ESSAYS ON QUANTITATIVE MARKETING MODELS AND MONTE CARLO INTEGRATION METHODS

ISBN 90 5170 991 9

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 348 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Essays on Quantitative Marketing Models and Monte Carlo Integration Methods

Essays over kwantitatieve marketing modellen en Monte Carlo integratie methoden

PROEFSCHRIFT

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

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 3 februari 2005 om 16:00 uur

door

Rutger Daniel van Oest
geboren te Leidschendam

Promotiecommissie

Promotoren:	Prof.dr. P.H.B.F. Franses
	Prof.dr. H.K. van Dijk
Overige leden:	Prof.dr. L. Bauwens
	Prof.dr. T.H.A. Bijmolt
	Prof.dr.ir. M.G. Dekimpe

Acknowledgements

Four years ago I started my PhD track at the Tinbergen Institute Rotterdam, finally resulting in this thesis. Several people have provided important guidance in the underlying research, have contributed to a pleasant working environment, and/or have offered some other form of support. Without them, realizing this thesis might not have been possible.

First of all, I would like to thank my two promoters Philip Hans Franses and Herman van Dijk. Philip Hans taught me academic efficiency, while Herman educated me with academic accuracy. Both aspects of doing research are very important without being mutually exclusive, of course. Philip Hans and Herman also offered me the input and enthusiasm necessary to finish this thesis. Additionally, credits should go to my other co-authors Luc Bauwens, Charles Bos, Lennart Hoogerheide and Richard Paap. In particular, I want to thank Richard, who seems to be an expert in almost every aspect of econometrics and is always willing to share his knowledge (and his other stories).

An important part of writing a thesis at the 16th floor of the H-building is the social component. I enjoyed the morning coffee breaks and the early afternoon lunches with discussion topics ranging from soccer and yesterday's television programs to politics and science. Thanks to Amy, Björn, Bram, Dennis, Erjen, Ivo, Jedid-Jah, Joost, Kevin, Klaas, Lenny, Michiel, Patrick, Pim, Reimer, Remco, René, Richard, Richard, Robin, Ward, Wilco, and several others. Furthermore, I would like to mention my appreciation for Lennart, who had to share a room with me for four years. Whenever our working hours coincided he was entertaining.

While working on my dissertation, I had the opportunity to present my research at several conferences in Canada, France, Germany, the Netherlands and the United States. Financial support from the Econometric Institute, the Tinbergen Institute and SAMSI is gratefully acknowledged. Furthermore I want to thank the staff at the Econometric and Tinbergen Institutes for excellent practical support. Also çok teşekkürler to Feray for her hospitality in Ann Arbor, Michigan, and for her contribution to the enjoyable holidays we had together.

Finally, I keep the last but most important "thank you" for my parents who have always supported me to succeed in the challenges I encounter.

Rutger van Oest
Rotterdam, September 2004

Contents

Chapter 1

Introduction

1.1 Motivation and structure of the thesis

The last few decades have led to an enormous increase in the availability of large detailed data sets and in the computing power needed to analyze such data. Additionally, new models and new computing techniques have been developed to exploit both sources. All of this has allowed for addressing research questions via analyses which were infeasible to carry out previously.

A discipline in which large data sets are quite common is marketing research. During the last twenty years, a vast literature has emerged in which scanner panel data sets are analyzed to get a better understanding of purchase behavior of households. Scanner panel data are obtained by recording the purchases and shopping trips of a large number of participating households over a long period of time. They contain observations on realized purchase behavior, together with all brand prices and other marketing-mix variables in the product category that households faced during their shopping trips. Hence, scanner panel data can be used to link purchase decisions to the promotional environment. By doing so, the effectiveness of, say, price promotions can be assessed.

The total effect of a price discount on own brand sales can be decomposed into several components. First, one can distinguish between changes in demand attributable to changes in each of the purchase timing, brand choice and purchase quantity decisions of households, see Gupta (1988). Here, purchase timing may be modeled as a "when to buy" decision by considering the interpurchase times of households, see Neslin *et al.* (1985), Gupta (1991), Jain and Vilcassim (1991) and Helsen and Schmittlein (1993). Alternatively, it can be represented by the "whether to buy" decision via purchase incidence, see Bucklin and Lattin (1991), Bucklin and Gupta (1992), Ailawadi and Neslin (1998) and Bucklin *et al.* (1998a). Wheat and Morrison (1990) compare both approaches and conclude that modeling purchase incidence is almost always preferable to modeling interpurchase time. As a second decomposition related to the duration of the promotion effect, one can distinguish between an immediate effect and an adjustment effect, see

Pauwels *et al.* (2002) who also consider a potentially permanent effect of price promotion but do not find evidence for it. A similar lack of evidence is reported by Dekimpe *et al.* (1999) and Nijs *et al.* (2001), suggesting that promotions usually only result in a *temporary* deviation in sales relative to the non-promoted sales level. The promotion effect during the period of deviation can be split up into an immediate effect and a subsequent adjustment effect lasting until the deviation in sales has vanished. The existence of adjustment effects may have far-reaching implications for the profitability of promotions. Chapter 2 and Chapter 3 in this thesis introduce two models for scanner panel data which can be thought of as dealing with such adjustment effects.

Scanner panel data provide detailed information about purchase behavior of households and allow for analyses which cannot be carried out using more aggregate data, typically store-level data containing the weekly marketing-mix and sales levels in a category in a store. However, for household scanner panel data there are some problems too. One of these is that scanner panel data are often expensive (in terms of effort and money) to acquire. A second problem is the representativeness of such data, depending on the size of the panel and on the way participating households have been selected. Store-level data provide less detail, but should be more representative, more widely available and much cheaper to collect. Hence, as a complement to the two scanner panel data models in Chapter 2 and Chapter 3, Chapter 4 presents a model for store-level data which can also be used to investigate the effectiveness of promotions on brand performance. Similar to the preceding two chapters, the model proposed in Chapter 4 accommodates for dynamics (although in a different way). These three quantitative marketing models constitute the first part of this thesis.

The second part of the thesis focusses on computing techniques developed to take advantage of the revolutionary increase in computing power of the last few decades. In particular, the discussed methods are suitable for Bayesian analysis in which model parameters are treated as random variables and one needs to get insight into the posterior density of these parameters. Chapter 5 provides an overview of some frequently used computing algorithms, and it also introduces a new methodology which extends current methods. To summarize, in this thesis we build on both the model/data developments and the computing techniques that have allowed for new computationally intensive analyses of relevant research questions. We intend to contribute to the literature by offering new models, new results, and new algorithms.

1.2 More detailed outline of the thesis

The outline of this thesis is as follows. In Chapter 2, we develop a utility-based model for purchase incidence, brand choice and purchase quantity. The model assumes that purchase behavior of households is rational, and it accounts for the possibility that households consume their inventory stock faster when the inventory level is higher. Inventory-driven

consumption can be expected for perishable products such as yogurt, see Ailawadi and Neslin (1998) and Bell *et al.* (1999). The existence of consumption acceleration may strongly affect the profitability of price promotions. If it would be largely absent, a promotion would still induce households to purchase more. However, this increase in current sales would mainly come at the expense of future sales in the category, as households purchase earlier but do not change their needs over time. On the other hand, such cannibalization effects would be much less if promotions (resulting in higher inventory levels) would also induce households to consume faster, effectively increasing demand in the category. Hence, consumption acceleration amounts to a positive adjustment effect following a promotion. It offers a possible explanation for the frequently reported lack of a postpromotion dip in store-level scanner data.

By applying our framework to a scanner panel data set on yogurt purchases, we find that the speed of consumption strongly depends on the inventory level. Furthermore, by running a simulation, we do find a dip in sales after a price promotion. However, this dip is very small, both in terms of duration and depth. Chapter 2 is a revised version of Van Oest *et al.* (2002b), and it is also partly based on Van Oest *et al.* (2002a).

In Chapter 3, we narrow our focus to the brand choice decision of households. The marketing literature provides much evidence that households do not only consider current prices when deciding which brand to buy in a category, but also take into account prices experienced in the past via the formation of internal reference prices. These anticipated prices "in the head of the consumer" serve as benchmarks for the current brand prices. Internal reference price formation implies that frequent price discounts may be beneficial in the short-run, but may also damage the brand in the long-run when households get used to these discounts and reference prices drop. Hence, such reference price effects amount to a negative adjustment after a promotion.

In Chapter 3, we propose a new reference price model for brand choice. By doing so, we attempt to bridge the gap between several survey studies consistently reporting that households have very limited price knowledge and hence may not always be able to construct a reference price, and current reference price models which do not account for this. Our model assumes that unobserved price recall of households evolves over time according to a first-order Markov process in which the "forgetting state" is absorbing. The reference price specification is flexible, allows for uncertainty in the unobserved process, allows that households may forget past prices, and even allows that households cannot construct a reference price at all.

We implement our model for a scanner panel data set on catsup purchases. We find that past prices are not frequently used for reference price formation. However, when a reference price is formed, its effect is stronger than suggested in the literature. By running a simulation, we find that the initial gain in sales due to a temporary price reduction is offset to a large extent by a decrease in subsequent sales. Chapter 3 is based on Van Oest and Paap (2004).

In Chapter 4, we develop a new model for store-level data which can be implemented to predict the market shares of brands and to link market shares with marketing instruments like price and promotion. The proposed model can be characterized as a Markov model, as it relates current market shares to previous market shares. It can be used to infer share-switching across brands when only store-level data are available, or to put it differently, it can address the managerially very relevant question "which brands gain share from which brands?". This provides new insights which cannot be obtained from the frequently considered market share attraction model. Additionally, the model structure allows for a decomposition of own and cross price elasticities. For example, a decrease of the own price is likely to result in an increase of the own market share. This increased market share is a result of (i) retaining more of the own share (less switching away from the own brand) and (ii) gaining additional share from the other brands (more switching towards the own brand). The decomposition from our model makes it possible to quantify these partial effects.

Application of our model to four store-level scanner data sets reveals some strong asymmetries in share-switching across brands. However, the resulting market shares seem balanced in the sense that for each brand the total share coming in and the total share going out are approximately equal. Furthermore, the price elasticity decompositions indicate that the retention effect is usually not the dominant factor in the reaction of market shares to price adjustments. Chapter 4 is based on Van Oest and Franses (2004).

In Chapter 5, we discuss Monte Carlo techniques which can be used in Bayesian analysis to get insight into the posterior density of model parameters. The basic idea behind the Monte Carlo approach is that an intractable posterior density can be investigated by generating a sample from it, and using the drawings to make inference. For example, the first moments of the collected sample may be reported as parameter estimates, whereas the (centered) second sample moments give an indication of parameter uncertainty. Unfortunately, most posterior densities do not have a standard form, so that sampling is not straightforward. Several techniques have been proposed to deal with such non-standard densities. However, they all have their specific problems. A common problem is that very large samples may be needed to obtain a representative coverage of the posterior density. Even worse, complete density regions might be overlooked and might never be sampled from. We start in Chapter 5 by providing an overview of some frequently applied methods. This overview draws from Van Oest *et al.* (2004). Furthermore, we put forward a new methodology in the remainder of the chapter. The corresponding discussion is based on Bauwens *et al.* (2004). Our approach extends current methods, and it has been designed to be flexible and robust with respect to the above mentioned problems. We consider three examples to demonstrate the benefits of our methods. One of the illustrations, the artificial example, has been borrowed from Bauwens *et al.* (2003).

Finally, we conclude in Chapter 6 with a summary and a brief discussion of the findings in this thesis.

Part I

Quantitative marketing models

Chapter 2

A model for purchase behavior with inventory-dependent consumption

2.1 Introduction and motivation

The immediate effect of a price discount on own brand sales can be decomposed into three components, that is, purchase time acceleration, brand switching and stockpiling, see Gupta (1988). Purchase time acceleration refers to the effect that the brand's promotion induces additional households to buy in the product category, brand switching means that the promoted brand may attract a larger fraction of the total number of buyers, and the stockpiling effect is due to households buying a larger quantity than they would have done without the price discount. Hence, the three components correspond to the household decisions on "whether to buy", "what to buy" and "how much to buy", respectively.

Purchase time acceleration and stockpiling have in common that they both result in larger inventory stocks for households. An important question concerning the profitability of price promotions is whether these increased inventory levels induce faster consumption, effectively increasing category demand, or whether consumption remains unaffected, so that category-wide promotional activity would merely result in earlier and larger purchases at the expense of future sales. Increased consumption after a price promotion is a possible explanation for the frequently reported lack of a postpromotion dip in store-level scanner data. Neslin and Schneider Stone (1996) provide an overview of existing theories.

The literature offers much support for the existence of consumption acceleration induced by a higher inventory level, see Folkes *et al.* (1993), Wansink and Deshpandé (1994), Ailawadi and Neslin (1998) and Chandon and Wansink (2002). Assuncão and Meyer (1993) conclude from a theoretical framework for optimal purchase and consumption behavior under Markovian price uncertainty that the consumption level should rationally increase with the inventory level. Bell *et al.* (1999) use the ratio of purchase quantity and subsequent interpurchase time as a proxy for the consumption rate. By comparing this ratio for promoted and non-promoted purchases, they find some evidence that promotions

may increase consumption. Nijs *et al.* (2001) find that product perishability (which can be expected to stimulate consumption acceleration) has a significant and positive effect on both the short-run and long-run effectiveness of price promotions. Bell *et al.* (2002) consider the effects of inventory-dependent consumption on the supply side of the market by developing a game theoretic model of price competition between firms. They conclude that such consumption causes more intense competition and lower profits.

The literature also provides several theories to motivate inventory-driven consumption. Folkes *et al.* (1993) focus on scarcity theory, stating that scarce products are perceived as more valuable, implying that consumption from a small inventory stock would be regarded more costly. Assuncão and Meyer (1993) take an economic perspective and consider the expected replacement costs of consumption. These costs are lower for households with higher inventory levels, as such households do not face serious stock-out constraints and therefore have the opportunity to wait for a good price deal. In contrast, a low inventory level might result in curbed consumption to avoid having to replenish at high prices. Wansink and Deshpandé (1994) relate the degree of consumption acceleration to usage-related salience, which is higher if products are perishable, can be used for multiple purposes, need refrigeration, and are stored in a visually prominent position. Chandon and Wansink (2002) find that a larger inventory stock increases consumption more for high-convenience products than for low-convenience products by triggering consumption incidence.

Although inventory-dependent consumption is an important phenomenon with potentially far-reaching implications for future sales and profitability, most scanner panel data models in the literature do no allow for this and assume that the consumption rate is fixed, see, for example, Neslin *et al.* (1985), Gupta (1988), Bucklin and Lattin (1991), and Bucklin *et al.* (1998a). Ailawadi and Neslin (1998) were the first to include an inventory-based consumption function in a model for purchase incidence, brand choice and purchase quantity. Their consumption function contains one inventory-sensitivity parameter, which has to be estimated from the data.

Several models proposed in the literature consider all three purchase decisions together, that is, purchase incidence (or purchase timing), brand choice and purchase quantity, see Gupta (1988) and Bucklin *et al.* (1998a), among others. However, besides not accounting for flexible consumption, most of these models do not allow that the purchase decisions might be interdependent (apart from the inclusion of common explanatory variables such as the "inventory level" and the "category value"). Krishnamurthi and Raj (1988) account for possible interdependencies between brand choice and purchase quantity by allowing the error terms to be correlated, but they do not consider the purchase incidence decision of households. Chiang (1991) and Chintagunta (1993) put forward models which capture all three purchase decisions and their interdependencies by assuming that these decisions result from a single utility maximization problem. However, both theory-based frameworks do not allow for increased consumption induced by a higher inventory level. Moreover, both models are based on the assumption that purchase quantities are con-

tinuous, while scanner panel data typically concern packaged consumer goods, which are only sold in discrete amounts.

In this chapter, we introduce a utility-based framework for purchase incidence, brand choice and purchase quantity, which does account for inventory-dependent consumption. The model can be used to infer the inventory-sensitivity of consumption and the joint effect of price discounts and consumption acceleration on future sales. Our model differs from the model of Ailawadi and Neslin (1998), as it is a utility-based framework for all three purchase decisions together, whereas Ailawadi and Neslin consider three separate models for purchase incidence, brand choice and purchase quantity. To put it differently, we intend to contribute to the literature by *integrating* inventory-dependent consumption and purchase behavior in a single model, while Ailawadi and Neslin only *include* an innovative consumption function in a standard model. In fact, it is demonstrated in Van Oest *et al.* (2002a) that the consumption function which is used in our model is more flexible and has an even better performance than the consumption function of Ailawadi and Neslin.

Our model also differs from the utility-based models of Chiang (1991) and Chintagunta (1993) in at least two respects. First, in our framework, utility derived from a purchase is determined by the increase in future consumption opportunities due to this purchase, and utility depends on the specific consumption path which is followed. In contrast, the models developed by Chiang and Chintagunta assume that utility is a direct function of the quantities purchased, and they do not consider a flexible consumption rate. Hence, the latter models do not link purchases and future consumption. The second difference is that our model explicitly takes into account that the quantities purchased are discrete and nonnegative for packaged consumer goods. The studies by Chiang and Chintagunta, which are based on the discrete choice/continuous quantity framework of Hanemann (1984), do not account for this.

The outline of this chapter is as follows. In Section 2.2, we put forward our model, and in Section 2.3, we discuss parameter estimation. We apply our model to an A.C. Nielsen scanner panel data set on yogurt purchases in Section 2.4. Finally, in Section 2.5, we conclude with a discussion of the implications and limitations of our model, and we provide some directions for further research.

2.2 The model

In this section, we develop our framework for purchase behavior and inventory-dependent consumption. We first define the inventory process. Next we derive the consumption function and the purchase conditions. Finally, we present the reduced form of our model.

2.2.1 Available inventory

The inventory level $S_{i,\tau}$ for household i at day τ is determined by the recursion

$$S_{i,\tau} = S_{i,\tau-1} + wQ_{i,\tau-1} - C_{i,\tau-1}, \tag{2.1}$$

where $Q_{i,\tau-1}$ is the number of units purchased in the category at day $\tau - 1$, $C_{i,\tau-1}$ is the consumption level for that day, and w is the package size for the product category expressed in inventory/consumption units. Hence, inventory is updated daily by adding a possible purchase and subtracting the consumed amount.

