog geen consensus over de beste methode voor deze aggregatie.

Wij bespreken een aantal bestaande aggregatietechnieken en verder stellen wij drie nieuwe alternatieven voor. Alle methoden worden zowel in discrete als in continue tijd

besproken. Naast een vergelijking op theoretische gronden, presenteren wij een vergelijking van de empirische prestaties van de modellen in continue tijd.

De resultaten wijzen uit dat wanneer niet-verklaarde verschillen in merkpreferenties, zogenaamde niet-waargenomen heterogeniteit, genegeerd worden, de standaard aanpak in de literatuur het beste werkt. In deze aanpak worden huishoudenspecifieke keuzeaandelen gebruikt om voor elk huishouden de marketinginstrumenten over de merken te aggregeren. Voor het voorspellen van tussenaankooptijden van nieuwe huishoudens werkt deze methode niet optimaal. Voor dit soort voorspellingen blijkt één van de nieuwe modellen het beste te werken. Dit model beschouwt de aankoopplanning en de merkkeuze tegelijk. Ook wanneer er wel rekening gehouden wordt met niet-waargenomen heterogeniteit blijkt dit model het beste te werken. Een andere conclusie is dat het populaire model gebaseerd op de zogenaamde "inclusive value" over het algemeen matig presteert.

In hoofdstuk 6 presenteren wij een meer inhoudelijke studie naar tussenaankooptijden. In dit hoofdstuk zijn wij vooral geïnteresseerd in mogelijke afhankelijkheden tussen opeenvolgende tussenaankooptijden. Een gebruikelijke aanname is dat de planning van opeenvolgende aankoopbeslissingen onafhankelijk zijn. Op basis van kennis uit andere onderzoeksgebieden lijkt deze aanname niet realistisch. Beschouw bijvoorbeeld een consument die door een promotie verleid wordt om een aankoop eerder te maken dan was voorzien. Deze consument heeft nu extra voorraad van dit product, en als de consumptiesnelheid constant blijft zal het langer dan gebruikelijk duren voordat deze voorraad is verbruikt. De volgende aankoop zal dus worden uitgesteld. In dit voorbeeld is er een negatieve correlatie tussen de twee opeenvolgende aankooptijden. Dit voorbeeld toont ook aan dat de effecten van marketinginstrumenten niet altijd beperkt zijn tot de lopende aankoopbeslissing. Om het effect van marketingacties goed te kunnen inschatten is het dus belangrijk om ook dynamische effecten in ogenschouw te nemen.

In dit hoofdstuk breiden wij het populaire "accelerated failure-time" model uit om ook afhankelijkheden tussen aankooptijden te kunnen beschrijven. Opnieuw gebruiken wij hiertoe een variant van het uit de tijdreeksanalyse bekende foutencorrectiemodel. Het gebruik van deze modelspecificatie leidt tot eenvoudig interpreteerbare parameters. In dit geval geven de parameters het directe effect van een marketingactie op de huidige tussenaankoopperiode en de som van de effecten op alle toekomstige aankoopbeslissingen. Dit laatste effect is gelijk aan het effect van een permanente promotie (bijvoorbeeld een permanente prijsverlaging) op tussenaankooptijden in de verre toekomst.

In dit proefschrift worden het model en de interpretatie ervan uitgebreid besproken. Daarnaast presenteren wij toepassingen van het model op drie verschillende productcategorieën. De resultaten tonen aan dat opeenvolgende tussenaankooptijden wel degelijk afhankelijk van elkaar zijn. Tevens vinden wij significante dynamische effecten van marketinginstrumenten. De directe effecten van marketinginstrumenten verschillen significant

van de lange termijn (cumulatieve) effecten. Over de productcategorieën heen blijken verder grote verschillen te bestaan met betrekking tot de dynamische effecten.

In hoofdstuk 7 staat de merkkeuzebeslissing van huishoudens centraal. Nadat een huishouden besloten heeft een aankoop te doen in een bepaalde productcategorie wordt de keuze voor een bepaald merk gemaakt. Deze keuze kan beïnvloed worden door de actuele prijzen van de merken en eventuele promotionele activiteiten. Het beslissingsproces dat gebruikt wordt, hoeft niet voor elk huishouden en op elke aankoopmoment hetzelfde te zijn. Wij onderscheiden twee verschillende beslissingsprocessen. Binnen het eerste proces weegt het huishouden alle verschillende merken actief tegen elkaar af. Hierbij houdt het huishouden rekening met zijn basispreferenties voor de verschillende merken, de prijzen en promotionele activiteiten. Deze strategie vergt echter relatief veel inspanning van het huishouden. Het is daarom niet waarschijnlijk dat elk huishouden bij elk winkelbezoek voor elke productcategorie zo'n weloverwogen beslissing zal maken. Wij introduceren een alternatief beslissingsproces dat minder inspanning vereist. Binnen dit proces laat het huishouden zich vooral sturen door gewoonte en basisvoorkeuren. Marketinginstrumenten spelen bij dit beslissingsproces geen rol. Voor beide beslissingsprocessen staan wij verder toe dat eerder gemaakte keuzes invloed hebben op de huidige keuze.

Uit de beschikbare aankoopgegevens is niet direct af te leiden welk beslissingsproces het huishouden daadwerkelijk heeft gebruikt. Wel kunnen we uitspraken doen over de waarschijnlijkheid dat een bepaald proces is gebruikt. Deze waarschijnlijkheid hangt af van eigenschappen van het huishouden en van eigenschappen van het winkelbezoek, zoals bijvoorbeeld de grootte van het huishouden, de tijd sinds de laatste aankoop en het bedrag dat tijdens het bezoek wordt uitgegeven.

In het hoofdstuk wordt het model uitgebreid besproken en er is bijzondere aandacht voor het schatten van de parameters. De combinatie van verschillen in het gebruikte beslissingsproces en mogelijke verschillen in basispreferenties over de huishoudens en aankoopmomenten leiden tot een complex model. Geavanceerde simulatietechnieken zijn nodig voor het schatten van de parameters van dit model. Eén van de bijdragen van dit hoofdstuk is een verbetering van de standaard simulatiemethode. Wij stellen een vernieuwde, efficiëntere, simulatietechniek op basis van "Importance Sampling" voor. Naast de technische punten, presenteren wij een toepassing op de merkkeuzes binnen de wasmiddelencategorie.

Uit de toepassing blijkt dat huishoudens tijdens de meeste, maar niet alle, winkelbezoeken gevoelig zijn voor promoties. Huishoudens die veel verschillende items tegelijk kopen, grote volumes aankopen of zeer frequent aankopen doen, zijn over het algemeen minder gevoelig. Naast deze gedragseigenschappen vinden wij dat ons model betere voorspellingen genereert dan concurrerende modellen waarin de gevoeligheid voor marketinginstrumenten niet is meegenomen.

Samenvattend, in dit proefschrift bespreken wij modellen voor verschillende marketing-maatstaven. Naast een bijdrage aan de marketing, bijvoorbeeld door de geboden inzichten in structurele heterogeniteit in keuzeprocessen en in dynamische effecten in marktaandelen en tussenaankooptijden, draagt dit proefschrift ook bij aan de econometrische literatuur door de uitbreiding en ontwikkeling van modellen en schattingstechnieken.

Bibliography

Ailawadi, K. L. and S. A. Neslin (1998), The Effect of Promotion on Consumption: Buying More and Consuming it Faster, *Journal of Marketing Research*, **35**, 390–398.

Allenby, G. M. and P. J. Lenk (1994), Modeling Household Purchase Behavior with Logistic Normal Regression, *Journal of the American Statistical Association*, **89**, 1218–1231.

Allenby, G. M. and P. E. Rossi (1999), Marketing Models of Consumer Heterogeneity, *Journal of Econometrics*, **89**, 57–78.

Allenby, G. M., R. P. Leone, and L. Jen (1999), A Dynamic Model of Purchase Timing with Application to Direct Marketing, *Journal of the American Statistical Association*, **94**, 365–374.

Basuroy, S. and D. Nguyen (1998), Multinomial Logit Market Share Models: Equilibrium Characteristics and Strategic Implications, *Management Science*, **44**, 1396–1408.

Bell, D. E., R. L. Keeney, and J. D. C. Little (1975), A Market Share Theorem, *Journal of Marketing Research*, **12**, 136–141.

Bell, D. R., J. Chiang, and V. Padmanabhan (1999), The Decomposition of Promotional Response: An Empirical Generalization, *Marketing Science*, **18**, 504–526.

Ben-Akiva, M. and S. R. Lerman (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, vol. 9 of *MIT Press Series in Transportation Studies*, MIT Press, Cambridge (MA).

Blattberg, R. C., G. C. Eppen, and J. Liebermann (1981), A Theoretical and Empirical Evaluation of Price Deals for Consumer Durables, *Journal of Marketing*, **45**, 116–129.

Böckenholt, U. (1999), Mixed INAR(1) Poisson Regression Models: Analyzing Heterogeneity and Serial Dependencies in Longitudinal Count Data, *Journal of Econometrics*, **89**, 317–338.

Bolton, R. N. (1989), The Relationship Between Market Characteristics and Promotional Price Elasticities, *Marketing Science*, **8**, 153–169.

Bowman, D. and H. Gatignon (1996), Order of Entry as a Moderator of the Effect of the Marketing Mix on Market Share, *Marketing Science*, **15**, 222–242.

Bradu, D. and Y. Mundlak (1970), Estimation in Lognormal Linear Models, *Journal of the American Statistical Association*, **65**, 198–211.

Brodie, R. J. and A. Bonfrer (1994), Conditions When Market Share Models are Useful for Forecasting: Further Empirical Results, *International Journal of Forecasting*, **10**, 277–285.