2.2.2 Consumption

For the consumption process, we assume that households are forward-looking utility maximizers with a finite planning horizon. Each household chooses its daily consumption levels such that total discounted utility within the planning period is maximal. As households cannot consume more than the size of their inventory stock, they have an inventory restriction. We note that if households would already know when to make a purchase in the category again, this point in time should define the planning horizon. However, although households may know in advance when to make their next *shopping trip* (for example, next Saturday), they may not know when their next *category purchase* will occur. In particular, this should hold for products which are not purchased on a regular basis. In such a case, the next purchase occasion is unknown, and hence it cannot be used to define the consumption planning horizon. On the other hand, the time of occurrence of the next shopping trip is not suitable either, as this would imply that households never hold any inventory when they go to the shop. This is implausible and would result in purchase models without any inventory effect, as the "whether to buy" decision is usually modeled conditional on a shopping occasion, see Bucklin and Lattin (1991), Chintagunta (1993), Ailawadi and Neslin (1998), Bucklin *et al.* (1998a), among many others. We return to the discussion on the planning horizon later on.

In the consumption model, we assume that households have Constant Relative Risk Aversion [CRRA] utility, that is, household i derives instantaneous utility

$$u(C_{i,\tau}) = \frac{C_{i,\tau}^{1-\eta}}{1-\eta} \qquad \text{with} \qquad 0 < \eta < 1 \tag{2.2}$$

from consuming $C_{i,\tau}$ units at day τ. This utility function involves one curvature (or concavity) parameter η, and it is, for example, described in Romer (1996, p. 40). The household's dynamic utility maximization problem for current and future consumption is

defined by

$$\max_{\{C_{i,\tau+h}\}_{h=0}^{H_{i,\tau}}} \sum_{h=0}^{H_{i,\tau}} \frac{1}{(1+\rho)^h}\, u(C_{i,\tau+h}),$$
$$\text{subject to} \quad \sum_{h=0}^{H_{i,\tau}} C_{i,\tau+h} \leq S_{i,\tau}\,,$$

where $H_{i,\tau}$ is the planning horizon at day τ, and $\rho > 0$ is a discount rate for time. It can be shown that the optimal consumption path is given by

$$C_{i,\tau} = S_{i,\tau}\, \frac{1-\nu}{1-\nu^{H_{i,\tau}+1}}\,, \tag{2.3}$$
$$C_{i,\tau+h} = \nu^h\, C_{i,\tau}, \qquad h = 1, \ldots, H_{i,\tau}, \tag{2.4}$$

where $\nu = (1+\rho)^{-\frac{1}{\eta}}$ is the consumption dampening factor, see Appendix 2.A. For a given inventory level $S_{i,\tau}$, this path provides total discounted utility

$$U(S_{i,\tau}) = \frac{S_{i,\tau}\, C_{i,\tau}^{-\eta}}{1-\eta}\,. \tag{2.5}$$

We complete our inventory-dependent consumption function by assuming a parametric form for the planning horizon $H_{i,\tau}$. We have already discussed why the amount of time until the next shopping trip or next purchase occasion are not suitable candidates. As a reasonable alternative, we assume that the planning horizon is proportional to the amount of time it takes for a household to deplete its current inventory stock at its average consumption rate. Hence, our planning horizon is defined as

$$H_{i,\tau} = \exp(\delta)\, \frac{S_{i,\tau}}{\overline{C}_i}\,, \tag{2.6}$$

where δ is the proportionality parameter and $\overline{C}_i$ is the average consumption rate of household i. For most product categories, we expect that $H_{i,\tau}$ is much smaller than the inventory depletion time $(S_{i,\tau}/\overline{C}_i)$, as the household's average consumption rate $\overline{C}_i$ also includes periods in which the household is out of stock. We do not claim that (2.6) is an accurate reflection of how far households plan ahead, but we believe that it can be used as a rough approximation to the expected time until the next purchase, given the household's current inventory level. Moreover, our assumed planning horizon results in a consumption function with many desirable analytical properties. By substituting (2.6) into (2.3), we obtain our final consumption function, that is,

$$C_{i,\tau} = S_{i,\tau}\, \frac{1-\nu}{1-\nu^{\frac{\exp(\delta) S_{i,\tau} + \overline{C}_i}{\overline{C}_i}}} \qquad \text{with} \qquad \nu = (1+\rho)^{-\frac{1}{\eta}}. \tag{2.7}$$

Together with (2.1), it describes how inventory and consumption interact with each other. The properties of our consumption function are as follows, that is,

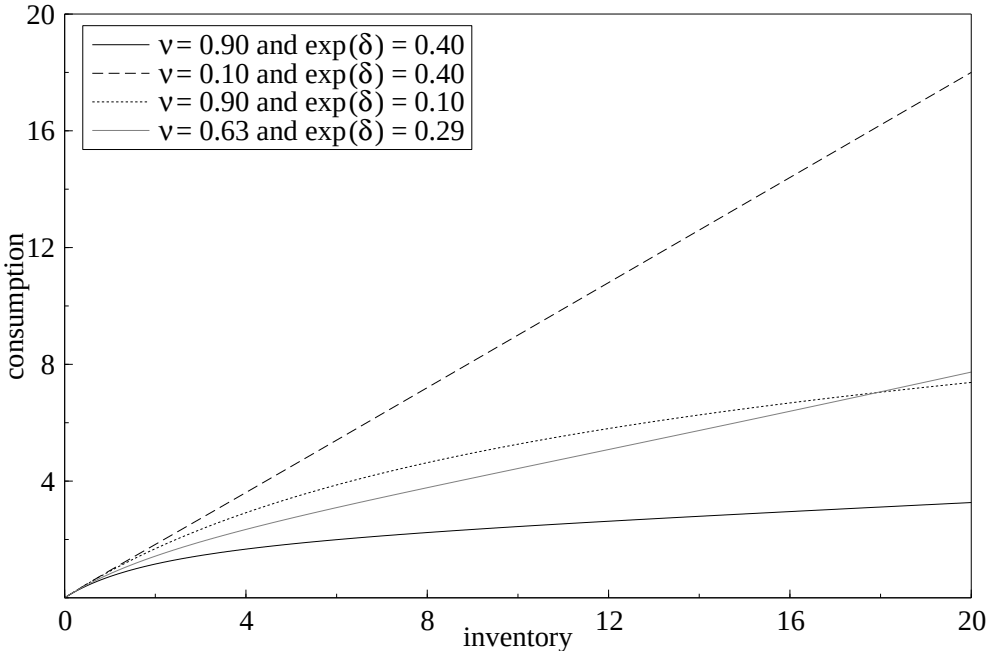

Figure 2.1: Illustration of some consumption patterns which can be reproduced by our model. The average consumption rate has been standardized to one.

- $C_{i,\tau}$ is always positive, but it never exceeds the inventory level $S_{i,\tau}$,
- $C_{i,\tau}$ is increasing in the inventory level $S_{i,\tau}$,
- $C_{i,\tau}$ is increasing in the household's average consumption rate $\overline{C}_i$,
- $C_{i,\tau}$ is homogeneous of degree one with respect to the inventory level $S_{i,\tau}$ and the average consumption rate $\overline{C}_i$, implying that the consumption process is invariant with respect to the unit of measurement (such as ounces, pounds, or even multiples of the average daily consumption rate).

Figure 2.1 displays some consumption patterns which can be reproduced by tuning the dampening factor ν and the proportionality factor $\exp(\delta)$ in (2.7). The average consumption rate $\overline{C}_i$ has been standardized to one. The solid line in the graphs mimics fixed consumption, the dashed line corresponds to a pattern such that almost the entire inventory stock is consumed at once, and the remaining two lines are patterns in between these two extreme cases. It can be seen that our consumption function is quite flexible. In Van Oest *et al.* (2002a), it is demonstrated that (2.7) performs better than a consumption function put forward by Ailawadi and Neslin (1998). However, they were the first to include an inventory-dependent consumption process in a model for purchase incidence, brand choice and purchase quantity.

Finally, although our consumption function is flexible and has nice analytical properties, we admit that it still describes a stylized process. Nonetheless, it has been derived

from economic principles and it allows us to integrate inventory-driven consumption and purchase behavior in a single model in a consistent way. A crucial implication of (2.7) for the development of our purchase model is that the maximum utility level (2.5), which can be achieved from consuming the inventory stock $S_{i,\tau}$, is concave with respect to $S_{i,\tau}$. This follows from a tedious derivation.

2.2.3 Purchase decision

Without loss of generality, we impose that at shopping days the household's consumption occurs prior to its purchase decision. Furthermore, we use a new time index t for the days at which a shopping trip is made. So, we have two time indices, that is, t refers to shopping trips and τ refers to calendar time. We assume that purchasing $q \in \{0, 1, \ldots, Q\}$ units of brand $j \in \{1, \ldots, J\}$ by household $i \in \{1, \ldots, N\}$ at shopping occasion $t \in \{1, \ldots, T_i\}$ provides net utility

$$\psi_{i,j,t|z}\Big[U(S_{i,t} - C_{i,t} + wq) - U(S_{i,t} - C_{i,t})\Big] - P_{i,j,t}\, q, \tag{2.8}$$

where $P_{i,j,t}$ is the price observed for brand j, $\psi_{i,j,t|z}$ is the perceived quality of the brand, and w is the package size for the product category expressed in inventory/consumption units. As different households might perceive different brands differently, the brand quality $\psi_{i,j,t|z}$ is allowed to vary across households. This brand quality might depend on both observed variables and unobserved factors, denoted by z and discussed in more detail later on. It follows from (2.8) that net utility from a purchase is the difference between the increase in consumption utility under optimal consumption planning and the expenditure. Hence, the household's purchase decision is based on the future consumption opportunities that a purchase would offer. The consumption utility part of (2.8) consists of two components, that is, a "core utility" for consuming in the product category and a brand-specific multiplier to account for differences across brands. For notational convenience, we rewrite (2.8) as

$$\psi_{i,j,t|z}\Big[\, V_{i,t}(q) - V_{i,t}(0)\,\Big] - P_{i,j,t}\, q, \tag{2.9}$$

where

$$V_{i,t}(q) \equiv U(S_{i,t} - C_{i,t} + wq). \tag{2.10}$$

Partly following Hanemann (1984), Chiang (1991) and Chintagunta (1993), we specify the brand quality $\psi_{i,j,t|z}$ in (2.8) as

$$\psi_{i,j,t|z} = \exp(\alpha_{j|z} + \beta_1 BL_{i,j,t} + \beta_2 PM_{i,j,t} + \beta_3 \ln(\overline{C}_i) + \varepsilon_{i,j,t}), \tag{2.11}$$

where the random disturbance $\varepsilon_{i,j,t}$ is assumed to be independently and identically distributed obeying a Type-I Extreme Value distribution with scaling parameter $\vartheta > 0$. Hence, $\varepsilon_{i,j,t}$ has cumulative density

$$F(\varepsilon_{i,j,t}) = \exp(-\exp(-\vartheta\, \varepsilon_{i,j,t})), \tag{2.12}$$

see, for example, also Allenby and Rossi (1991). In (2.11), $\alpha_{j|z}$ captures the intrinsic brand preferences of the household, $BL_{i,j,t}$ is a variable to deal with observed heterogeneity, such as the brand loyalty measure of Guadagni and Little (1983), $PM_{i,j,t}$ is a 0/1 promotion indicator (feature or display), and $\overline{C}_i$ is the household's average consumption rate. The intrinsic preferences are allowed to vary across households by assuming that households can be grouped into an unknown number of unobserved segments. Households within the same segment have the same intrinsic preferences, whereas households belonging to different segments have different preferences. Basically, this is the segmentation approach of Kamakura and Russell (1989). Chintagunta *et al.* (1991) and Chintagunta (1993) also allow the intrinsic preferences to be different across households.

We note that (2.5), (2.7), (2.8) and (2.11) together imply that our purchase model is invariant with respect to the unit of measurement for inventory and consumption. For example, the processes for inventory and consumption may be defined in ounces, pounds, or even in multiples of the household's average daily consumption rate $\overline{C}_i$, so that the original $\overline{C}_i$ is standardized to one for each household[1], see Gupta (1991) and Neslin and Schneider Stone (1996). The invariance holds, as the intrinsic preferences $\alpha_{j|z}$ and the parameter β_3 in (2.11) can fully adjust to such scaling.

The first decision a household has to make at a shopping occasion is whether to purchase in the category. Clearly, not buying is optimal if this gives a higher net utility level than buying any positive quantity from any of the available brands, or equivalently, any positive choice/quantity combination should result in negative net utility. This would imply that many conditions have to be tested to determine the purchase incidence outcome. However, as the concavity property of the utility function U carries over to net utility (2.9) as a function of purchase quantity q, a necessary and sufficient condition for not purchasing is that, for all brands, not buying is preferable to buying one unit. Hence, the no-purchase condition translates into

$$\frac{\psi_{i,k,t|z}\,[\,V_{i,t}(1) - V_{i,t}(0)\,]}{P_{i,k,t}} < 1, \qquad k = 1, \ldots, J, \tag{2.13}$$

stating that one dollar kept in pocket (having value 1) provides more utility than one dollar spent on any of the brands (still assuming that an integer amount is purchased).

The second decision, the brand choice decision, is conditional on purchase incidence. So, it is given that one dollar spent on the selected brand offers more value than one dollar kept in pocket. However, it would still be suboptimal to choose that brand if there would exist another brand, providing even more utility for the same dollar. Hence, household i selects brand j at shopping trip t if and only if

$$\frac{\psi_{i,j,t|z}\,[\,V_{i,t}(\tilde{q}) - V_{i,t}(0)\,]}{P_{i,j,t}\,\tilde{q}} > \frac{\psi_{i,k,t|z}\,[\,V_{i,t}(\tilde{q}) - V_{i,t}(0)\,]}{P_{i,k,t}\,\tilde{q}}, \qquad k \neq j, \tag{2.14}$$

[1]However, in this case, the variable $\overline{C}_i$ showing up in the brand quality specification (2.11) is still required to be given in ounces, pounds, or some other household-invariant size. The reason is that here $\overline{C}_i$ is also used to distinguish between light and heavy users, so that standardization would affect the estimation results.

where $\tilde{q} \geq 1$.

The third decision, the purchase quantity decision, is conditional on both purchase incidence and brand choice. Clearly, it is optimal to buy q units of brand j if and only if this gives a higher net utility level than buying any other positive amount. As for the purchase incidence outcome, many conditions would have to be checked to verify that q is indeed the optimal quantity. Fortunately, we can again take advantage of the concavity property of the utility function U.

The net utility function (2.9) implies that household i prefers purchasing $\tilde{q} \geq 1$ units of brand j to purchasing $\tilde{q} - 1$ units if and only if

$$\frac{\psi_{i,j,t|z}\,[\,V_{i,t}(\tilde{q}) - V_{i,t}(\tilde{q}-1)\,]}{P_{i,j,t}} > 1, \tag{2.15}$$

meaning that a dollar spent on the $\tilde{q}$-th unit should offer more value than a dollar kept in pocket. As U is a concave function and hence $V_{i,t}$ is concave with respect to $\tilde{q}$, the left-hand side of (2.15) is decreasing in $\tilde{q}$. The optimal strategy therefore consists of purchasing additional units as long as the consumption value of the dollar spent on the marginal unit is larger than its monetary value. This amounts to the following two optimality conditions for q, that is,

$$\frac{\psi_{i,j,t|z}\,[\,V_{i,t}(q) - V_{i,t}(q-1)\,]}{P_{i,j,t}} > 1, \tag{2.16}$$

$$\frac{\psi_{i,j,t|z}\,[\,V_{i,t}(q+1) - V_{i,t}(q)\,]}{P_{i,j,t}} < 1. \tag{2.17}$$

In our model, we consider an upper bound Q for the quantity purchased, as in practice households have such an upper bound. For example, extremely large purchase quantities may result in transportation and storage problems. Moreover, scanner panel data sets also suggest that households never buy above a certain level. By assuming an upper bound Q, we effectively create a preference restriction that buying Q units is always preferred to buying $Q+1$ units, that is,

$$\frac{\psi_{i,k,t|z}\,[\,V_{i,t}(Q+1) - V_{i,t}(Q)\,]}{P_{i,k,t}} < 1, \qquad k = 1, \ldots, J. \tag{2.18}$$

2.2.4 Reduced form

In this subsection, we present the reduced form of our model. A detailed derivation can be found in Appendix 2.B. We define $Y_{i,t}$ as a 0/1 purchase indicator for household i at shopping trip t. Moreover, if a purchase occurs, $B_{i,t} \in \{1, \ldots, J\}$ denotes the brand choice and $Q_{i,t} \in \{1, \ldots, Q\}$ denotes the purchase quantity. The no-purchase probability

for household i (belonging to segment z) at shopping trip t is given by

$$\Pr(Y_{i,t}=0|z) = \exp\Bigg(\sum_{k=1}^{J} \exp\Big(\mu_{i,k,t|z} + \vartheta \ln\Big[V_{i,t}(Q+1) - V_{i,t}(Q)\Big] \Big) - \sum_{k=1}^{J} \exp\Big(\mu_{i,k,t|z} + \vartheta \ln\Big[V_{i,t}(1) - V_{i,t}(0)\Big] \Big) \Bigg), \tag{2.19}$$

where

$$\mu_{i,k,t|z} \equiv \alpha_{k|z} + \beta_1 BL_{i,k,t} + \beta_2 PM_{i,k,t} + \beta_3 \ln(\overline{C}_i) - \vartheta \ln(P_{i,k,t}). \tag{2.20}$$

Two implications of (2.19) are that the no-purchase probability increases when the inventory level $S_{i,t}$ increases, whereas it decreases when the category value $\ln(\sum_{k=1}^{J} \exp(\mu_{i,k,t|z}))$ increases, that is, when the category becomes more attractive, see Ben-Akiva and Lerman (1985). Both relationships are plausible.