Brodie, R. J. and C. A. de Kluyver (1984), Attraction versus Linear and Multiplicative Market Share Models: An Empirical Evaluation, *Journal of Marketing Research*, **21**, 194–201.

Brodie, R. J., P. J. Danaher, V. Kumar, and P. S. H. Leeflang (2001), Econometric Models for Forecasting Market Share, in J. S. Armstrong (ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Kluwer, Norwell, MA, pp. 597–611.

Bronnenberg, B. J., V. Mahajan, and W. R. Vanhonacker (2000), The Emergence of New Repeat-Purchase Categories: The Interplay of Market Share and Retailer Distribution, *Journal of Marketing Research*, **37**, 16–31.

Bucklin, R. E. and S. Gupta (1992), Brand Choice, Purchase Incidence, and Segmentation: An Integrated Approach, *Journal of Marketing Research*, **29**, 201–215.

Bucklin, R. E. and J. M. Lattin (1991), A Two-State Model of Purchase Incidence and Brand Choice, *Marketing Science*, **10**, 24–39.

Bucklin, R. E., S. Gupta, and S. Siddarth (1998), Determining Segmentation in Sales Response Across Consumer Purchase Behaviors, *Journal of Marketing Research*, **35**, 189–197.

Carpenter, G. S., L. G. Cooper, D. M. Hanssens, and D. Midgley (1988), Modeling Asymmetric Competition, *Marketing Science*, **7**, 393–412.

Casella, G. and E. I. George (1992), Explaining the Gibbs Sampler, *American Statistician*, **46**, 167–174.

Chen, Y., V. Kanetkar, and D. L. Weiss (1994), Forecasting Market Shares with Disaggregate of Pooled Data: A Comparison of Attraction Models, *International Journal of Forecasting*, **10**, 263–276.

Chintagunta, P. K. (1993), Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households, *Marketing Science*, **12**, 184–208.

Chintagunta, P. K. (1999), Measuring the Effects of New Brand Introduction on Inter-Brand Strategic Interaction, *European Journal of Operational Research*, **118**, 315–331.

Chintagunta, P. K. and S. Haldar (1998), Investigating Purchase Timing Behavior in Two Related Product Categories, *Journal of Marketing Research*, **35**, 43–53.

Chintagunta, P. K. and A. R. Prasad (1998), An Empirical Investigation of the "Dynamic McFadden" Model of Purchase Timing and Brand Choice: Implications for Market Structure, *Journal of Business & Economic Statistics*, **16**, 2–12.

Chintagunta, P. K., D. C. Jain, and N. J. Vilcassim (1991), Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data, *Journal of Marketing Research*, **28**, 417–428.

Chow, G. (1960), Tests of Equality Between Sets of Coefficients in Two Linear Regressions, *Econometrica*, **28**, 591–605.

Cooper, L. G. (1993), Market-Share Models, in J. Eliashberg and G. L. Lilien (eds.), *Handbook in Operations Research and Management Science*, vol. 5, chap. 6, North-Holland, Amsterdam, pp. 259–314.

Cooper, L. G. and M. Nakanishi (1988), *Market Share Analysis: Evaluating Competitive Marketing Effectiveness*, Kluwer Academic Publishers, Boston.

Cubbin, J. and S. Domberger (1988), Advertising and Post-Entry Oligopoly Behavior, *Journal of Industrial Economics*, **37**, 123–140.

Danaher, P. J. (1994), Comparing Naive with Econometric Market Share Models when Competitors' Actions are Forecast, *International Journal of Forecasting*, **10**, 287–294.

Davies, R. B. (1977), Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative, *Biometrika*, **64**, 247–254.

Dekimpe, M. G. and D. M. Hanssens (1995a), Empirical Generalizations about Market Evolution and Stationarity, *Marketing Science*, **14**, G109–G121.

Dekimpe, M. G. and D. M. Hanssens (1995b), The Persistence of Marketing Effects on Sales, *Marketing Science*, **14**, 1–21.

Dekimpe, M. G., D. M. Hanssens, and J. M. Silva-Risso (1999), Long-Run Effects of Price Promotions in Scanner Markets, *Journal of Econometrics*, **89**, 269–291.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977), Maximum Likelihood from Incomplete Data Via the EM Algorithm (with Discussion), *Journal of the Royal Statistical Society B*, **39**, 1–38.

Doornik, J. A. (2002), *Object-Oriented Matrix Programming using Ox*, 3 edn., Timberlake Consultants Press and Oxford, `www.nuff.ox.ac.uk/Users/Doornik`, London.

Engle, R. F. and J. R. Russell (1998), Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data, *Econometrica*, **66**, 1127–1162.

Erdem, T., S. Imai, and M. P. Keane (2003), Brand and Quantity Choice Dynamics Under Price Uncertainty, *Quantitative Marketing and Economics*, **1**, 5–64.