The conditional brand choice probability directly follows from the logit framework developed by McFadden (1974). It is given by

$$\Pr(B_{i,t}=j|Y_{i,t}=1,z) = \frac{\exp(\mu_{i,j,t|z})}{\sum_{k=1}^{J} \exp(\mu_{i,k,t|z})}, \tag{2.21}$$

where the brand-invariant $\beta_3 \ln(\overline{C}_i)$ term in $\mu_{i,k,t|z}$ cancels out. The subsequent probability that q units are purchased turns out to be

$$\begin{aligned} &\Pr\Big(Q_{i,t}=q \Big| Y_{i,t}=1, B_{i,t}=j, z\Big) \\ &= \Big(\exp\big(-\exp\big(\mu_{i,j,t|z} + \vartheta \ln\big[V_{i,t}(q+1) - V_{i,t}(q) \big] \big) \big) \\ &\qquad - \exp\big(-\exp\big(\mu_{i,j,t|z} + \vartheta \ln\big[V_{i,t}(q) - V_{i,t}(q-1) \big] \big) \big) \Big) \Big/ \\ &\Big(\exp\big(-\exp\big(\mu_{i,j,t|z} + \vartheta \ln\big[V_{i,t}(Q+1) - V_{i,t}(Q) \big] \big) \big) \\ &\qquad - \exp\big(-\exp\big(\mu_{i,j,t|z} + \vartheta \ln\big[V_{i,t}(1) - V_{i,t}(0) \big] \big) \big) \Big). \end{aligned} \tag{2.22}$$

We note that the quantity purchased is allowed to depend on the brand loyalty variable $BL_{i,j,t}$. Households might tend to purchase more units of the selected brand if this brand is relatively familiar, see Bucklin *et al.* (1998a).

For completeness, we also recall some expressions and equalities, which we have obtained earlier and which are needed to evaluate the probabilities above, that is,

$$V_{i,t}(q) = U(S_{i,t} - C_{i,t} + wq), \tag{2.23}$$

$$U(S_{i,\tau}) = \frac{S_{i,\tau}\, C_{i,\tau}^{-\eta}}{1-\eta}, \tag{2.24}$$

$$C_{i,\tau} = S_{i,\tau} \frac{1-\nu}{1-\nu^{\frac{\exp(\delta) S_{i,\tau} + \overline{C}_i}{\overline{C}_i}}} \quad \text{with} \quad \nu = (1+\rho)^{-\frac{1}{\eta}}, \tag{2.25}$$

$$S_{i,\tau} = S_{i,\tau-1} + wQ_{i,\tau-1} - C_{i,\tau-1}. \tag{2.26}$$

As a possible extension of the model, one can allow for different package sizes of the various brands. By doing so, the weights w in (2.23) and (2.26) become brand-specific, and the brand choice probability becomes (heuristically)

$$\Pr(B_{i,t}=j|Y_{i,t}=1,z)=\frac{\exp(\,\mu_{i,j,t|z}+\vartheta\ln[\,U(S_{i,t}-C_{i,t}+w_j)-U(S_{i,t}-C_{i,t})\,]\,)}{\sum_{k=1}^{J}\exp(\,\mu_{i,k,t|z}+\vartheta\ln[\,U(S_{i,t}-C_{i,t}+w_k)-U(S_{i,t}-C_{i,t})\,]\,)}\,, \tag{2.27}$$

where it is important that $P_{i,k,t}$ is the shelf price per purchase unit, not per ounce. The probabilities (2.19) and (2.22) for purchase incidence and purchase quantity do not change.

2.3 Parameter estimation

The parameters of our model can be estimated using maximum likelihood [ML]. The parameter estimates result from maximization of the log-likelihood function, which is given by

$$\ln\mathcal{L}=\sum_{i=1}^{N}\ln\!\left(\sum_{z=1}^{Z}\pi_z\,\mathcal{L}_{i|z}\right), \tag{2.28}$$

where π_z is the size of segment z, and $\mathcal{L}_{i|z}$ is the likelihood for household i given segment membership z, that is,

$$\begin{aligned}\mathcal{L}_{i|z}=\prod_{t=1}^{T_i}\Bigg[&\Big[\Pr(Y_{i,t}=1|z)\Big]^{y_{i,t}}\Big[\Pr(Y_{i,t}=0|z)\Big]^{(1-y_{i,t})}\\&\Bigg[\prod_{j=1}^{J}\Big(\Pr(B_{i,t}=j|Y_{i,t}=1,z)\Pr(Q_{i,t}=q_{i,t}|Y_{i,t}=1,B_{i,t}=j,z)\Big)^{I\{b_{i,t}=j\}}\Bigg]\Bigg],\end{aligned} \tag{2.29}$$

In (2.29), $y_{i,t}$, $b_{i,t}$ and $q_{i,t}$ denote the realizations of $Y_{i,t}$, $B_{i,t}$ and $Q_{i,t}$, respectively, and $I\{\cdot\}$ is the 0/1 indicator function.

Standard errors are obtained by taking the square roots of the diagonal elements of the estimated parameter covariance matrix, which, in turn, can be computed as minus the inverse of the Hessian of (2.28) evaluated for the optimal parameter values. Numerical techniques, such as the BFGS algorithm or the Newton-Raphson algorithm, have to be used to get the ML parameter estimates.

2.4 Empirical analysis

In this section, we apply our purchase model with inventory-dependent consumption to an A.C. Nielsen scanner panel data set on yogurt purchases in the Sioux Falls SD market. For this product category, we expect substantial consumption acceleration after a price promotion, see Ailawadi and Neslin (1998) and Bell *et al.* (1999). In the analysis, we consider five brands, which together account for about 75% of the category sales in units.

These brands are Dannon, Nordica, W.B.B., Weight Watchers and Yoplait. The sample period consists of 92 weeks from November 1986 to August 1988. The first 46 weeks are used for initialization purposes and are discarded in the log-likelihood evaluation, while the remaining 46 weeks are used for either parameter estimation or out-of-sample model validation. Only households which limited their purchases to the five brands mentioned above, made at least one shopping trip every four weeks, and made at least three purchases in both the initialization and the estimation period are considered in the analysis. The estimation sample contains 80% of these households. The remaining 20% is assigned to a hold-out sample. The total sample consists of 171 households who made 16660 shopping trips, resulting in 3935 purchases in the yogurt category.

We measure purchase quantities in multiples of six ounces. They range from 1 to $Q = 17$ units. Furthermore, we define the household's average consumption rate $\overline{C}_i$ as the number of units purchased in the initialization period divided by the number of days, and we initialize the household's inventory process by assuming that any household starts with an inventory level equal to its average purchase quantity in the initialization period. However, the estimation results are not sensitive to this initialization condition, as the inventory process "renews" itself as soon as the level becomes (virtually equal to) zero. For all households in the sample, such renewal occurs at least once in the initialization stage.

2.4.1 Parameter estimates

We consider two variants of our model, that is, the model without unobserved heterogeneity and the model with an unknown number of latent segments to capture unobserved heterogeneity. For the latter variant, we use the AIC-3 criterion, proposed by Bozdogan (1994), to determine the optimal number of segments. This measure is defined as $-2\ln\mathcal{L} + 3K$ with $\mathcal{L}$ being the maximum likelihood value and K being the number of parameters to be estimated. The model corresponding to the smallest AIC-3 value is selected. The AIC-3 criterion suggests more parsimonious models than the standard AIC criterion, but if the number of observations exceeds 20 it is less parsimonious than BIC, the other popular measure in the literature. An extensive simulation study by Andrews and Currim (2003) indicates that AIC-3, applied to multinomial choice data, performs better than several other criteria including AIC and BIC. For our model, we find five segments (with different intercept parameters) in the data set. The parameter estimates are reported in Table 2.1.

For both the model without unobserved heterogeneity and the five-segments model, all response parameters have the expected sign. Furthermore, almost all effects are significant at the 1% level. The only exception is the non-price promotion parameter β_2 in the latent segments model, which is however still significant at the 5% level. The most important difference between the estimation results before and after accounting for unobserved heterogeneity concerns the price parameter ϑ. This parameter increases from 1.70 to 2.69,

Table 2.1: Parameter estimates for our model. The estimated standard errors are given in parentheses.

	no heterogeneity		5-segments model	
β_1 (brand loyalty)	0.324^{***}	(0.008)	0.305^{***}	(0.011)
β_2 (promotion)	0.068^{***}	(0.020)	0.043^{**}	(0.020)
β_3 (consumption rate)	0.491^{***}	(0.034)	0.245^{***}	(0.048)
ϑ (price)	1.702^{***}	(0.439)	2.689^{***}	(0.445)
ρ (discount rate)	0.021^{**}	(0.009)	0.024^{***}	(0.006)
η (curvature)	0.034^{**}	(0.016)	0.050^{***}	(0.013)
δ (horizon)	-1.424^{***}	(0.167)	-1.232^{***}	(0.127)

** significant at 5%.
*** significant at 1%.

indicating that the model without unobserved heterogeneity severely underestimates the price effect. This finding is consistent with the studies by Chintagunta *et al.* (1991) and Jain *et al.* (1994).

The estimates of the discount rate ρ and the curvature parameter η indicate that households are not very willing to postpone consumption. For the five-segments model, it is implied that the consumption dampening factor $\nu = (1+\rho)^{-\frac{1}{\eta}}$ is 0.63 and the proportionality factor for the planning horizon equals 0.29. Figure 2.1 in Subsection 2.2.2 contains the corresponding consumption function. It shows that consumption is clearly increasing in the inventory level. As $\exp(\delta)$ is substantially smaller than one, the consumption planning horizon $H_{i,\tau}$ is much shorter than the inventory depletion time $(S_{i,\tau}/\overline{C}_i)$.

2.4.2 Model performance

To see how well our model performs for the considered data set, we report the hit rates for the purchase incidence and brand choice components, and we show the Root Mean Squared Prediction Error [RMSPE] for the purchase quantity component. These numbers are given in the top part of Table 2.2. In both the estimation sample and the validation sample, the hit rate for incidence is higher than 70%, and the hit rate for choice is higher than 75%. Furthermore, the RMSPE values for purchase quantity are smaller than three.

The bottom part of Table 2.2 contains the prediction-realization tables underlying the hit rates for purchase incidence. These tables provide a more complete view, as they take into account that it is much more difficult to predict the few purchase occasions in a typical scanner panel data set correctly than to explain the many non-purchases. The hit rate measure does not recognize this. For example, for our data set, one can obtain a

Table 2.2: Empirical performance of our model, represented by the prediction-realization table and corresponding hit rate for purchase incidence, the hit rate for brand choice, and the Root Mean Squared Prediction Error for purchase quantity.

	in-sample		out-of-sample	
hit rate for incidence (in %)	73.7		71.9	
hit rate for brand choice (in %)	77.1		78.5	
RMSPE for purchase quantity	2.91		2.40	
	purchase predicted		purchase predicted	
	no	yes	no	yes
purchase observed: no	58.1	16.8	60.9	18.5
purchase observed: yes	9.5	15.6	9.7	11.0

Table 2.3: Empirical performance of the habit-persistence model, represented by the prediction-realization table and corresponding hit rate for purchase incidence, the hit rate for brand choice, and the Root Mean Squared Prediction Error for purchase quantity.

	in-sample		out-of-sample	
hit rate for incidence (in %)	72.0		72.1	
hit rate for brand choice (in %)	72.5		73.6	
RMSPE for purchase quantity	2.59		2.10	
	purchase predicted		purchase predicted	
	no	yes	no	yes
purchase observed: no	60.7	14.1	65.4	14.0
purchase observed: yes	13.9	11.3	14.0	6.7

hit rate of 75% by claiming that households never buy in the category, which is of course very unrealistic. It can be seen from the table that actual purchase occasions are classified correctly at a rate of 62% in-sample, whereas this rate is 53% out-of-sample.

As a benchmark for our model, we consider a model featuring habit-persistence in the sense that the next prediction for each of the three purchase decisions coincides with the most recent outcome. The performance measures for this model are shown in Table 2.3. It can be seen that our model clearly outperforms the habit-persistence model as far as it concerns the purchase incidence and brand choice components. In particular, this holds for the classification of actual purchase occasions as such. The habit-persistence model correctly classifies 45% of the purchases in the estimation sample, whereas this classification rate is only 32% in the validation sample. However, the purchase quantity

Table 2.4: The estimated own price elasticities for purchase incidence, brand choice and purchase quantity.

	incidence	choice	quantity	total
Dannon	−0.35	−2.24	−0.03	−2.62
Nordica	−0.80	−1.66	−0.06	−2.52
W.B.B.	−0.36	−2.22	−0.02	−2.60
Weight Watchers	−0.14	−2.18	−0.01	−2.33
Yoplait	−0.58	−1.70	−0.05	−2.33

component of our model performs worse, as the RMSPE values for the benchmark model are about 12% lower. Overall, we conclude that our model has a reasonable performance.

2.4.3 Price elasticity analysis

In this subsection, we consider the immediate effects of price changes on sales. Gupta (1988) shows that the overall price elasticity for a brand can be decomposed into the sum of elasticities concerning purchase incidence, brand choice and purchase quantity separately. This convenient decomposition is frequently applied in the literature, see, for example, Bucklin *et al.* (1998a) and Bell *et al.* (1999). The results for our model, applied to the data set on yogurt purchases, are reported in Table 2.4. It can be seen that the majority of the increase in own brand sales induced by a 1% price discount is caused by brand switching. Furthermore, the effect of purchase time acceleration is still quite important. However, the effect of stockpiling by purchasing larger quantities is only marginal. The price elasticities suggest a 22/76/2 decomposition for incidence/choice/quantity. It is, however, important to note that this decomposition, with brand switching accounting for 76%, does not imply that the other brands lose 76 units if the promoted brand would gain 100 units. The reason is that the elasticity decomposition does not consider the increase in category sales induced by the brand's promotion. Such category expansion would also benefit the non-promoted brands, making the incurred loss substantially smaller than 76 units. This is pointed out by Van Heerde *et al.* (2003) who propose a primary/secondary demand decomposition based on unit sales.

2.4.4 Impact of increased consumption on future sales

In the previous subsection, we have considered the immediate effects of price changes. Here, we extend the analysis by considering the impact on future sales. The results obtained so far indicate that consumption increases substantially after a price promotion. First, households react to price discounts in the category by purchasing more, hence increasing their inventory levels. Second, they react to the larger inventory stock at home

by consuming it faster. An important question that remains is whether this consumption acceleration following a price promotion is sufficient for brands to avoid a clear sales dip in the subsequent weeks.

To answer this question, we perform a simulation study in which we consider the effects of a temporary category-wide 10% price decrease on current and future category sales. The price cut has a duration of only one week, and it applies to all five brands. By comparing the weekly sales levels for this price-cut setting with the sales levels that would result when all prices are kept at their original level, one can get insight into the current and future effects of temporary category-wide price discounts.

For each week after the 46-weeks initialization period for which at least four subsequent weeks are available in the data set, we impose a temporary 10% price cut, while keeping the prices in the other weeks at their original level. We simulate the purchases for all households in the estimation sample, and we add the expected purchase quantities for all shopping trips occurring in the promotion week. This is done for all 42 weeks for which we impose the temporary promotion, resulting in the total current sales level. At the same time and analogously, we compute the total sales levels for each of the four weeks following the promotion week. We perform ten repetitions of this procedure in which the segment memberships of households are drawn from the corresponding posterior distributions, and we add the sales values obtained from all runs. In a similar way, we also obtain the current and future sales levels for the case that prices are not adjusted.

The results of the simulation are as follows. A temporary 10% price reduction increases current category demand by 25.22%. However, during the next week in which prices have returned to their original level, demand is 0.24% lower than it would have been without the past price discount. The promotion effect does not exist anymore for subsequent weeks. Hence, for the yogurt category, we find a very short and small dip in sales after a price promotion. This "postpromotion dip" only lasts for one week, and it is about 1% of the current sales effect.

2.5 Conclusions

In this chapter, we have developed a utility-based model for purchase incidence, brand choice and purchase quantity, which accounts for inventory-driven consumption. By doing so, we have combined a flexible consumption specification, as in Ailawadi and Neslin (1998), with a rational framework for purchase behavior, as in Chiang (1991) and Chintagunta (1993). Inventory-dependent consumption is an important phenomenon, as its existence would imply that price promotions are effective tools to increase current category demand without serious cannibalization of future demand. Such consumption also provides a possible explanation for the frequently reported lack of a postpromotion dip in store-level scanner data. Besides developing a new purchase model, we have introduced a flexible consumption function, which can be used to capture consumption acceleration in a broad class of models for purchase incidence and purchase quantity.

Applied to A.C. Nielsen scanner panel data on yogurt purchases, our model indicates that the speed of consumption strongly depends on the inventory level. This is consistent with other studies in the literature. Furthermore, a decomposition of the price elasticities suggests that 22% of the total effect of a price reduction on own brand sales is attributable to purchase time acceleration, 76% is attributable to brand switching, and only 2% is attributable to stockpiling. For the considered data set, we find a very short and small dip in category sales after a category price promotion. This dip lasts for one week, and it is about 1% of the current sales effect. Hence, for the yogurt category, the cannibalization effect of promotional activity is only marginal.

We conclude this chapter by mentioning some limitations of our framework and we provide some suggestions for further research. One feature of our model is that the consumption process only concerns the product category as a whole, and does not depend on the particular brand which is consumed. However, it might be more plausible that consumption utility does depend on the household's preferences for the consumed brand. This would also imply that households may want to buy multiple brands at a single purchase occasion. Kim *et al.* (2002) put forward a model of household demand with an additive but not necessarily linear utility structure, allowing for the selection of multiple alternatives at the same time. Another issue is that consumption is not observed in the scanner panel data set, so that it cannot be inferred without error. As we model consumption as a "deterministic" process, our framework does not account for this uncertainty. A third limitation is that, for mathematical tractability, our model does not allow that the error terms in the brand qualities are correlated, so that our model is prone to the so-called Independence of Irrelevant Alternatives property. Finally, an interesting suggestion for further research would be to develop a model for scanner panel data, integrating inventory-dependent consumption and the rational response of households to uncertainty about future prices. This kind of research would build on the studies by Assunção and Meyer (1993) and Erdem *et al.* (2003). Assunção and Meyer show that, under Markovian price uncertainty, consumption should rationally increase with the inventory level, but they do not translate the theoretical framework to a model for scanner panel data. On the other hand, Erdem *et al.* do develop a model for scanner panel data accounting for how inventories and expectations about future prices affect current purchase decisions. However, they assume that households have an exogenous (stochastic) usage requirement for the product such that consumption above this level does not provide any additional utility. Hence, they do not allow consumption to depend on the inventory level beyond possible stock outs.

2.A Derivation of the consumption function

The dynamic consumption optimization problem

$$\begin{aligned} &\max_{\{C_{i,\tau+h}\}_{h=0}^{H_{i,\tau}}} \sum_{h=0}^{H_{i,\tau}} \frac{1}{(1+\rho)^h}\, u(C_{i,\tau+h}), \\ &\text{subject to} \quad \sum_{h=0}^{H_{i,\tau}} C_{i,\tau+h} \leq S_{i,\tau}\,, \end{aligned} \tag{2.30}$$

can be solved by employing the Euler equation approach. The Lagrangian of the problem is defined by

$$\sum_{h=0}^{H_{i,\tau}} \left[\frac{1}{(1+\rho)^h}\, \frac{C_{i,\tau+h}^{1-\eta}}{1-\eta} \right] - \zeta \left[\sum_{h=0}^{H_{i,\tau}} C_{i,\tau+h} - S_{i,\tau} \right], \tag{2.31}$$

where ζ is the so-called shadow price of inventory. The first-order conditions at day $\tau+h$ and day $\tau+h-1$ ($h=1,\ldots,H_{i,\tau}$) are given by

$$\frac{1}{(1+\rho)^h}\, C_{i,\tau+h}^{-\eta} = \zeta, \tag{2.32}$$

$$\frac{1}{(1+\rho)^{h-1}}\, C_{i,\tau+h-1}^{-\eta} = \zeta, \tag{2.33}$$

respectively. It immediately follows from (2.32) and (2.33) that

$$\frac{1}{(1+\rho)^h}\, C_{i,\tau+h}^{-\eta} = \frac{1}{(1+\rho)^{h-1}}\, C_{i,\tau+h-1}^{-\eta}, \tag{2.34}$$

which can be rewritten as

$$C_{i,\tau+h} = \nu\, C_{i,\tau+h-1} \quad \text{with} \quad \nu = (1+\rho)^{-\frac{1}{\eta}}, \tag{2.35}$$

implying that

$$C_{i,\tau+h} = \nu^h\, C_{i,\tau}, \qquad h=1,\ldots,H_{i,\tau}, \tag{2.36}$$

where $\nu = (1+\rho)^{-\frac{1}{\eta}}$ is the consumption dampening factor. Following this consumption path, total consumption within the planning period equals

$$\sum_{h=0}^{H_{i,\tau}} C_{i,\tau+h} = C_{i,\tau} \sum_{h=0}^{H_{i,\tau}} \nu^h = C_{i,\tau}\, \frac{1-\nu^{H_{i,\tau}+1}}{1-\nu} \quad \text{with} \quad \nu = (1+\rho)^{-\frac{1}{\eta}} \tag{2.37}$$

for any given current consumption level $C_{i,\tau}$. Finally, as utility is strictly increasing in consumption, the inventory restriction (2.30) should be binding, so that optimal current consumption is given by

$$C_{i,\tau} = S_{i,\tau}\, \frac{1-\nu}{1-\nu^{H_{i,\tau}+1}} \quad \text{with} \quad \nu = (1+\rho)^{-\frac{1}{\eta}}. \tag{2.38}$$

The optimal consumption path, defined by (2.36) and (2.38) for a given inventory level $S_{i,\tau}$, provides total utility $U(S_{i,\tau})$, that is,

$$\begin{aligned} U(S_{i,\tau}) &= \sum_{h=0}^{H_{i,\tau}} \frac{1}{(1+\rho)^h} \frac{(\nu^h\, C_{i,\tau})^{1-\eta}}{1-\eta} \\ &= \frac{C_{i,\tau}^{1-\eta}}{1-\eta} \sum_{h=0}^{H_{i,\tau}} \left(\frac{\nu^{1-\eta}}{1+\rho}\right)^h \\ &= \frac{C_{i,\tau}^{1-\eta}}{1-\eta} \sum_{h=0}^{H_{i,\tau}} \nu^h \\ &= \frac{C_{i,\tau}^{1-\eta}}{1-\eta} \frac{1-\nu^{H_{i,\tau}+1}}{1-\nu} \\ &= \frac{S_{i,\tau}\, C_{i,\tau}^{-\eta}}{1-\eta}\,. \end{aligned} \tag{2.39}$$

2.B Derivation of the purchase probabilities

In this appendix, we derive the reduced form of our model. To do so, we frequently use the auxiliary result that, for any brand k and any quantity $\tilde{q} \geq 1$,

$$\begin{aligned} & \frac{\psi_{i,k,t|z}\,[\,V_{i,t}(\tilde{q}) - V_{i,t}(\tilde{q}-1)\,]}{P_{i,k,t}} < 1 \\ \Leftrightarrow\ & \frac{\exp(\alpha_{k|z} + \beta_1 BL_{i,k,t} + \beta_2 PM_{i,k,t} + \beta_3 \ln(\overline{C}_i) + \varepsilon_{i,k,t})\,[\,V_{i,t}(\tilde{q}) - V_{i,t}(\tilde{q}-1)\,]}{P_{i,k,t}} < 1 \\ \Leftrightarrow\ & \exp(\varepsilon_{i,k,t}) < \frac{P_{i,k,t}}{\exp(\alpha_{k|z} + \beta_1 BL_{i,k,t} + \beta_2 PM_{i,k,t} + \beta_3 \ln(\overline{C}_i))\,[\,V_{i,t}(\tilde{q}) - V_{i,t}(\tilde{q}-1)\,]} \\ \Leftrightarrow\ & \exp(\varepsilon_{i,k,t}) < \frac{1}{\exp(\,\tilde{\mu}_{i,k,t|z} + \ln[\,V_{i,t}(\tilde{q}) - V_{i,t}(\tilde{q}-1)\,]\,)} \\ \Leftrightarrow\ & \varepsilon_{i,k,t} < -\Big(\,\tilde{\mu}_{i,k,t|z} + \ln\Big[V_{i,t}(\tilde{q}) - V_{i,t}(\tilde{q}-1)\Big]\,\Big), \end{aligned} \tag{2.40}$$

where $\tilde{\mu}_{i,k,t|z} \equiv \alpha_{k|z} + \beta_1 BL_{i,k,t} + \beta_2 PM_{i,k,t} + \beta_3 \ln(\overline{C}_i) - \ln(P_{i,k,t})$. The corresponding probability of occurrence is given by

$$\begin{aligned} & \Pr\Big(\varepsilon_{i,k,t} < -\Big(\,\tilde{\mu}_{i,k,t|z} + \ln\Big[V_{i,t}(\tilde{q}) - V_{i,t}(\tilde{q}-1)\Big]\,\Big)\Big) \\ &= \exp\Big(-\exp\Big(\vartheta\Big(\,\tilde{\mu}_{i,k,t|z} + \ln\Big[V_{i,t}(\tilde{q}) - V_{i,t}(\tilde{q}-1)\Big]\,\Big)\Big)\Big) \\ &= \exp\Big(-\exp\Big(\,\mu_{i,k,t|z} + \vartheta \ln\Big[V_{i,t}(\tilde{q}) - V_{i,t}(\tilde{q}-1)\Big]\,\Big)\Big), \end{aligned} \tag{2.41}$$

where $\mu_{i,k,t|z} = \vartheta\tilde{\mu}_{i,k,t|z} = \alpha_{k|z} + \beta_1 BL_{i,k,t} + \beta_2 PM_{i,k,t} + \beta_3 \ln(\overline{C}_i) - \vartheta \ln(P_{i,k,t})$ after reparameterizing the intrinsic preferences $\alpha_{k|z}$ and the response parameters β_1, β_2, and β_3.

The probability that household i (belonging to segment z) does not make a purchase at shopping trip t is given by

$$
\begin{aligned}
&\Pr\Big(Y_{i,t}=0\Big|z\Big)\\
&=\Pr\Bigg(\frac{\psi_{i,k,t|z}\,[\,V_{i,t}(1)-V_{i,t}(0)\,]}{P_{i,k,t}}<1,\ k=1,\ldots,J\\
&\quad\Bigg|\ \frac{\psi_{i,k,t|z}\,[\,V_{i,t}(Q+1)-V_{i,t}(Q)\,]}{P_{i,k,t}}<1,\ k=1,\ldots,J\Bigg)\\
&=\frac{\Pr\Big(\frac{\psi_{i,k,t|z}\,[\,V_{i,t}(1)-V_{i,t}(0)\,]}{P_{i,k,t}}<1,\ k=1,\ldots,J\Big)}{\Pr\Big(\frac{\psi_{i,k,t|z}\,[\,V_{i,t}(Q+1)-V_{i,t}(Q)\,]}{P_{i,k,t}}<1,\ k=1,\ldots,J\Big)}\\
&=\frac{\Pr\big(\varepsilon_{i,k,t}<-(\,\tilde{\mu}_{i,k,t|z}+\ln[\,V_{i,t}(1)-V_{i,t}(0)\,]\,),\ k=1,\ldots,J\big)}{\Pr\big(\varepsilon_{i,k,t}<-(\,\tilde{\mu}_{i,k,t|z}+\ln[\,V_{i,t}(Q+1)-V_{i,t}(Q)\,]\,),\ k=1,\ldots,J\big)}\\
&=\frac{\prod_{k=1}^{J}\Pr\big(\varepsilon_{i,k,t}<-(\,\tilde{\mu}_{i,k,t|z}+\ln[\,V_{i,t}(1)-V_{i,t}(0)\,]\,)\big)}{\prod_{k=1}^{J}\Pr\big(\varepsilon_{i,k,t}<-(\,\tilde{\mu}_{i,k,t|z}+\ln[\,V_{i,t}(Q+1)-V_{i,t}(Q)\,]\,)\big)}\\
&=\frac{\prod_{k=1}^{J}\exp(-\exp(\,\mu_{i,k,t|z}+\vartheta\ln[\,V_{i,t}(1)-V_{i,t}(0)\,]\,))}{\prod_{k=1}^{J}\exp(-\exp(\,\mu_{i,k,t|z}+\vartheta\ln[\,V_{i,t}(Q+1)-V_{i,t}(Q)\,]\,))}\\
&=\frac{\exp(-\sum_{k=1}^{J}\exp(\,\mu_{i,k,t|z}+\vartheta\ln[\,V_{i,t}(1)-V_{i,t}(0)\,]\,))}{\exp(-\sum_{k=1}^{J}\exp(\,\mu_{i,k,t|z}+\vartheta\ln[\,V_{i,t}(Q+1)-V_{i,t}(Q)\,]\,))}\\
&=\exp\Bigg(\sum_{k=1}^{J}\exp\Big(\,\mu_{i,k,t|z}+\vartheta\ln\Big[V_{i,t}(Q+1)-V_{i,t}(Q)\Big]\,\Big)\\
&\quad-\sum_{k=1}^{J}\exp\Big(\,\mu_{i,k,t|z}+\vartheta\ln\Big[V_{i,t}(1)-V_{i,t}(0)\Big]\,\Big)\Bigg).
\end{aligned}
\tag{2.42}
$$

However, if a purchase occurs, the probability of selecting brand j becomes

$$
\begin{aligned}
&\Pr\Big(B_{i,t}=j\Big|Y_{i,t}=1,z\Big)\\
&=\Pr\Bigg(\frac{\psi_{i,j,t|z}\,[\,V_{i,t}(\tilde{q})-V_{i,t}(0)\,]}{P_{i,j,t}\,\tilde{q}}>\frac{\psi_{i,k,t|z}\,[\,V_{i,t}(\tilde{q})-V_{i,t}(0)\,]}{P_{i,k,t}\,\tilde{q}},\ k\neq j\Bigg)\\
&=\Pr\Bigg(\frac{\psi_{i,j,t|z}}{P_{i,j,t}}>\frac{\psi_{i,k,t|z}}{P_{i,k,t}},\ k\neq j\Bigg)\\
&=\Pr\Big(\tilde{\mu}_{i,j,t|z}+\varepsilon_{i,j,t}>\tilde{\mu}_{i,k,t|z}+\varepsilon_{i,k,t},\ k\neq j\Big)\\
&=\frac{\exp(\vartheta\tilde{\mu}_{i,j,t|z})}{\sum_{k=1}^{J}\exp(\vartheta\tilde{\mu}_{i,k,t|z})}\\
&=\frac{\exp(\mu_{i,j,t|z})}{\sum_{k=1}^{J}\exp(\mu_{i,k,t|z})},
\end{aligned}
\tag{2.43}
$$

and the conditional probability that q units are purchased becomes

$$
\begin{aligned}
&\Pr\Big(Q_{i,t} = q \Big| Y_{i,t} = 1, B_{i,t} = j, z\Big) \\
&= \Pr\Bigg(\frac{\psi_{i,j,t|z}\,[\,V_{i,t}(q) - V_{i,t}(q-1)\,]}{P_{i,j,t}} > 1, \frac{\psi_{i,j,t|z}\,[\,V_{i,t}(q+1) - V_{i,t}(q)\,]}{P_{i,j,t}} < 1 \\
&\quad \Bigg| \frac{\psi_{i,j,t|z}\,[\,V_{i,t}(1) - V_{i,t}(0)\,]}{P_{i,j,t}} > 1, \frac{\psi_{i,j,t|z}\,[\,V_{i,t}(Q+1) - V_{i,t}(Q)\,]}{P_{i,j,t}} < 1 \Bigg) \\
&= \Pr\Big(-\Big(\tilde{\mu}_{i,j,t|z} + \ln\Big[V_{i,t}(q) - V_{i,t}(q-1)\Big]\Big) < \varepsilon_{i,j,t} < -\Big(\tilde{\mu}_{i,j,t|z} + \ln\Big[V_{i,t}(q+1) - V_{i,t}(q)\Big]\Big) \\
&\quad \Big| -\Big(\tilde{\mu}_{i,j,t|z} + \ln\Big[V_{i,t}(1) - V_{i,t}(0)\Big]\Big) < \varepsilon_{i,j,t} < -\Big(\tilde{\mu}_{i,j,t|z} + \ln\Big[V_{i,t}(Q+1) - V_{i,t}(Q)\Big]\Big)\Big) \\
&= \frac{\Pr\big(-\big(\tilde{\mu}_{i,j,t|z} + \ln\big[V_{i,t}(q) - V_{i,t}(q-1)\big]\big) < \varepsilon_{i,j,t} < -\big(\tilde{\mu}_{i,j,t|z} + \ln\big[V_{i,t}(q+1) - V_{i,t}(q)\big]\big)\big)}{\Pr\big(-\big(\tilde{\mu}_{i,j,t|z} + \ln\big[V_{i,t}(1) - V_{i,t}(0)\big]\big) < \varepsilon_{i,j,t} < -\big(\tilde{\mu}_{i,j,t|z} + \ln\big[V_{i,t}(Q+1) - V_{i,t}(Q)\big]\big)\big)} \\
&= \Big(\exp\big(-\exp\big(\mu_{i,j,t|z} + \vartheta \ln\big[\,V_{i,t}(q+1) - V_{i,t}(q)\,\big]\big)\big) \\
&\qquad - \exp\big(-\exp\big(\mu_{i,j,t|z} + \vartheta \ln\big[\,V_{i,t}(q) - V_{i,t}(q-1)\,\big]\big)\big)\Big) \Big/ \\
&\quad \Big(\exp\big(-\exp\big(\mu_{i,j,t|z} + \vartheta \ln\big[\,V_{i,t}(Q+1) - V_{i,t}(Q)\,\big]\big)\big) \\
&\qquad - \exp\big(-\exp\big(\mu_{i,j,t|z} + \vartheta \ln\big[\,V_{i,t}(1) - V_{i,t}(0)\,\big]\big)\big)\Big). \qquad (2.44)
\end{aligned}
$$

Chapter 3

A reference price model for brand choice

3.1 Introduction and motivation

The literature provides ample evidence that households do not only consider current prices when deciding which brand to buy in a category, but also take into account past prices via the formation of internal reference points, see Krishnamurthi *et al.* (1992), Putler (1992), Kalyanaram and Little (1994), among many others. A conceptual basis for such reference price formation is provided by the Adaptation-Level Theory of Helson (1964). Kalyanaram and Winer (1995) translate the cumulative evidence into an empirical generalization.

The idea behind reference price effects is as follows. If the price of a brand is below its reference price, the observed price is lower than anticipated, resulting in a perceived gain. This would make the brand more attractive. Similarly, the opposite situation would result in a perceived loss, reducing the probability that the brand is purchased. An important consequence of internal reference price formation is that although frequent price discounts may be beneficial in the short-run, they may damage the brand in the long-run when households get used to these discounts and reference prices drop. The reduced price become anticipated and loses its effectiveness, whereas the non-promoted price becomes unanticipated and would be perceived as a loss.

As the existence of internal reference prices would imply a tradeoff between current and future brand sales, it is important to have a good understanding of how reference prices are formed from past price experience. This includes questions such as (i) what is the size of the reference effect?, (ii) what is the duration of the reference effect?, and (iii) how often are households able to construct a reference price? Still, current brand choice models dealing with reference price effects do not pay much attention to the operationalization of the reference price construct.