Finney, D. J. (1941), On the Distribution of a Variate Whose Logarithm is Normally Distributed, *Supplement to the Journal of the Royal Statistical Association*, **7**, 155–161.

Fok, D. and P. H. Franses (2001), Forecasting Market Shares from Models for Sales, *International Journal of Forecasting*, **17**, 121–128.

Fok, D. and P. H. Franses (2002), Modeling the Diffusion of Scientific Publications, unpublished manuscript, Erasmus University Rotterdam.

Fok, D. and R. Paap (2003), Modeling Category-Level Purchase Timing with Brand-Level Marketing Variables, Econometric Institute Report EI-2003-15, Erasmus University Rotterdam.

Fok, D., P. H. Franses, and R. Paap (2001), Incorporating Responsiveness to Marketing Efforts when Modeling Brand Choice, ERIM Report Series in Management ERS-2001-47-MKT, Erasmus University Rotterdam.

Fok, D., P. H. Franses, and R. Paap (2002a), Econometric Analysis of the Market Share Attraction Model, in P. H. Franses and A. L. Montgomery (eds.), *Advances in Econometrics*, vol. 16, chap. 10, JAI Press, pp. 223–256.

Fok, D., P. H. Franses, and R. Paap (2002b), Modeling Dynamic Effects of Promotion on Interpurchase Times, Econometric Institute Report EI-2002-37, Erasmus University Rotterdam.

Fok, D., P. H. Franses, and R. Paap (2003), Modeling Dynamic Effects of the Marketing Mix on Market Shares, ERIM Report Series in Management ERS-2003-44-MKT, Erasmus University Rotterdam.

Franses, P. H. (1994), Modeling New Product Sales: An Application of Cointegration Analysis, *International Journal of Research in Marketing*, **11**, 491–502.

Franses, P. H. and R. Paap (2001), *Quantitative Models in Marketing Research*, Cambridge University Press, Cambridge.

Franses, P. H., S. Srinivasan, and P. Boswijk (2001), Testing for Unit Roots in Market Shares, *Marketing Letters*, **12**, 351–364.

Gatignon, H., B. Weitz, and P. Bansal (1990), Brand Introduction Strategies and Competitive Environments, *Journal of Marketing Research*, **27**, 390–401.

Geman, S. and D. Geman (1984), Stochastic Relaxations, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Geweke, J. (1989a), Bayesian Inference in Econometric Models Using Monte Carlo Integration, *Econometrica*, **57**, 1317–1339.

Geweke, J. (1989b), Exact Predictive Densities for Linear Models with ARCH Disturbances, *Journal of Econometrics*, **40**, 63–86.

Ghosh, A., S. A. Neslin, and R. Shoemaker (1984), A Comparison of Market Share Models and Estimation Procedures, *Journal of Marketing Research*, **21**, 202–210.

Gönül, F. and K. Srinivasan (1993a), Consumer Purchase Behavior in a Frequently Bought Product Category: Estimation Issues and Managerial Insights from a Hazard Function Model with Heterogeneity, *Journal of the American Statistical Association*, **88**, 1219–1227.

Gönül, F. and K. Srinivasan (1993b), Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues, *Marketing Science*, **12**, 213–229.

Gourieroux, C. and A. Montfort (1993), Simulation-based Inference: A Survey with Special Reference to Panel Data Model, *Journal of Econometrics*, **59**, 5–33.

Greene, W. H. (1993), *Econometric Analysis*, Prentice-Hall Inc., New Jersey.

Gruca, T. S., K. R. Kumar, and D. Sudharshan (1992), An Equilibrium Analysis of Defensive Response to Entry Using a Coupled Response Function Model, *Marketing Science*, **11**, 348–358.

Gruca, T. S., D. Sudharshan, and K. R. Kumar (2001), Marketing Mix Response to Entry in Segmented Markets, *International Journal of Research in Marketing*, **18**, 53–66.

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

Gupta, S. (1988), Impact of Sales Promotions on When, What, and How Much to Buy, *Journal of Marketing Research*, **25**, 342–355.

Gupta, S. (1991), Stochastic Models of Interpurchase Time With Time-Dependent Covariates, *Journal of Marketing Research*, **28**, 1–15.

Hajivassiliou, V. and P. Ruud (1994), Classical Estimation Methods for LDV Models using Simulation, in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, North-Holland, Amsterdam, pp. 2383–2411.

Hanssens, D. M., L. J. Parsons, and R. L. Schultz (1989), *Market Response Models: Econometric and Time Series Analysis*, Kluwer Academic Publishers, Norwell, USA.

Helsen, K. and D. C. Schmittlein (1992), How Does a Product Market's Typical Price-Promotion Pattern Affect the Timing of Households' Purchases? An Empirical Study Using UPC Scanner Data, *Journal of Retailing*, **68**, 316–338.

Helsen, K. and D. C. Schmittlein (1993), Analyzing Duration Times in Marketing: Evidence for the Effectiveness of Hazard Rate Models, *Marketing Science*, **11**, 395–414.

Hendry, D. F. (1995), *Dynamic Econometrics*, Oxford University Press, Oxford.

Hendry, D. F., A. R. Pagan, and J. D. Sargan (1984), Dynamic Specification, in Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, chap. 18, North-Holland, Amsterdam, pp. 1023–1100.

Hobert, J. P. and G. Casella (1996), The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models, *Journal of the American Statistical Association*, **91**, 1461–1473.

Hsu, A. and R. T. Wilcox (2000), Stochastic Prediction in Multinomial Logit Models, *Management Science*, **46**, 1137–1144.

Jain, D. C. and N. J. Vilcassim (1991), Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach, *Marketing Science*, **10**, 1–23.

Jain, D. C., N. J. Vilcassim, and P. K. Chintagunta (1994), A Random-Coefficients Logit Brand-Choice Model Applied to Panel Data, *Journal of Business & Economic Statistics*, **12**, 317–328.

Jedidi, K., C. F. Mela, and S. Gupta (1999), Managing Advertising and Promotion for Long-Run Profitability, *Marketing Science*, **19**, 1–22.

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

Judge, G. G., W. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee (1985), *The Theory and Practice of Econometrics*, 2nd edn., John Wiley & Sons, New York.

Kalbfleisch, J. and R. Prentice (1980), *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, New York.

Kamakura, W. A. and G. J. Russell (1989), A Probabilistic Choice Model for Market Segmentation and Elasticity Structure, *Journal of Marketing Research*, **26**, 379–390.

Kamakura, W. A., B.-D. Kim, and J. Lee (1996), Modeling Preference and Structural Heterogeneity in Consumer Choice, *Marketing Science*, **15**, 152–172.

Karnani, A. (1985), Strategic Implications of Market Share Attraction Models, *Management Science*, **31**, 536–547.

Keane, M. P. (1997), Modeling Heterogeneity and State Dependence in Consumer Choice Behavior, *Journal of Business & Economic Statistics*, **15**, 310–327.

Kiefer, N. M. (1988), Economic Duration Data and Hazard Functions, *Journal of Economic Literature*, **26**, 646–679.

Klapper, D. and H. Herwartz (2000), Forecasting Market Share Using Predicited Values of Competitor Behavior: Further Empirical Results, *International Journal of Forecasting*, **16**, 399–421.

Kloek, T. and H. K. van Dijk (1978), Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte-Carlo, *Econometrica*, **44**, 345–351.

Kumar, V. (1994), Forecasting Performance of Market Share Models: An Assessment, Additional Insights, and Guidelines, *International Journal of Forecasting*, **10**, 295–312.

Lal, R. and V. Padmanabhan (1995), Competitive Response and Equilibria, *Marketing Science*, **14**, G101–G108.

Lancaster, T. (1979), Econometric methods for the Duration of Unemployment, *Econometrica*, **47**, 939–956.

Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, vol. 17 of *Econometric Society Monographs*, Cambridge University Press, Cambridge.

Lee, L. (1995), Asymptotic Bias in Simulated Maximum Likelihood Estimation of Discrete Choice Models, *Econometric Theory*, **11**, 437–483.

Leeflang, P. S. H. and J. C. Reuyl (1984), On the Predictive Power of Market Share Attraction Model, *Journal of Marketing Research*, **21**, 211–215.

Leeflang, P. S. H., D. R. Wittink, M. Wedel, and P. A. Naert (2000), *Building Models for Marketing Decisions*, Kluwer Academic Publishers, Dordrecht.

Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, 2nd edn., Springer-Verlag, Berlin.

Maddala, G. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, vol. 3 of *Econometric Society Monographs*, Cambridge University Press, Cambridge.

McFadden, D. L. (1973), Conditional Logit Analysis of Qualitative Choice Behavior, in P. Zarembka (ed.), *Frontiers in Econometrics*, chap. 4, Academic Press, New York, pp. 105–142.

McFadden, D. L. (1981), Econometric Models of Probabilistic Choice, in C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data: With Econometric Applications*, MIT Press, Cambridge, MA, pp. 197–272.

McFadden, D. L. and K. E. Train (2000), Mixed MNL Models for Discrete Response, *Journal of Applied Econometrics*, **15**, 447–470.

Mela, C. F., S. Gupta, and D. Lehmann (1997), The Long-term Impact of Promotions and Advertising on Consumer Brand Choice, *Journal of Marketing Research*, **34**, 248–261.

Mela, C. F., S. Gupta, and K. Jedidi (1998), Assessing Long-term Promotional Influences on Market Structure, *International Journal of Research in Marketing*, **15**, 89–107.

Naert, P. A. and M. Weverbergh (1981), On the Prediction Power of Market Share Attraction Models, *Journal of Marketing Research*, **18**, 146–153.