In this chapter, we focus on internal reference prices, which are constructed from past prices, and we do not consider external reference prices, based on the in-store prices during the purchase occasion. The two most popular specifications for internal reference price

are the price observed at the previous purchase occasion and an exponentially smoothed average of previously observed prices. Examples of the former reference price include Krishnamurthi *et al.* (1992), Mayhew and Winer (1992) and Chang *et al.* (1999), whereas examples of the latter specification include Lattin and Bucklin (1989), Kalyanaram and Little (1994) and Erdem *et al.* (2001). By assuming that the reference price of a brand equals its previous price, a reference effect of only one period is imposed, while multi-period dynamics may be more appropriate. An argument which is often made in favor of one-period dynamics is that the literature consistently shows that households have a very limited ability to recall prices, see, for example, the price knowledge surveys by Dickson and Sawyer (1990) and Vanhuele and Drèze (2002). However, this argument would imply that households might not even be able to recall the previously observed price. On the other hand, the exponentially-smoothed-average measure for reference price often results in extremely long price dynamics. For example, Briesch *et al.* (1997) conclude for several product categories that the lag in the formation of reference price is about six periods.

Two limitations of current reference price specifications are that (i) the dynamics underlying the reference price are determined a priori (possibly up to a tuning parameter), and (ii) it is implicitly assumed that households always have sufficient price knowledge to form a reference price. Both might result in underestimation of the size of the actual reference effect. Briesch *et al.* (1997) note that a misspecified reference price model can obscure the reference effect even when it may actually exist. Similarly, by not accounting for the possibility that households forget past prices, one might confound households not *forming* a reference price with households not *reacting* to it, resulting in lower response estimates. A third limitation of current practice is that the assumed processes usually do not account for uncertainty in reference price formation. However, as reference prices are not observed, their existence cannot be inferred without error.

In this chapter, we propose a new reference price model, which (i) does not a priori impose a rigid dynamic price structure, (ii) accounts for uncertainty in reference price formation, (iii) allows that a household may forget past prices, and (iv) even allows that a household cannot construct a reference price at all. Hence, the model is an attempt to bridge the gap between survey studies, such as Dickson and Sawyer (1990) and Vanhuele and Drèze (2002) which report that households have very limited price knowledge and hence may not be able to construct a reference price, and current reference price models which do not account for this. In the proposed framework, price recall of households is modeled as a hidden Markov-switching process with an absorbing state, and a reference price is constructed from the prices the household is able to recall. Our model can be used to get insight into the extent to which past price information is used in individual brand choice decisions, that is, it can address the question "which past prices do consumers use in forming a reference price, and how many are used?", raised by Kalyanaram and Winer (1995).

The structure of this chapter is as follows. In Section 3.2, we put forward our model. Next, in Section 3.3, we discuss parameter estimation, and in Section 3.4, we explain how

the estimation results can be used to analyze the dynamic structure underlying reference price formation and to analyze the extent to which internal reference prices are formed. We apply our model to an A.C. Nielsen scanner panel data set on catsup purchases in Section 3.5. Finally, in Section 3.6, we conclude with a discussion of the implications of our model, and we provide some directions for further research.

3.2 The model

In this section, we develop our reference price model and we show how it can be incorporated in a scanner panel data model for brand choice.

3.2.1 Reference price

We allow that households may forget prices observed in the past, as suggested by several price knowledge surveys in the literature. To this end, we introduce the unobserved 0/1 price recall variable $S_{i,t}^{\tau}$ such that

$$S_{i,t}^{\tau} = \begin{cases} 1 & \text{if at time } t \text{ household } i \text{ remembers the prices encountered at time } \tau \leq t \\ 0 & \text{otherwise,} \end{cases}$$

where the time indices t and τ correspond to purchase occasions. As $S_{i,t}^{\tau}$ does not contain a brand-index, price recall is the same across brands. Hence, a household is either able to recall the prices of all brands or cannot recall any brand price. To keep implementation of the model feasible, we impose that $S_{i,t}^{\tau} = 0$ if $t - \tau > L$, that is, we assume that households always forget prices which were observed more than L purchase occasions ago. The value of L has to be chosen by the researcher. As it does not affect the number of parameters in the model, a possible selection strategy would be to set $L \in \{1, 2, \ldots\}$ such that it gives the highest maximum likelihood value. The price knowledge of household i at purchase occasion t can be summarized in one *composite* price recall variable $\tilde{S}_{i,t} = (S_{i,t}^{t-L}, \ldots, S_{i,t}^{t-2}, S_{i,t}^{t-1})$, indicating which of the past L prices are recalled at purchase occasion t.

The second step in the development of our model concerns the description of how price recall of households evolves over time. We assume that, starting from purchase occasion τ when prices were observed, the subsequent price recall variables $\{S_{i,t}^{\tau}\}_{t=\tau,\tau+1,\ldots,\tau+L}$ obey a first-order Markov process with states 0 and 1. Hence, the probability that prices are recalled at the current purchase occasion depends on whether these prices were still in memory at the previous purchase occasion. Moreover, as it is plausible that prices are never recalled again once they have been forgotten, we let the "forgetting state" 0 be an

absorbing state. The Markov transition probabilities are given by

$$\Pr(S^{\tau}_{i,t} = 0 | S^{\tau}_{i,t-1} = 0) = 1, \tag{3.1}$$
$$\Pr(S^{\tau}_{i,t} = 1 | S^{\tau}_{i,t-1} = 0) = 0, \tag{3.2}$$
$$\Pr(S^{\tau}_{i,t} = 0 | S^{\tau}_{i,t-1} = 1) = 1 - p^{\tau}_{i,t}, \tag{3.3}$$
$$\Pr(S^{\tau}_{i,t} = 1 | S^{\tau}_{i,t-1} = 1) = p^{\tau}_{i,t}. \tag{3.4}$$

The process is initialized by setting $S^{\tau}_{i,\tau} = 1$, meaning that households are aware of the prices of brands at the moment of purchase. The memory processes defined by (3.1)−(3.4) are independent across households i as well as independent across purchase occasions τ during which prices were observed.

We define the transition probability $p^{\tau}_{i,t}$ as

$$p^{\tau}_{i,t} = \frac{1}{1 + \exp(-\left[\,\gamma_0 + \gamma_1(t-\tau)\,\right])}\,. \tag{3.5}$$

This *conditional* probability of price recall depends on the number of purchase occasions $t - \tau$ which have passed since the prices were observed at purchase occasion τ. The *unconditional* probability that the prices at purchase occasion τ are still remembered at purchase occasion t is given by

$$\Pr(S^{\tau}_{i,t} = 1) = \prod_{\tilde{\tau}=\tau+1}^{t} p^{\tau}_{i,\tilde{\tau}} = \prod_{\tilde{\tau}=1}^{t-\tau} \frac{1}{1 + \exp(-\left[\,\gamma_0 + \gamma_1\tilde{\tau}\,\right])}\,, \tag{3.6}$$

which is a decreasing function in $t - \tau$. Figure 3.1 displays some price memory patterns which can be reproduced by tuning the parameters γ_0 and γ_1 in (3.5). It shows a geometrically decaying pattern, a 1-period full memory pattern which amounts to the previously observed brand price being the reference price, a 2-periods full memory pattern, and a pattern such that prices are sometimes recalled at the next purchase occasion after which they are kept in memory forever. The graph illustrates that our model is able to mimic a wide variety of possible memory structures. Finally, we note that it is straightforward to extend (3.5) such that the conditional price recall probability $p^{\tau}_{i,t}$ also depends on, for example, the number of days elapsed since the previous purchase occasion $t - 1$ and the degree of promotional activity at purchase occasion τ when the prices were observed.

In our model, a reference price is only based on the prices which a household is able to recall, where price recall develops according to a first-order Markov process with absorbing state. We define the reference price $R_{i,j,t}$ of household $i \in \{1, \ldots, N\}$ for brand $j \in \{1, \ldots, J\}$ at purchase occasion $t \in \{1, \ldots, T_i\}$ as

$$R_{i,j,t} = \begin{cases} \frac{\sum_{\tau=t-L}^{t-1} S^{\tau}_{i,t} P_{i,j,\tau}}{\sum_{\tau=t-L}^{t-1} S^{\tau}_{i,t}} & \text{if} \quad \sum_{\tau=t-L}^{t-1} S^{\tau}_{i,t} > 0 \\ P_{i,j,t} & \text{if} \quad \sum_{\tau=t-L}^{t-1} S^{\tau}_{i,t} = 0. \end{cases} \tag{3.7}$$

Hence, this reference price equals the average of the prices which are recalled provided that at least one price observed in the past is still available in memory. However, it might also

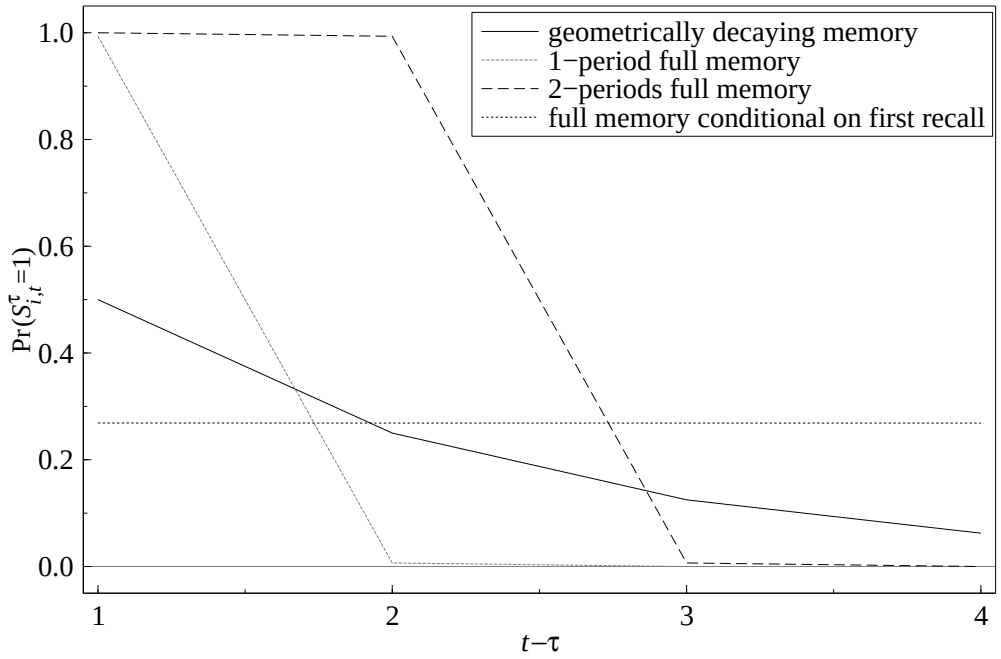

Figure 3.1: Illustration of some price memory patterns which can be reproduced by our model. The unconditional probability of price recall is plotted against the number of purchase occasions which have passed since the prices were observed.

happen that a household has forgotten all past prices. In such a case, the household does not have enough price information to form an internal reference price, and it perceives neither a price gain nor a loss. This is captured by the condition $P_{i,j,t} - R_{i,j,t} = 0$ in case $\sum_{\tau=t-L}^{t-1} S_{i,t}^{\tau} = 0$. The proposed reference price is an unweighted average from the perspective of the household, reflecting that households may have forgotten the order in which the recalled prices were observed. From the perspective of the researcher, it is however a weighted average. The weights are the unobserved 0/1 price recall variables for which the probability distributions have to be inferred from the data. Note that the weights differ across purchase occasions and households.

3.2.2 Brand choice

Our reference price can be incorporated in a standard model for brand choice. In this chapter, we consider a conditional logit model which (i) allows for asymmetric response of households to price gains and losses, and (ii) accounts for heterogeneity across households via the latent segments approach of Kamakura and Russell (1989). By doing so, we follow the mainstream literature on modeling brand choice and reference price effects, see, for example, Bell and Lattin (2000).

Prospect Theory, developed by Kahneman and Tversky (1979), predicts that households react more negatively to price losses than they react positively to price gains of equal size. Indeed, many reference price studies support this hypothesis, see Mayhew

and Winer (1992), Putler (1992), Kalyanaram and Little (1994) and Erdem *et al.* (2001), among others. Kalyanaram and Winer (1995) propose this gain-loss asymmetry as an empirical generalization. However, the literature also contains some studies which report either a lack of significance or even an opposite effect. Such examples include Briesch *et al.* (1997), Chang *et al.* (1999) and Bell and Lattin (2000). Besides allowing for separate response parameters for price gains and losses, it is also important to account for household heterogeneity. Several studies show that ignoring heterogeneity might result in biased response estimates, see, for example, Chintagunta *et al.* (1991), Jain *et al.* (1994) and Chang *et al.* (1999).

We assume that household i with unobserved memory state $\tilde{s}_{i,t} = (s_{i,t}^{t-L}, \ldots, s_{i,t}^{t-2}, s_{i,t}^{t-1})$, belonging to the unobserved response segment $z \in \{1, \ldots, Z\}$, perceives utility

$$\begin{aligned} U_{i,j,t|z,\tilde{s}_{i,t}} &= \alpha_{j|z} + \beta_{1|z} BL_{i,j,t} + \beta_{2|z} PM_{i,j,t} + \beta_{3|z} P_{i,j,t} + \beta_{4|z} G_{i,j,t|z,\tilde{s}_{i,t}} + \beta_{5|z} L_{i,j,t|z,\tilde{s}_{i,t}} + \varepsilon_{i,j,t} \\ &\equiv \tilde{U}_{i,j,t|z,\tilde{s}_{i,t}} + \varepsilon_{i,j,t} \end{aligned} \tag{3.8}$$

from purchasing brand j at purchase occasion t, where the random disturbance $\varepsilon_{i,j,t}$ is assumed to be independently and identically distributed obeying a Type-I Extreme Value distribution. In (3.8), $\alpha_{j|z}$ captures the intrinsic brand preferences of the household, $BL_{i,j,t}$ is the brand loyalty measure of Guadagni and Little (1983), $PM_{i,j,t}$ is a 0/1 promotion indicator (feature or display), $P_{i,j,t}$ is the shelf price, $G_{i,j,t|z,\tilde{s}_{i,t}}$ is the price gain relative to the reference price, and $L_{i,j,t|z,\tilde{s}_{i,t}}$ is the price loss. The gain and loss variables are defined as

$$G_{i,j,t|z,\tilde{s}_{i,t}} = I\{P_{i,j,t} < R_{i,j,t|z,\tilde{s}_{i,t}}\}(R_{i,j,t|z,\tilde{s}_{i,t}} - P_{i,j,t}), \tag{3.9}$$

$$L_{i,j,t|z,\tilde{s}_{i,t}} = I\{P_{i,j,t} > R_{i,j,t|z,\tilde{s}_{i,t}}\}(P_{i,j,t} - R_{i,j,t|z,\tilde{s}_{i,t}}), \tag{3.10}$$

where $I\{\cdot\}$ is the 0/1 indicator function. We note that the reference price $R_{i,j,t|z,\tilde{s}_{i,t}}$ is conditional on both the price recall state $\tilde{s}_{i,t}$ and segment membership z, as we allow the price recall parameters γ_0 and γ_1 in (3.5) to be different across the latent segments.

As suggested by the mental accounting framework of Thaler (1985), total utility $U_{i,j,t|z,\tilde{s}_{i,t}}$ derived from purchasing brand j consists of two components, that is, acquisition utility and transaction utility. The acquisition component corresponds to the "monetary value" of the deal. It is related to the discrepancy between the value of the brand for the household and the brand price. Additionally, the transaction component corresponds to the "psychological value" of the deal, which is determined by the discrepancy between the brand price and the corresponding reference price. In (3.8), transaction utility is captured by the price gain and price loss components. Brand choice models which do not consider reference effects, such as the model of Guadagni and Little (1983), assume acquisition utility but do not account for transaction utility. In his seminal paper on reference price modeling, Winer (1986) refers to the discrepancy between the observed price and the reference price as a sticker shock effect.

It immediately follows from the distributional assumptions on $\varepsilon_{i,j,t}$ that brand j is chosen, providing maximum utility, with probability

$$\Pr(B_{i,t} = j | Z_i = z, \tilde{S}_{i,t} = \tilde{s}_{i,t}) = \frac{\exp(\tilde{U}_{i,j,t|z,\tilde{s}_{i,t}})}{\sum_{k=1}^{J} \exp(\tilde{U}_{i,k,t|z,\tilde{s}_{i,t}})}, \tag{3.11}$$

where Z_i describes the unobserved segment membership variable for household i, see McFadden (1974).

3.3 Parameter estimation

In this section, we discuss how the parameters of our model can be estimated. Parameter estimation is not straightforward, as the model contains two kinds of unobserved variables, that is, segment membership and the price recall states. The price recall variables obey a first-order Markov process, and hence are not independent. We use the EM algorithm of Dempster *et al.* (1977) to deal with the unobserved response segments, see McLachlan and Krishnan (1997) for a textbook discussion. Mixture models, such as our model, provide a natural application area for this algorithm, see Wedel *et al.* (1993), Ramaswamy *et al.* (1994) and Böckenholt (1999), among others. Within the EM algorithm, we apply an iterative filter, put forward in Hamilton (1989) and also described in Hamilton (1994, p. 692−693), to sum out the unobserved price memory states $\tilde{s}_{i,t}$.

Using the shorthand notation $B_{i,1:\tilde{t}}$ to denote the sequence of brand choices $B_{i,1}$, $B_{i,2}, \ldots, B_{i,\tilde{t}}$, the (unconditional) likelihood function is given by

$$\begin{aligned} \mathcal{L} &= \prod_{i=1}^{N} \Pr(B_{i,1:T_i} = b_{i,1:T_i}) \\ &= \prod_{i=1}^{N} \sum_{z=1}^{Z} \pi_z \Pr(B_{i,1:T_i} = b_{i,1:T_i} | Z_i = z), \end{aligned} \tag{3.12}$$

where $b_{i,1:T_i}$ corresponds to realized brand choice and $\pi_z = \Pr(Z_i = z)$ is the size of segment z.

3.3.1 Applying the Hamilton filter

To evaluate the likelihood function (3.12), one needs to compute

$$\begin{aligned} &\Pr(B_{i,1:T_i} = b_{i,1:T_i} | Z_i = z) \\ &= \sum_{\tilde{s}_{i,1}} \cdots \sum_{\tilde{s}_{i,T_i}} \Pr(B_{i,1:T_i} = b_{i,1:T_i}, \tilde{S}_{i,1} = \tilde{s}_{i,1}, \ldots, \tilde{S}_{i,T_i} = \tilde{s}_{i,T_i} | Z_i = z), \end{aligned} \tag{3.13}$$

which involves 2^{LT_i} summations to get rid of all unobserved memory states. We note that the Markov property of $\tilde{S}_{i,1}, \ldots, \tilde{S}_{i,T_i}$ implies that all possible price recall paths have to be

considered. For example, even if price recall is restricted to $L = 4$ periods and household i has only made $T_i = 6$ purchases in the category, (3.13) would already consist of more than 16 million components. Fortunately, the Hamilton (1989) filter turns out to be a very useful tool to evaluate this objective function, as it avoids such infeasible summations.