Neslin, S. A., C. Henderson, and J. Quelch (1985), Consumer Promotions and the Acceleration of Product Purchases, *Marketing Science*, **4**, 147–165.

Newey, W. K. and D. L. McFadden (1994), Large Sample Estimation and Hypothesis Testing, in R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, chap. 36, Elsevier Science B.V., Amsterdam, pp. 2111–2245.

Newton, M. A. and A. E. Raftery (1994), Approximate Bayesian Inference with the Weighted Likelihood Bootstrap, *Journal of the Royal Statistical Society B*, **56**, 3–48.

Nijs, V. R., M. G. Dekimpe, J.-B. E. M. Steenkamp, and D. M. Hanssens (2001), The Category-Demand Effects of Price Promotions, *Marketing Science*, **20**, 1–22.

Paap, R. (2002), What are the Advantages of MCMC Based Inference in Latent Variable Models?, *Statistica Neerlandica*, **56**, 1–21.

Paap, R. and P. H. Franses (2000), A Dynamic Multinomial Probit Model for Brand Choice with Different Long-run and Short-run Effects of Marketing-Mix Variables, *Journal of Applied Econometrics*, **15**, 717–744.

Pauwels, K., D. M. Hanssens, and S. Siddarth (2002), The Long-Term Effects of Price Promotions on Category Incidence, Brand Choice, and Purchase Quantity, *Journal of Marketing Research*, **39**, 421–439.

Raju, J. S. (1992), The Effect of Price Promotions on Variability in Product Category Sales, *Marketing Science*, **11**, 207–220.

Ridder, G. (1990), The Non-Parametric Identification of Generalized Accelerated Failure-Time Models, *Review of Economic Studies*, **57**, 167–182.

Robinson, W. T. (1988), Marketing Mix Reactions by Incumbents to Entry, *Marketing Science*, **7**, 368–385.

Schwarz, G. (1978), Estimating the Dimension of a Model, *Annals of Statistics*, **6**, 461–464.

Seetharaman, P. B. and P. K. Chintagunta (2003), The Proportional Hazard Model for Purchase Timing: A Comparison of Alternative Specifications, *Journal of Business & Economic Statistics*, forthcoming.

Shankar, V. (1997), Pioneers' Marketing Mix Reactions to Entry in Different Competitive Game Structures: Theoretical Analysis and Empirical Illustration, *Marketing Science*, **16**, 271–293.

Shankar, V. (1999), New Product Introduction and Incumbent Response Strategies: Their Interrelationship and the Role of Multimarket Contact, *Journal of Marketing Research*, **36**, 327–344.

Shankar, V., G. S. Carpenter, and L. Krishnamurthi (1999), The Advantages of Entry in the Growth Stage of the Product Life Cycle: An Empirical Analysis, *Journal of Marketing Research*, **36**, 269–276.

Smith, A. F. M. and G. O. Roberts (1993), Bayesian Computing via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods, *Journal of the Royal Statistical Society B*, **55**, 3–23.

Srinivasan, S. and F. M. Bass (2000), Cointegration Analysis of Brand Sales and Category Sales: Stationarity and Long-run Equilibrium in Market Shares, *Applied Stochastic Models in Business and Industry*, **16**, 159–177.

Srinivasan, S., P. T. L. Popkowski Leszczyc, and F. M. Bass (2000), Market Share Response and Competitive Interaction: The Impact of Temporary, Evolving and Structural Changes in Prices, *International Journal of Research in Marketing*, **17**, 281–305.

Srinivasan, S., K. Pauwels, D. M. Hanssens, and M. G. Dekimpe (2001), Do Promotions Benefit Manufacturers, Retailers, or Both?, Marketing Science Institute Report 01-120.

Takada, H. and F. M. Bass (1998), Multiple Time Series Analysis of Competitive Marketing Behavior, *Journal of Business Research*, **43**, 97–107.

Tanner, M. and W. Wong (1987), The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association*, **82**, 528–550.

Tierney, L. (1994), Markov Chains for Exploring Posterior Distributions, *Annals of Statistics*, **22**, 1701–1762.

Tobin, J. (1958), Estimation of Relationships for Limited Dependent Variables, *Econometrica*, **26**, 24–36.

Train, K. E. (2003), *Discrete Choice Models with Simulation*, Cambridge University Press, Cambridge.

Vakratsas, D. and F. M. Bass (2002), A Segment-Level Hazard Approach to Studying Household Purchase Timing Decisions, *Journal of Applied Econometrics*, **17**, 49–59.

van Heerde, H. J., P. S. H. Leeflang, and D. R. Wittink (2000), The Estimation of Pre- and Postpromotion Dips with Store-Level Scanner Data, *Journal of Marketing Research*, **37**, 383–395.

van Nierop, J. E. M., D. Fok, and P. H. Franses (2002), Sales Models for Many Items Using Attribute Data, ERIM Report Series in Management, ERS-2002-65-MKT.