Hamilton (1989) originally developed his nonlinear filter to make inference on changes in economic regimes using time series. He applied the filter to establish the dates of historical business cycles, assuming that the underlying recession-expansion regimes follow a first-order Markov process. Econometric models with latent Markov-switching processes were first introduced by Goldfeld and Quandt (1973), and our reference price framework can be considered a member of this class of models as well.

In our application of the Hamilton filter, the unobserved components are the *composite* price recall variables $\tilde{S}_{i,t}$, $t = 1, \ldots, T_i$, for which the transition probabilities still have to be derived. It follows from the independence assumptions made in Subsection 3.2.1 and the assumption of complete price information at the moment of purchase that

$$\begin{aligned} &\Pr\Big(\tilde{S}_{i,t} = \tilde{s}_{i,t} \Big| \tilde{S}_{i,t-1} = \tilde{s}_{i,t-1}\Big) \\ &= \prod_{\tau=t-L}^{t-1} \Pr(S_{i,t}^{\tau} = s_{i,t}^{\tau} | S_{i,t-1}^{\tau} = s_{i,t-1}^{\tau}) \qquad \text{with} \qquad S_{i,t-1}^{t-1} = 1, \end{aligned} \tag{3.14}$$

where the transition probabilities $\Pr(S_{i,t}^{\tau} = s_{i,t}^{\tau} | S_{i,t-1}^{\tau} = s_{i,t-1}^{\tau})$ are defined by (3.1)−(3.5). We initialize the process at the first purchase occasion $t = 1$ by setting

$$\Pr\Big(\tilde{S}_{i,1} = \tilde{s}_{i,1}\Big) = \prod_{\tau=1-L}^{0} \Pr\Big(S_{i,1}^{\tau} = s_{i,1}^{\tau}\Big), \tag{3.15}$$

where, analogous to (3.6),

$$\Pr(S_{i,1}^{\tau} = 1) = \prod_{\tilde{\tau}=1}^{1-\tau} \frac{1}{1 + \exp(-\left[\,\gamma_0 + \gamma_1 \tilde{\tau}\,\right])}\,. \tag{3.16}$$

We note that, for notational convenience, we suppress the dependence on segment membership z.

The Hamilton filter allows for inference on the unobserved memory state $\tilde{s}_{i,t}$ of household i at purchase occasion t, while taking into account the household's purchase history $b_{i,1:t}$ up to purchase occasion t. As a by-product, it also provides an evaluation of the brand choice probability $\Pr(B_{i,t} = b_{i,t} | B_{i,1:t-1} = b_{i,1:t-1})$ *unconditional* on $\tilde{s}_{i,t}$, which has been summed out. We note that current brand choice may depend on previous brand choices via the brand loyalty variable $BL_{i,j,t}$. By iteratively applying the filter, one can obtain the *unconditional* likelihood value for each household.

In the sequel, we use the shorthand notation $\xi_{i,t|\tilde{t}} = \overline{\Pr}(\tilde{S}_{i,t} = \tilde{s}_{i,t} | B_{i,1:\tilde{t}} = b_{i,1:\tilde{t}})$ to denote the $(2^L \times 1)$ vector containing the probabilities of *all* possible states $\tilde{s}_{i,t}$, given the household's observed purchase history $b_{i,1:\tilde{t}}$. Basically, an iteration of the Hamilton filter

can be split up into two steps. Starting from $\xi_{i,t|t-1}$, the first step consists of updating $\xi_{i,t|t-1}$ to $\xi_{i,t|t}$, that is, updating the inference on $\tilde{s}_{i,t}$ by including the information contained in the most recent purchase observation $b_{i,t}$. It immediately follows from Bayes' theorem that

$$\xi_{i,t|t} = \frac{\xi_{i,t|t-1} \odot \overline{\Pr}(B_{i,t} = b_{i,t}|\tilde{S}_{i,t} = \tilde{s}_{i,t}, B_{i,1:t-1} = b_{i,1:t-1})}{\iota' \left[\xi_{i,t|t-1} \odot \overline{\Pr}(B_{i,t} = b_{i,t}|\tilde{S}_{i,t} = \tilde{s}_{i,t}, B_{i,1:t-1} = b_{i,1:t-1}) \right]}, \tag{3.17}$$

where ι is the $(2^L \times 1)$ vector consisting of ones, and $\odot$ denotes element-by-element multiplication. We note that the vector of probabilities $\overline{\Pr}(B_{i,t} = b_{i,t}|\tilde{S}_{i,t} = \tilde{s}_{i,t}, B_{i,1:t-1} = b_{i,1:t-1})$ is conditional on the memory states $\tilde{s}_{i,t}$, so that it can be evaluated immediately using (3.7)−(3.11). Furthermore, it is crucial to note that the denominator in (3.17) amounts to the *unconditional* probability $\Pr(B_{i,t} = b_{i,t}|B_{i,1:t-1} = b_{i,1:t-1})$. This is the by-product, which allows us to evaluate (3.13). In the second step of the iteration, $\xi_{i,t|t}$ is updated to $\xi_{i,t+1|t}$, that is, the next memory state $\tilde{s}_{i,t+1}$ is inferred given the currently available purchase observations $b_{i,1:t}$. As price recall obeys a first-order Markov process, the basic Markov identity

$$\xi_{i,t+1|t} = \Lambda_{i,t+1|t}{}'\xi_{i,t|t} \tag{3.18}$$

holds. Here, $\Lambda_{i,t+1|t}$ is the row-conditional $(2^L \times 2^L)$ transition probability matrix for $\tilde{S}_{i,t}$. The elements $\Pr(\tilde{S}_{i,t+1} = \tilde{s}_{i,t+1}|\tilde{S}_{i,t} = \tilde{s}_{i,t})$ of $\Lambda_{i,t+1|t}$ are defined by (3.14) and hence (3.1)−(3.5). The output $\xi_{i,t+1|t}$ of (3.18) can be used as the input for (3.17) in the next iteration. This iterative procedure is initialized by setting $\xi_{i,1|0} = \overline{\Pr}(\tilde{S}_{i,1} = \tilde{s}_{i,1})$ in accordance with (3.15) and (3.16).

In sum, the following algorithm results in an evaluation of the *unconditional* probability that the sequence of brand choices $b_{i,1:T_i}$ is observed for household i, that is,

Initialize the unconditional probability: $\Pr_i = 1$.

Initialize $\xi_{i,1|0}$ using (3.15) and (3.16).

Do for purchase occasion $t = 1, \ldots, T_i$:

Compute $\overline{\Pr}(B_{i,t} = b_{i,t}|\tilde{S}_{i,t} = \tilde{s}_{i,t}, B_{i,1:t-1} = b_{i,1:t-1})$ using (3.7)−(3.11).

Compute $\xi_{i,t|t}$ using (3.17).

Update the unconditional probability: $\Pr_i = \Pr_i \cdot [\,\text{denominator of (3.17)}\,]$.

Compute row-conditional $\Lambda_{i,t+1|t}$ using (3.14) and hence (3.1)−(3.5).

Compute $\xi_{i,t+1|t}$ using (3.18).

Return $\Pr_i$.

3.3.2 Applying the EM algorithm

To obtain the parameter estimates for our model we maximize the likelihood function (3.12), which can be evaluated *for given parameter values* using the Hamilton filter. Although numerical techniques such as the BFGS algorithm can be applied directly to find the optimal parameter values, we consider the EM algorithm to take advantage of the specific structure of the problem. The EM algorithm quickly moves to reasonable (but not yet optimal) parameter values and it is quite robust with respect to the starting values, see, for example, Hamilton (1990). As final convergence of the algorithm is usually slow, one can decide to first apply the EM algorithm until the steps in the parameter space become rather small, and next one can do direct optimization of (3.12) starting from the parameter location obtained from the EM stage.

The EM algorithm considers the *complete* data likelihood of our model, that is, the *joint* likelihood of both observed brand choices and unobserved segment membership. The complete data likelihood is given by

$$\mathcal{L}_c = \prod_{i=1}^{N} \prod_{z=1}^{Z} \Big(\pi_z \Pr(B_{i,1:T_i} = b_{i,1:T_i} | Z_i = z) \Big)^{I\{Z_i = z\}}. \tag{3.19}$$

After taking logarithms we obtain

$$\ln \mathcal{L}_c = \sum_{i=1}^{N} \sum_{z=1}^{Z} I\{Z_i = z\} \ln(\pi_z) + \sum_{i=1}^{N} \sum_{z=1}^{Z} I\{Z_i = z\} \ln(\Pr(B_{i,1:T_i} = b_{i,1:T_i} | Z_i = z)). \tag{3.20}$$

The EM algorithm contains an Expectation step and a Maximization step, which are performed iteratively. In the E-step, the expectation of the log complete data likelihood (3.20) is taken with respect to the unobserved segment membership variables Z_i, $i = 1, \ldots, N$, *given* the household's observed purchase history $b_{i,1:T_i}$, $i = 1, \ldots, N$, and *given* the current parameter values. This results in

$$E[\ln \mathcal{L}_c] = \sum_{i=1}^{N} \sum_{z=1}^{Z} \Pi_{i,z} \ln(\pi_z) + \sum_{i=1}^{N} \sum_{z=1}^{Z} \Pi_{i,z} \ln(\Pr(B_{i,1:T_i} = b_{i,1:T_i} | Z_i = z)), \tag{3.21}$$

where, using Bayes' theorem,

$$\Pi_{i,z} \equiv \Pr(Z_i = z | B_{i,1:T_i} = b_{i,1:T_i}) = \frac{\pi_z \Pr(B_{i,1:T_i} = b_{i,1:T_i} | Z_i = z)}{\sum_{\tilde{z}=1}^{Z} \pi_{\tilde{z}} \Pr(B_{i,1:T_i} = b_{i,1:T_i} | Z_i = \tilde{z})}. \tag{3.22}$$

Next, in the M-step, (3.21) is maximized over all segment sizes $\pi = (\pi_1, \ldots, \pi_Z)$ and all model parameters denoted by $\theta = (\theta_1, \ldots, \theta_Z)$, *given* the posterior segment probabilities $\Pi_{i,z}$ computed in the E-step. This amounts to $Z + 1$ *separate* subproblems, that is,

$$\max_{\pi} \quad \sum_{i=1}^{N} \sum_{z=1}^{Z} \Pi_{i,z} \ln(\pi_z), \tag{3.23}$$

$$\max_{\theta_z} \quad \sum_{i=1}^{N} \Pi_{i,z} \ln(\Pr(B_{i,1:T_i} = b_{i,1:T_i} | Z_i = z)), \qquad z = 1, \ldots, Z. \tag{3.24}$$

The resulting parameter updates can again be used as input for the E-step, see (3.22), after which a new iteration starts. It can be shown that subproblem (3.23) has a closed-form solution, which is given by

$$\pi_z = \frac{1}{N}\sum_{i=1}^{N} \Pi_{i,z}, \qquad z = 1, \ldots, Z. \tag{3.25}$$

In sum, an iteration of the EM algorithm consists of the follows steps, that is,

Compute the posterior segment probabilities $\Pi_{i,z}$ using (3.22).

Update the segment sizes $\pi = (\pi_1, \ldots, \pi_Z)$ using (3.25).

Update the model parameters $\theta = (\theta_1, \ldots, \theta_Z)$ by solving (3.24).

This iterative procedure converges and the resulting parameter values correspond to the maximum likelihood [ML] estimates of (3.12), see Dempster *et al.* (1977). In the first iteration, we initialize the EM algorithm by randomizing the $\Pi_{i,z}$ under the restriction that $\sum_z \Pi_{i,z} = 1$. By considering multiple runs, this is an effective and easily implemented approach to try out different starting points.

3.4 Interpretation

In this section, we discuss how the estimation results for our model can be used to analyze (i) the dynamic structure underlying reference price formation, and to analyze (ii) the extent to which households are able to form an internal reference price.

Ideally, for given segment membership z, inference on the unobserved memory state $\tilde{s}_{i,t}$ should be based on *all* available information $b_{i,1:T_i}$ in the data set. This would give the most reliable results. However, the Hamilton filter described in the previous section only allows for inference on $\tilde{s}_{i,t}$ using the information $b_{i,1:t}$ available up to purchase occasion $t \leq T_i$. So, again using the notation $\xi^z_{i,t|\tilde{t}} = \overline{\Pr}(\tilde{S}_{i,t} = \tilde{s}_{i,t} | B_{i,1:\tilde{t}} = b_{i,1:\tilde{t}}, Z_i = z)$ but now with explicit segment membership index z, we can only estimate $\xi^z_{i,t|t}$, while we are actually more interested in $\xi^z_{i,t|T_i}$. Fortunately, the smoothing algorithm of Kim (1994), also described in Hamilton (1994, p. 694), allows us to compute $\xi^z_{i,t|T_i}$ from $\xi^z_{i,t|t}$. This algorithm is characterized by the recursion

$$\xi^z_{i,t|T_i} = \xi^z_{i,t|t} \odot \left[\Lambda^z_{i,t+1|t}\left(\xi^z_{i,t+1|T_i} \oslash \xi^z_{i,t+1|t} \right) \right], \tag{3.26}$$

where $\oslash$ denotes element-by-element division. Using that $\xi^z_{i,T_i|T_i}$ is already available from the Hamilton filter and starting from $t = T_i - 1$, the recursion (3.26) is iterated backward to obtain all $\xi^z_{i,t|T_i}$, $t = 1, \ldots, T_i$.

Finally, one can obtain the posterior distribution of $\tilde{S}_{i,t} = (S^{t-L}_{i,t}, \ldots, S^{t-2}_{i,t}, S^{t-1}_{i,t})$ *unconditional* on segment membership z by weighting the segment-conditional $\xi^z_{i,t|T_i}$,

$z = 1, \ldots, Z$, with the corresponding posterior segment membership probabilities $\Pi_{i,z}$, $z = 1, \ldots, Z$. This amounts to computing

$$\Pr(\tilde{S}_{i,t} = \tilde{s}_{i,t} | B_{i,1:T_i} = b_{i,1:T_i}) = \sum_{z=1}^{Z} \Pi_{i,z} \, \xi^{z}_{i,t|T_i}, \tag{3.27}$$

where $\Pi_{i,z}$ is defined by (3.22).

To increase the interpretability of (3.27), one can derive the probability that household i is still able to recall at purchase occasion t the prices observed at purchase occasion τ as

$$\Pr(S^{\tau}_{i,t} = 1 | B_{i,1:T_i} = b_{i,1:T_i}) = \sum_{\tilde{s}_{i,t}:\tilde{s}^{\tau}_{i,t}=1} \Pr(\tilde{S}_{i,t} = \tilde{s}_{i,t} | B_{i,1:T_i} = b_{i,1:T_i}), \tag{3.28}$$

and one can compute the probability that household i has enough price information to form an internal reference price at purchase occasion t as

$$\Pr(\sum_{\tau=t-L}^{t-1} S^{\tau}_{i,t} > 0 | B_{i,1:T_i} = b_{i,1:T_i}) = 1 - \Pr(\tilde{S}_{i,t} = (0, \ldots, 0, 0) | B_{i,1:T_i} = b_{i,1:T_i}). \tag{3.29}$$

We emphasize that (3.28) and (3.29) take into account *all* information contained in the household's purchase history.

3.5 Empirical analysis

In this section, we apply our reference price model to an A.C. Nielsen scanner panel data set on catsup purchases in the Sioux Falls SD market. The considered period consists of 114 weeks from June 1986 to August 1988. The first 57 weeks are used for initialization purposes and are discarded in the log-likelihood evaluation, while the remaining 57 weeks are used for either parameter estimation or out-of-sample model validation. Only households which made at least three purchases in both periods are considered in the analysis. The estimation sample contains 80% of these households. The remaining 20% is assigned to a hold-out sample. The total sample consists of 619 households who made together 9416 purchases in the catsup category. The brands in our data set are Heinz, Hunts, Del Monte and Private Label.

3.5.1 Parameter estimates

We consider two variants of our model, that is, the model without unobserved heterogeneity and the model with an unknown number of latent response segments to capture unobserved heterogeneity. For the latter variant, we use the AIC-3 criterion, proposed by Bozdogan (1994), to determine the optimal number of segments. This measure is defined as $-2 \ln \mathcal{L} + 3K$ with $\mathcal{L}$ being the maximum likelihood value and K being the number of parameters to be estimated. The model corresponding to the smallest AIC-3 value is

Table 3.1: Parameter estimates for our model. The estimated standard errors are given in parentheses.

	no heterogeneity	2-segments model		
		averaged	segment 1	segment 2
β_1 (brand loyalty)	0.232^{***}	0.200^{***}	0.299^{***}	0.120^{***}
	(0.010)	(0.013)	(0.037)	(0.040)
β_2 (promotion)	0.456^{***}	0.469^{***}	0.372^{***}	0.548^{***}
	(0.034)	(0.037)	(0.076)	(0.093)
β_3 (price)	-1.130^{***}	-1.077^{*}	-0.777	-1.321
	(0.421)	(0.654)	(1.375)	(0.893)
β_4 (price gain)	0.949^{***}	0.913^{***}	0.203	1.494^{***}
	(0.255)	(0.330)	(0.277)	(0.578)
β_5 (price loss)	-1.092^{***}	-1.010^{***}	-1.001	-1.017^{***}
	(0.187)	(0.275)	(0.612)	(0.322)
γ_0 (intercept)	-2.105^{***}	-2.007^{***}	-2.142	-1.897^{*}
	(0.509)	(0.696)	(1.971)	(1.006)
γ_1 $(t-\tau)$	0.653^{*}	0.686	0.688	0.684
	(0.347)	(0.634)	(1.816)	(0.687)
segment size			0.449	0.551

* significant at 10%.
** significant at 5%.
*** significant at 1%.

selected. The AIC-3 criterion suggests more parsimonious models than the standard AIC criterion, but if the number of observations exceeds 20 it is less parsimonious than BIC, the other popular measure in the literature. An extensive simulation study by Andrews and Currim (2003) indicates that AIC-3, applied to multinomial choice data, performs better than several other criteria including AIC and BIC. For our model, after setting $L = 4$ (this lag in price recall gives the highest maximum likelihood value), we find two response segments in the data set. Our parameter estimates are reported in Table 3.1. For the two-segments model, we show both the segment-specific parameters and the segment-

averaged parameters for which the standard errors have been obtained using the delta method.[1]

For both the model without unobserved heterogeneity and the two-segments model, all response parameters have the expected sign. Furthermore, all parameters are significant at the 1% level in the first model, whereas in the second model this holds for all parameters except the price parameter. However, the parameter estimates do not indicate a strong response asymmetry concerning price gains and price losses. In the two-segments model, the first segment contains about 45% of the households and the second segment contains the remaining 55%. Comparison of the segment-specific parameter estimates suggests that households in segment 2 react stronger to promotional activities of brands, while households in segment 1 are more driven by state dependence, which is represented by the brand loyalty variable. For both variants of our model, the estimates of the price recall parameters γ_0 and γ_1 in (3.5) indicate that price memory is very limited, as $\gamma_0 + \gamma_1$ is much smaller than zero, and hence even price recall related to the previous purchase occasion is already far below 50%.