Vilcassim, N. J. and D. C. Jain (1991), Modeling Purchase-Timing and Brand-Switching Behavior Incorporating Explanatory Variables and Unobserved Heterogeneity, *Journal of Marketing Research*, **28**, 29–41.

Wedel, M. and W. A. Kamakura (1999), *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers, Dordrecht.

Wedel, M., W. A. Kamakura, N. Arora, A. Bemmaor, J. Chiang, T. Elrod, R. Johnson, P. J. Lenk, S. A. Neslin, and C. S. Poulsen (1999), Discrete and Continuous Representations of Unobserved Heterogeneity in Choice Modelling, *Marketing Letters*, **10**, 219–232.

Yang, S. and G. M. Allenby (2000), A Model for Observation, Structural, and Household Heterogeneity in Panel Data, *Marketing Letters*, **11**, 137–149.

Zellner, A. (1962), An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias, *Journal of the American Statistical Association*, **57**, 348–368.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

Author index

Curriculum Vitea

Dennis Fok (1977) obtained his master's degree in econometrics with honors from the Erasmus University Rotterdam in 1999. In the same year he started as a Ph.D. student with marketing and econometrics as the primary fields of interest at the same university. His research on econometric models in marketing has resulted in several publications in international journals. In September 2003, he started a research position at the Econometric Institute, Erasmus University Rotterdam.

Erasmus Research Institute of Management

ERIM Ph.D. Series Resesearch in Management

1. J. Robert van der Meer, **Operational Control of Internal Transport**
   Promotor(es): Prof.dr. M.B.M. de Koster, Prof.dr.ir. R. Dekker
   Defended: September 28, 2000, ISBN: 90-5892-004-6

2. Moritz Fleischmann, **Quantitative Models for Reverse Logistics**
   Promotor(es): Prof.dr.ir. J.A.E.E. van Nunen, Prof.dr.ir. R. Dekker
   Defended: October 5, 2000, Lecture Notes in Economics and Mathematical Systems, Volume 501, 2001, Springer Verlag, Berlin, ISBN: 3540-417-117

3. Dolores Romero Morales, **Optimization Problems in Supply Chain Management**
   Promotor(es): Prof.dr.ir. J.A.E.E. van Nunen, dr. H.E. Romeijn
   Defended: October 12, 2000, ISBN: 90-9014-078-6

4. Kees Jan Roodbergen, **Layout and Routing Methods for Warehouses**
   Promotor(es): Prof.dr. M.B.M. de Koster, Prof.dr.ir. J.A.E.E. van Nunen
   Defended: May 10, 2001, ISBN: 90-5892-005-4

5. Rachel Campbell, **Rethinking Risk in International Financial Markets**
   Promotor(es): Prof.dr. C.G. Koedijk
   Defended: September 7, 2001, ISBN: 90-5892-008-9

6. Yongping Chen, **Labour flexibility in China's companies: an empirical study**
   Promotor(es): Prof.dr. A. Buitendam, Prof.dr. B. Krug
   Defended: October 4, 2001, ISBN: 90-5892-012-7

7. Pursey P.M.A.R. Heugens, **Strategic Issues Management: Implications for Corporate Performance**
   Promotor(es): Prof.dr.ing. F.A.J. van den Bosch, Prof.dr. C.B.M. van Riel
   Defended: October 19, 2001, ISBN: 90-5892-009-7

8. Roland F. Speklé, **Beyond Generics; A closer look at Hybrid and Hierarchical Governance**
   Promotor(es): Prof.dr. M.A. van Hoepen RA
   Defended: October 25, 2001, ISBN: 90-5892-011-9

9. Pauline Puvanasvari Ratnasingam, **Interorganizational Trust in Business to Business E-Commerce**
   Promotor(es): Prof.dr. K. Kumar, Prof.dr. H.G. van Dissel
   Defended: November 22, 2001, ISBN: 90-5892-017-8

10. Michael M. Mol, **Outsourcing, Supplier-relations and Internationalisation: Global Source Strategy as a Chinese puzzle**
    Promotor(es): Prof.dr. R.J.M. van Tulder
    Defended: December 13, 2001, ISBN: 90-5892-014-3

11. Matthijs J.J. Wolters, **The Business of Modularity and the Modularity of Business**
    Promotor(es): Prof.mr.dr. P.H.M. Vervest, Prof.dr.ir. H.W.G.M. van Heck
    Defended: February 8, 2002, ISBN: 90-5892-020-8

12. J. van Oosterhout, **The Quest for Legitimacy; On Authority and Responsibility in Governance**
    Promotor(es): Prof.dr. T. van Willigenburg, Prof.mr. H.R. van Gunsteren
    Defended: May 2, 2002, ISBN: 90-5892-022-4