3.5.2 Impact of purchase timing and promotion on price recall

To see whether interpurchase times and (non-price) promotional activity affect price recall, we also estimate our model after including these variables in the price recall probability (3.5). Here, we define promotional activity as the average promotion rate in the category at purchase occasion τ when the prices were observed. Hence, promotional activity is measured as $\frac{1}{J}\sum_{k=1}^{J} PM_{i,k,\tau}$. We note that it is conceptually straightforward to include such time-varying variables in the memory process, but a practical drawback is that the transition probability matrix $\Lambda_{i,t+1|t}$ has to be computed for each purchase occasion to evaluate the likelihood function (3.12). This makes parameter estimation very time-consuming. In contrast, for the original model in which (3.5) just depends on $t-\tau$, $\Lambda_{i,t+1|t}$ only needs the be evaluated once (for each response segment) per likelihood evaluation.

It turns out that interpurchase time does not have a significant impact on price recall. For the model without unobserved heterogeneity, the z-score is -0.668. For the two-segments model, the z-score becomes 0.019. This lack of significance is consistent with the price knowledge study by Vanhuele and Drèze (2002), which reports that purchase recency does not affect price knowledge. On the other hand, the effect of promotional activity on price recall is positive and significant at the 1% level for both models, with z-scores of 3.011 and 2.699, respectively. So, promotional activities such as features and displays

[1] The segment-averaged parameter values should not be interpreted as the parameter values for "the average household", as any household belongs to exactly one of the unobserved segments and hence cannot have preferences and response parameters corresponding to a convex combination. For an appropriate meaning, we have to reformulate our model as a random effects model in which the parameters for each household are drawn from a continuous heterogeneity distribution. In a mixture model, the heterogeneity distribution is represented by a discrete distribution. This subtle difference is pointed out by Jain *et al.* (1994). However, as the difference is only conceptual and not methodological, we continue to interpret the segment-averaged parameter values as aggregate estimates.

Table 3.2: Empirical comparison of our model and three competing models. From left to right, we report the in-sample log-likelihood value without accounting for unobserved heterogeneity, the number of segments identified by the AIC-3 criterion, the corresponding number of parameters, the in-sample log-likelihood, the AIC-3 values, and the out-of-sample log-likelihood.

	LL no het.	# seg.	# par.	LL in	AIC-3	LL out[a]
our reference price	−2992.3	2	23	−2961.0	5991.0	−716.2
exp. weighted average	−3009.6	3	32	−2951.5	5999.0	−725.8
previous price	−3011.8	3	29	−2952.7	5992.3	−718.7
no reference price	−3019.0	3	23	−2973.1	6015.2	−722.4

a: The LL-out values for the final three models with two segments are −726.4, −723.1, and −723.7, respectively.

substantially increase the probability that corresponding prices are kept in memory, and hence affect both current and future brand choices. This result is, in a sense, consistent with the study by Lattin and Bucklin (1989), which also demonstrates that promotional activity has significant reference effects. Still, there is an important difference. We model the reference effect of promotion on brand choice as an *indirect* effect via the reference effect of price, whereas Lattin and Bucklin model the reference effect of promotion as an effect in itself. However, it is not clear why such promotional activities, which only aim at drawing the attention of households, should directly affect future brand choice utility.

3.5.3 Model performance

To see how well our model performs in-sample and out-of-sample, we compare it with three popular alternatives in the literature. The competing models are a brand choice model which does not account for any reference price effect, a model in which the reference price of a brand is the price observed at the previous purchase occasion, and a model in which the reference price is defined as an exponentially smoothed average of previously observed prices.

Table 3.2 contains the results of the empirical comparison. The in-sample log-likelihood values without unobserved heterogeneity show that our reference price is the most flexible one. Interestingly, after accounting for heterogeneity, the AIC-3 criterion identifies three segments for the three competing models, whereas it suggests only two segments for our model. Hence, our model has one response segment less. This is probably due to its richer dynamical structure, allowing for additional household heterogeneity to be picked up. The AIC-3 value of our model is slightly lower than it is for the previous-price model, but it is clearly lower than the AIC-3 values of the remaining two models. Our model also has the highest out-of-sample log-likelihood, although the difference with the previous-

Table 3.3: The top part shows the estimated response parameters of the price gain and price loss variables. For the multi-segment models, these response parameters amount to segment-weighted averages. The middle part reports the corresponding z-scores. The bottom part contains the p-values of the LR test for the null hypothesis "no gain-loss asymmetry".

		no heterogeneity		multi-segment	
		gain	loss	gain	loss
param.	our reference price	0.949	−1.092	0.913	−1.010
	exp. weighted average	0.045	−0.215	0.083	−0.269
	previous price	−0.050	−0.115	0.046	−0.111
z-score	our reference price	3.722	−5.852	2.772	−3.674
	exp. weighted average	0.646	−3.899	0.781	−2.238
	previous price	−1.008	−3.419	0.586	−2.631
p-value	our reference price	0.642		0.740	
	exp. weighted average	0.013		0.000	
	previous price	0.003		0.000	

price model is again quite small. The out-of-sample comparison would even become more favorable if we set the number of response segments for all models equal to two, as in our model. Overall, we conclude that our model has a good performance relative to three popular alternative models.

3.5.4 Size and symmetry of the reference effect

The parameter estimates in Table 3.1 suggest that the effect of reference price on brand choice is substantial, but without a clear asymmetry between price gains and losses. It would however be interesting to consider these results in more detail, and to compare them with the sticker shock effects resulting from the other two reference price specifications that we have considered in the empirical comparison.

The top part of Table 3.3 shows the estimated price gain and price loss parameters for the three models, and the middle part of the table provides the corresponding z-scores. Clearly, both in an absolute sense and in terms of significance, the reference effect in our model is largest. This is not surprising, as our reference price is the most flexible one, and only our model is able to distinguish between households not forming a reference price and households not reacting to it. Neglecting this distinction would result in a downward bias of the estimated response to price gains and losses *given* that a reference price is available. As the reference effect resulting from our new model is substantially larger than the size

we would obtain from applying current practice, the impact of internal reference price formation on brand choice might be even larger than is currently believed.

Finally, the bottom part of Table 3.3 shows the p-values of the Likelihood Ratio [LR] test for the null hypothesis "no gain-loss asymmetry", that is, $\beta_4 = -\beta_5$. These results are interesting too. For our model the null hypothesis cannot be rejected at any reasonable significance level, whereas it is strongly rejected at the 5% level for the other two models. So, although applying current practice would result in finding an asymmetric sticker shock effect, as predicted by Prospect Theory, this asymmetry seems to disappear when our more advanced reference price model is considered.

3.5.5 Analyzing reference price formation

Our model can be used to get insight into (i) the probabilities that each of the past prices is used for reference price formation, and (ii) the probability that households have sufficient price knowledge to form an internal reference price. Current reference price models in the literature do not allow for such inference from scanner panel data.

We discuss the results for our model with two response segments. Figure 3.2 contains histograms revealing the extent to which households are still able to recall the prices which were observed $t-\tau$ purchase occasions ago, where $t-\tau = 1, 2, 3, 4$. These histograms are obtained by computing (3.28) for all purchase occasions in the estimation sample. Figure 3.2 clearly demonstrates that price memory of households is very limited. For example, even the recall probabilities for prices observed at the previous purchase occasion are seldom larger than 50%. The average percentages of price recall are 21.0%, 7.3%, 3.7% and 2.4%, for 1, 2, 3 and 4 purchase occasions ago.

Figure 3.3 shows an analogous histogram indicating to what extent households form internal reference prices. This histogram is obtained by computing (3.29) for all purchase occasions in the estimation sample. It can be seen that the probability of reference price formation is usually smaller than 60%. The average probability is 31.1%. We note that this low percentage is a direct consequence of the low price recall probabilities.

3.5.6 Price elasticity analysis

The existence of internal reference prices implies a tradeoff between current and future brand sales. Related to this tradeoff, Greenleaf (1995) and Kopalle *et al.* (1996) develop dynamic programming models to investigate the impact of reference effects on the profitability of price promotions, and to obtain optimal pricing schemes. They define reference prices at the market level. However, in our model for scanner panel data, we consider reference prices at the household level and we can only analyze the impact on household demand, and not the impact on profitability.

Our model suggests a strong and significant effect of reference price on brand choice, but the managerially relevant implications for current and future brand sales are still unclear. To get a better understanding of the effects of a price change, we perform a

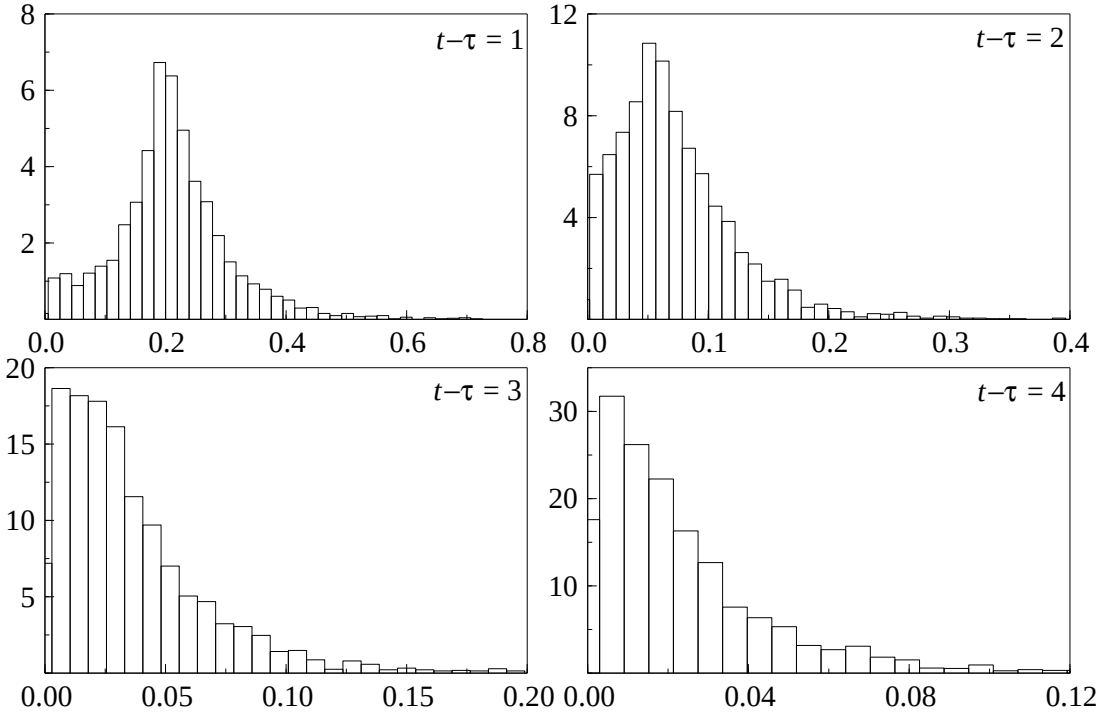

Figure 3.2: The probability that a household is able to recall the prices observed $t - \tau$ purchase occasions ago, $t - \tau = 1, 2, 3, 4$. The histograms are based on all purchase occasions in the estimation sample.

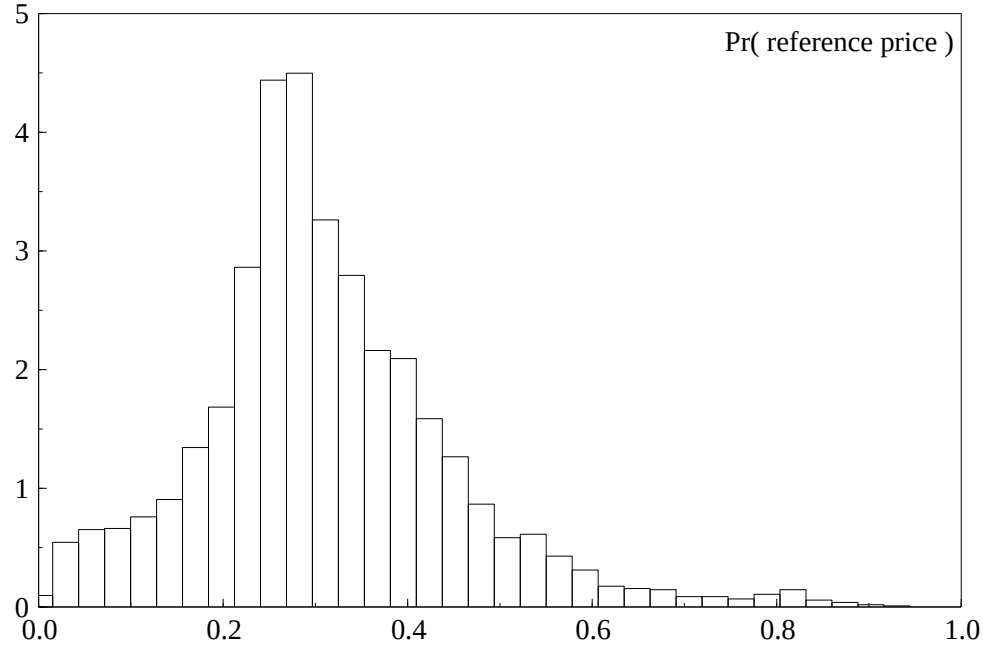

Figure 3.3: The probability that a household has enough price knowledge to form an internal reference price. The histogram is based on all purchase occasions in the estimation sample.

Table 3.4: Estimated price elasticities indicating how current and future "sales" are affected by a temporary own price change.

		t	$t+1$	$t+2$	$t+3$	$t+4$	"total"
				our reference price			
	Heinz	−4.45	1.84	0.59	0.28	0.17	−1.65
t	Hunts	−6.07	2.56	0.80	0.40	0.26	−1.86
	Del Monte	−8.54	4.49	1.47	0.74	0.39	−1.37
	Rest	−9.08	3.70	1.41	0.77	0.53	−2.67
				exp. weighted average			
	Heinz	−5.28	1.58	0.75	0.33	0.15	−2.55
t	Hunts	−7.45	2.17	1.02	0.50	0.26	−3.31
	Del Monte	−8.77	2.57	1.27	0.65	0.29	−3.92
	Rest	−8.33	1.19	0.69	0.42	0.26	−5.71

simulation study. We are not aware of studies in the literature which do something similar, except for a study by Erdem *et al.* (2003) in which the effects of future price expectations on household demand are considered. Households forming price expectations are forward-looking, whereas households forming internal reference prices are backward-looking.

In the simulation study we investigate two scenarios. In the first situation the price of the considered brand is *temporarily* adjusted by 1% at the current purchase occasion, whereas in the second setting this price is kept at its original level. For both scenarios, future prices are unchanged. Moreover, the prices of all other brands remain unaffected. By comparing the current and future brand choice probabilities for the two settings, one can obtain estimates of both the instantaneous sensitivity of own brand choice and the lagged response via the reference price effect. For each purchase occasion in the estimation sample which is followed by at least $L = 4$ future purchase occasions, we compute the current brand choice probabilities and the brand choice probabilities for $L = 4$ periods ahead. Proxies for current and future sales are obtained by adding the corresponding probabilities over different purchase occasions. We repeat this procedure 1000 times, and we add the "sales" values obtained from all runs. For each run, segment membership of each household is drawn from its posterior distribution, and the price recall states are drawn from the corresponding first-order Markov processes.

Table 3.4 reports the percent change in sales between the two scenarios, where we distinguish between current sales and the sales levels up to $L = 4$ periods ahead. Furthermore, the table shows the "total effect" of the temporary price change. This effect is in percents too. It is computed as the ratio of the net change in sales over time and the current sales level for the scenario without price adjustment. For comparison, Table 3.4

also contains the results for the exponentially-smoothed-average reference price model. We note that the previous-price model cannot capture dynamic effects lasting for more than one period, while a brand choice model without any reference price component can only capture instantaneous effects.

Table 3.4 illustrates that the impact of internal reference price formation on sales is substantial, even though price recall of households is very limited. Although the instantaneous own price elasticities of all brands are located in between -10 and -4, the net effects after $L = 4$ subsequent purchase occasions are rather close to zero, that is, all net price elasticities are larger than -3. For all brands, the net effect of a price change on current and future sales is less than 40% of its instantaneous effect. A comparison with the results obtained for the exponentially-weighted-average reference price model illustrates that the reference effect implied by our model is much stronger, in an absolute sense as well as in a relative sense.