13. Otto R. Koppius, **Information Architecture and Electronic Market Performance**
    Promotor(es): Prof.dr. P.H.M. Vervest, Prof.dr.ir. H.W.G.M. van Heck
    Defended: May 16, 2002, ISBN: 90-5892-023-2

14. Iris F.A. Vis, **Planning and Control Concepts for Material Handling Systems**
    Promotor(es): Prof.dr. M.B.M. de Koster, Prof.dr.ir. R. Dekker
    Defended: May 17, 2002, ISBN: 90-5892-021-6

15. Jos Bijman, **Essays on Agricultural Co-operatives; Governance Structure in Fruit and Vegetable Chains**
    Promotor(es): Prof.dr. G.W.J. Hendrikse
    Defended: June 13, 2002, ISBN: 90-5892-024-0

16. Linda H. Teunter, **Analysis of Sales Promotion Effects on Household Purchase Behavior**
    Promotor(es): Prof.dr.ir. B. Wierenga, Prof.dr. T. Kloek
    Defended: September 19, 2002, ISBN: 90-5892-029-1

17. Joost Loef, **Incongruity between Ads and Consumer Expectations of Advertising**
    Promotor(es): Prof.dr. W.F. van Raaij, Prof.dr. G. Antonides
    Defended: September 26, 2002, ISBN: 90-5892-028-3

18. Andrea Ganzaroli, **Creating Trust between Local and Global Systems**
    Promotor(es): Prof.dr. K. Kumar, Prof.dr. R.M. Lee
    Defended: October 10, 2002, ISBN: 90-5892-031-3

19. Paul C. van Fenema, **Coordination and Control of Globally Distributed Software Projects**
    Promotor(es): Prof.dr. K. Kumar
    Defended: October 10, 2002, ISBN: 90-5892-030-5

20. Dominique J.E. Delporte–Vermeiren, **Improving the flexibility and profitability of ICT-enabled business networks: an assessment method and tool.**
Promotor(es): Prof.mr.dr. P.H.M. Vervest, Prof.dr.ir. H.W.G.M. van Heck
Defended: May 9, 2003, ISBN: 90-5892-040-2

21. Raymond van Wijk, **Organizing Knowledge in Internal Networks. A Multilevel Study**
Promotor(es): Prof.dr.ing. F.A.J. van den Bosch
Defended: May 22, 2003, ISBN: 90-5892-039-9

22. Leon W.P. Peeters, **Cyclic Railway Timetable Optimization**
Promotor(es): Prof.dr. L.G. Kroon, Prof.dr.ir. J.A.E.E. van Nunen
Defended: June 6, 2003, ISBN: 90-5892-042-9

23. Cyriel de Jong, **Dealing with Derivatives: Studies on the role, informational content and pricing of financial derivatives**
Promotor(es): Prof.dr. C.G. Koedijk
Defended: June 19, 2003, ISBN: 90-5892-043-7

24. Ton de Waal, **Processing of Erroneous and Unsafe Data**
Promotor(es): Prof.dr.ir. R. Dekker
Defended: June 19, 2003, ISBN: 90-5892-045-3

25. Reggy Hooghiemstra, **The Construction of Reality**
Promotor(es): Prof.dr. L.G. van der Tas RA, Prof.dr. A.Th.H. Pruyn
Defended: September 25, 2003, ISBN: 90-5892-047-X

26. Marjolijn Dijksterhuis, **Organizational dynamics of cognition and action in the changing Dutch and U.S. banking industries**
Promotor(es): Prof.dr. F.A.J. van den Bosch, Prof.dr. H.W. Volberda
Defended: September 18, 2003, ISBN: 90-5892-048-8

Advanced Econometric Marketing Models

The present availability of large databases in marketing, concerning, for example, store-level sales or individual purchases, has led to an increased demand for appropriate econometric models to deal with these data. The typical database contains information on revealed preferences, measured by for example sales, market shares or brand choices. In this thesis we study econometric models for some of these series, to be precise, we consider market shares, purchase timing and brand choices. These models allow us to, for example, gain insight into the effect of marketing instruments on consumer behavior. Examples of topics we discuss are heterogeneity in decision processes and the development of easily interpretable models to capture dynamical features in market shares and interpurchase times. Additionally, we contribute to the econometric literature by extending and developing models and estimation techniques.

ERIM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are the Rotterdam School of Management and the Rotterdam School of Economics. ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focussed on the management of the firm in its environment, its intra- and inter-firm relations, and its business processes in their interdependent connections. The objective of ERIM is to carry out first rate research in management, and to offer an advanced graduate program in Research in Management. Within ERIM, over two hundred senior researchers and Ph.D. candidates are active in the different research programs. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.

The ERIM PhD Series contains Dissertations in the field of Research in Management defended at Erasmus University Rotterdam. The Dissertations in the Series are available in two ways, printed and electronical. ERIM Electronic Series Portal: http://hdl.handle.net/1765/1.

ERIM ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

www.research-in-management.nl/

ISBN 90-5892-049-6