3.6 Conclusions

In this chapter, we have proposed a new reference price framework for brand choice. Price recall of households is modeled as a hidden Markov-switching process with an absorbing state, and an internal reference price is constructed from the prices the household is able to recall. Features of our reference price model include that (i) it does not a priori impose a rigid dynamic price structure, (ii) it accounts for uncertainty in reference price formation, (iii) it allows that a household may forget past prices, and (iv) it even allows that a household cannot construct a reference price at all. Our model can be used to analyze how many prices observed in the past are considered for reference price formation, and to what extent households have sufficient price knowledge to form an internal reference price.

Applied to A.C. Nielsen scanner panel data on catsup purchases, our model has a good performance relative to two popular reference price specifications. Our main findings are as follows. First, a price supported by a feature/display has a higher probability to be recalled than a price which is not supported by such promotional activity. Second, contrary to the two competing reference price models, our model does not indicate asymmetry between price gains and losses. However, our model suggests a stronger and more significant reference price effect. This is an indication that the impact of internal reference price formation on brand choice might be even larger than is currently believed. Third, we find that the prices observed at the previous purchase occasion have an average recall probability of about 20%. We estimate the average probability that a household has sufficient price knowledge to form an internal reference price at about 30%. Fourth, even though price recall is very limited, the impact of reference price formation on brand choice is very substantial. For all brands in the data set, the net effect of a price discount on current and future sales is less than 40% of its instantaneous effect.

We conclude by mentioning some limitations of our framework and we provide some suggestions for further research. A conceptual remark about our model is that we interpret "recalling a price" as being equivalent to "using this price for reference price formation". We report price recall probabilities, but in fact we can only make inference on the probabilities that past prices show up in the reference price. It might be possible that a household is able to recall prices observed in the past, but still does not use this price information to form a reference price. However, this is only a matter of interpretation. A second issue is that households might have different degrees of price knowledge, see, for example, Vanhuele and Drèze (2002) who distinguish between recallable price knowledge, price recognition and deal spotting. Our model does not account for this. An interesting suggestion for further research would be to allow for such limited price knowledge in a reference price model for brand choice. Finally, we have focussed on internal reference prices, constructed from past prices, and we have not allowed for an external reference price, based on the in-store prices during the purchase occasion. However, several studies indicate that both are important and can exist at the same time, see Mayhew and Winer (1992), Rajendran and Tellis (1994) and Mazumdar and Papatla (2000).

Chapter 4

A Markov model for market shares

4.1 Introduction and motivation

In the previous two chapters, we have analyzed the effects of inventory-dependent consumption and internal reference price formation on household demand. We have inferred these effects from household scanner panel data. Such micro data provide detailed information about purchase behavior of households and allow for analyses which cannot be carried out using more aggregate data, typically store-level data containing the weekly marketing-mix and sales levels in a category in a store. However, for household scanner panel data there are some problems too. One of these is the availability, that is, scanner panel data are relatively scarce, and it is often expensive (in terms of effort and money) to acquire them. A second problem is the representativeness of such data, depending on the size of the panel and on the way participating households have been selected. The representativeness issue for household scanner panel data is addressed by Gupta *et al.* (1996). Russell and Kamakura (1994) propose an integrated approach to take advantage of both the information richness of micro data and the representativeness of aggregate data. However, their approach does not avoid the availability of scanner panel data.

In this chapter, we turn our focus to weekly store-level data. A useful measure for the performance of a brand relative to the other brands in a market is the weekly share in category sales. To predict market shares and to link market shares with marketing instruments like price and promotion, one often considers a regression type model. A popular format for such a model is the market share attraction model which takes into account that shares are positive and sum to unity, see Cooper and Nakanishi (1988) and Cooper (1993) for detailed early treatments of the model and Fok *et al.* (2002) for a recent review of its econometric properties. Such an attraction model delivers, among other things, estimates of own and cross price elasticities, which in turn can be used to draw inference on the competitive structure. With these elasticities, one can infer what might happen to the own brand market share if the own price is lowered or when competitors lower their prices.

Market share models are useful to sketch competitive structures. However, there is one thing these models cannot do, and that is, to help to understand which brands lose share to which brands. To make this statement more precise, think about the following situation. There are two brands, A and B, and we observe weekly sales of these brands. Suppose for the moment that the same number of households visits the store in each week. If the price of, say, A is lowered in a certain week, we might observe more sales of A and less of B. This information might be useful to see if lower prices generate more sales, but it is insufficient to understand whether the change in sales of A can be fully attributed to households changing from B to A, or whether some buyers of A have also switched to B. Hence, the question where the new sales come from cannot be addressed by simply looking at weekly sales or shares data. Nonetheless, it is a managerially very relevant question.

In this chapter, we therefore develop a method to infer share-switching from store-level scanner data. Share-switching is related to aggregate market shares in a similar fashion as brand-switching is related to individual brand choice behavior. Our model can be used to get a better understanding of the competitive structure. For example, consider again two brands A and B. It might happen that the cross price elasticities suggest that A and B are not strong price competitors, while still much switching occurs between these two brands. In that case, competition between A and B might be due to low customer loyalty for both brands. An explanation for such competition is that households behave as variety-seekers, that is, households might derive utility from brand-switching itself, besides the utility resulting from the selected brand. In such a case, increasing customer loyalty should be more important than pricing strategies. McAlister and Pessemier (1982) provide a classification scheme of the variety-seeking literature.

A second contribution of our model is that it allows for a decomposition of own and cross price elasticities. For example, a decrease of the own price is likely to result in an increase of the own market share. This increased market share is a result of (i) retaining more of the own share (less switching away from the own brand) and (ii) gaining additional share from the other brands (more switching towards the own brand). The decomposition from our model makes it possible to quantify these partial effects. This is an additional insight which cannot be obtained from market share attraction models.

The structure of this chapter is as follows. In Section 4.2, we outline in a little more detail the situation, which was already briefly touched upon above, concerning weekly sales of brands and week-to-week share-switching. Next, in Section 4.3, we put forward our model, and in Section 4.4, we explain how the model can be used to get insight into the competitive structure. Parameter estimation is discussed in Section 4.5. In Section 4.6, we apply our method to four store-level data sets concerning two product categories. Finally, we conclude in Section 4.7 with a discussion of the implications of our method. We also provide some directions for further research.

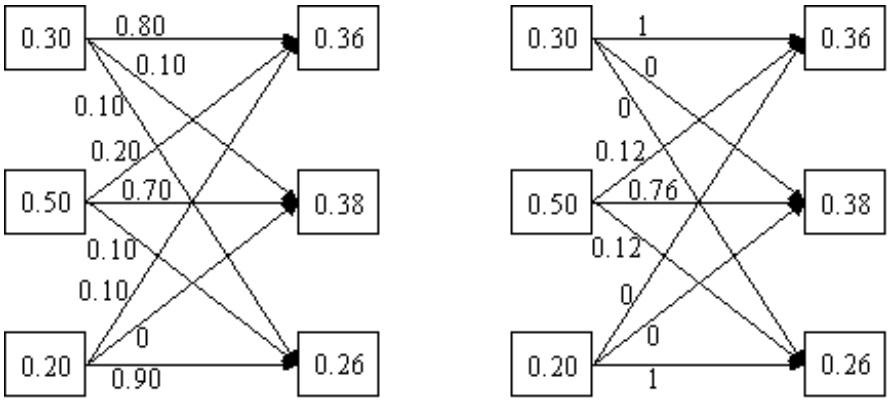

Figure 4.1: A numerical example of share-switching.

4.2 On store-level data and share-switching

Consider the case where we have weekly store-level scanner data, for $t = 1, \ldots, T$. For the moment, we assume that $T = 2$. Figure 4.1 sketches a possible situation involving the market shares of three brands A, B and C. At $t = 1$ the three brands have shares of 30%, 50% and 20%, while the shares at $t = 2$ are given by 36%, 38% and 26%, respectively. These observed changes in market share from $t = 1$ to $t = 2$ are the result of unobserved share-switching from and to each of the three brands, which is represented by the arrows. The numbers on the arrows indicate which fraction of the market share at $t = 1$ is transferred. So, in the left panel of Figure 4.1, brand A retains 80% of its market share, it loses 10% of its share to brand B, and it loses another 10% to brand C. Similarly, brand B retains 70% of its share, and it loses 20% and 10% to brand A and brand C, respectively. It is further seen that brand C retains 90% of its share, whereas the remaining 10% moves to brand A. It is easily checked that these switching fractions indeed give the market shares at time $t = 2$. For example, at $t = 2$ brand A has a share of 36%, which is $0.80 \times 30\% + 0.20 \times 50\% + 0.10 \times 20\%$. However, these share-switching fractions do not uniquely determine the market shares at $t = 2$, and the right panel gives another set of share-switching fractions resulting in the same shares. When the length T of the observed period becomes sufficiently large, unique share-switching estimates can be obtained. Share-switching may also be related to marketing activity. For example, a price decrease for brand A at $t = 2$ might result in additional share-switching from each of the three brands at $t = 1$ to brand A at $t = 2$, where increased share-switching from A to A amounts to increased customer retention.

4.3 The model

In this section, we develop our share-switching model. To this end, we define the market share of brand j at time t by $M_{j,t}$, $j = 1, \ldots, J$, $t = 1, \ldots, T$. Furthermore, we define

the fraction of the market share $M_{l,t-1}$ of brand l moving to brand k at time t by $\lambda_{l,k,t}$. It is convenient to collect the market shares observed at time t in the vector $M_t = (M_{1,t}, \ldots, M_{J,t})'$ and to collect the share-switching fractions at time t in the matrix

$$\Lambda_t = \begin{pmatrix} \lambda_{1,1,t} & \cdots & \lambda_{1,J,t} \\ \vdots & \ddots & \vdots \\ \lambda_{J,1,t} & \cdots & \lambda_{J,J,t} \end{pmatrix}.$$

For example, in the left panel of Figure 4.1 we would have

$$M_{t-1} = \begin{pmatrix} 0.30 \\ 0.50 \\ 0.20 \end{pmatrix}, \qquad M_t = \begin{pmatrix} 0.36 \\ 0.38 \\ 0.26 \end{pmatrix}, \qquad \Lambda_t = \begin{pmatrix} 0.80 & 0.10 & 0.10 \\ 0.20 & 0.70 & 0.10 \\ 0.10 & 0 & 0.90 \end{pmatrix}.$$

4.3.1 General structure of the model

The share-switching fractions as contained in Λ_t satisfy two conditions. First, these share-switching fractions just redistribute each brand's market share, so that they have to sum to one for each given brand, that is,

$$\sum_{k=1}^{J} \lambda_{l,k,t} = 1, \qquad l = 1, \ldots, J, \tag{4.1}$$

or in matrix notation,

$$\Lambda_t \, \iota_J = \iota_J, \tag{4.2}$$

where ι_J denotes the $J \times 1$ vector consisting of ones. Second, by definition, the current market share of each brand is the sum of the portions carried over to that brand from each brand's previous market share. This leads to the first-order Markov equation

$$M_{k,t} = \sum_{l=1}^{J} \lambda_{l,k,t} M_{l,t-1}, \qquad k = 1, \ldots, J, \tag{4.3}$$

which is also the key equation in the aggregate Markov model of Leeflang (1974), see also Leeflang *et al.* (2000). In matrix notation, it is given by

$$M_t = \Lambda_t' M_{t-1}. \tag{4.4}$$

We note that our model differs from the model of Leeflang (1974) in various respects. Advantages of our model include that (i) estimation results are invariant with respect to which market share equation is disregarded to account for the sum constraint, (ii) parameter estimation is based on simple maximum likelihood and does not require optimization under inequality conditions, and (iii) the involved covariance matrix is positive definite by construction.

We have already mentioned that the share-switching fractions $\lambda_{l,k,t}$ might be correlated to marketing activity. However, even if such observable variables are included, it cannot be expected that share-switching is fully explained. To account for this, we decompose the share-switching matrix Λ_t into a deterministic component $\tilde{\Lambda}_t$ and a random component E_t, that is,

$$\Lambda_t = \tilde{\Lambda}_t + E_t, \tag{4.5}$$

where

$$\tilde{\Lambda}_t = \begin{pmatrix} \tilde{\lambda}_{1,1,t} & \cdots & \tilde{\lambda}_{1,J,t} \\ \vdots & \ddots & \vdots \\ \tilde{\lambda}_{J,1,t} & \cdots & \tilde{\lambda}_{J,J,t} \end{pmatrix}, \qquad E_t = \begin{pmatrix} e_{1,1,t} & \cdots & e_{1,J,t} \\ \vdots & \ddots & \vdots \\ e_{J,1,t} & \cdots & e_{J,J,t} \end{pmatrix}.$$

All elements of the error matrix E_t have expectation zero, so that the expectation of the share-switching matrix Λ_t is $\tilde{\Lambda}_t$. Substituting (4.5) into (4.4) gives

$$M_t = \tilde{\Lambda}_t' M_{t-1} + E_t' M_{t-1}, \tag{4.6}$$

decomposing market shares into an explained and an unexplained component. This constitutes the first part of our model.

To interpret the elements of $\tilde{\Lambda}_t$ (which have to be estimated using actual data) as share-switching fractions, we require that the "redistribution condition" is maintained, that is,

$$\tilde{\Lambda}_t \, \iota_J = \iota_J. \tag{4.7}$$

By postmultiplying both sides of (4.5) by ι_J and substituting (4.2) and (4.7), it follows that the rows of the error matrix E_t have to sum to zero, that is,

$$E_t \, \iota_J = 0, \tag{4.8}$$

or

$$\sum_{k=1}^{J} e_{l,k,t} = 0, \qquad l = 1, \ldots, J. \tag{4.9}$$

Next, by transposing (4.6), postmultiplying both sides by ι_J, and incorporating the conditions (4.7) and (4.8), it follows that

$$\begin{aligned} & M_t' \, \iota_J = M_{t-1}' \tilde{\Lambda}_t \, \iota_J + M_{t-1}' E_t \, \iota_J \\ \Leftrightarrow \; & M_t' \, \iota_J = M_{t-1}' \, \iota_J + M_{t-1}' \, 0 \\ \Leftrightarrow \; & 1 = 1 + 0. \end{aligned} \tag{4.10}$$

So, both the market shares $M_t = \Lambda_t' M_{t-1}$, *generated* from our model given M_{t-1}, and the market shares $\tilde{\Lambda}_t' M_{t-1}$, *predicted* from our model, sum to unity. That is, the sum condition for market shares is always met. On the other hand, the range condition that market shares are restricted between zero and one does not necessarily hold. This range condition is satisfied for *predicted* market shares, provided that the elements of $\tilde{\Lambda}_t$ are non-negative. However, there is no guarantee for *generated* market shares, because of the

additive way the disturbance E_t is incorporated in Λ_t, see (4.5). Although an additive structure might result in violation of the range condition for generated market shares, we still opt for it, as (i) it allows for closed-form expressions, keeping estimation of the model parameters fast and simple, and (ii) predicted market shares in which one is often interested still satisfy both the sum condition and the range condition. It follows from (4.10) that one of the J market share equations in (4.6) is redundant, so that only the parameters of $J-1$ equations have to be estimated.

4.3.2 Detailed structure of the model

We complete the model by specifying the elements of $\tilde{\Lambda}_t$ and by assuming a distribution for the elements of the error matrix E_t. First, we define the elements of $\tilde{\Lambda}_t$ by the logit structure

$$\tilde{\lambda}_{l,k,t} = \frac{\exp(\alpha_{l,k} + x'_{k,t}\beta)}{\sum_{j=1}^{J} \exp(\alpha_{l,j} + x'_{j,t}\beta)}, \qquad l = 1, \ldots, J, \qquad k = 1, \ldots, J, \tag{4.11}$$

so that the redistribution condition (4.7) and the non-negativity requirement for $\tilde{\Lambda}_t$ are indeed satisfied. Similar to the conditional logit model of McFadden (1974), not all intercept parameters $\alpha_{l,k}$ are identified. For identification purposes, we set $\alpha_{l,J} = 0$, $l = 1, \ldots J$. These parameters correspond to the redundant J-th market share equation. It is seen from (4.11) that each share-switching fraction $\tilde{\lambda}_{l,k,t}$ from brand l to brand k has its own intercept with parameter $\alpha_{l,k}$, and that $\tilde{\lambda}_{l,k,t}$ depends on the marketing-mix variables of the receiving brand k through the shared response parameters β.

We consider all brand-pair specific intercepts to relate current market shares to previous market shares. This is similar to what Lattin and McAlister (1985) do by including all brand-pair specific similarity parameters in a first-order Markov model for brand choice. Related to that research, Che *et al.* (2003) operationalize a similarity variable whose response parameter indicates the extent to which households are seeking or avoiding variety in their brand choice decision.

We further note that $x_{k,t}$ in (4.11) only contains the index k of the receiving brand, so that share-switching is not affected by the marketing-mix of the supplying brand l. This is consistent with the brand-switching models of Givon (1984) and Seetharaman and Chintagunta (1998), which aim to quantify variety-seeking and inertia in brand choice. In (4.11), the marketing-mix variables in $x_{k,t}$ are in first-differences, as share-switching corresponds to *changes* in market shares, rather than absolute market share levels. We emphasize that, as in the conditional logit model, the estimation results of our share-switching model are invariant with respect to which brand is taken as the base brand J.

For the share-switching errors $e_{l,k,t}$ we would want to consider a distribution which is, on one hand, parsimonious in its parameterization, and which is, on the other hand, flexible enough to reflect uncertainty patterns in share-switching. We assume that the $e_{l,k,t}$

are distributed independently as $N(0, \sigma^2_{l,k,t})$ with $\sigma^2_{l,k,t} = \sigma^2 M^{\gamma_1}_{l,t-1} M^{\gamma_2}_{k,t-1}$. So, the share-switching errors obey a normal distribution, and they are allowed to be heteroscedastic in the sense that the variance might depend on the market shares of the supplying brand l and the receiving brand k. One might expect that $\gamma_1 < 0$ and $\gamma_2 > 0$, that is, there is more uncertainty for share-switching from a weak brand with a small market share to a strong brand with a large market share, and there is less uncertainty for share-switching in the opposite direction.

Several restrictions can be imposed on the heteroscedasticity structure. For example, if one believes that only the market share of the receiving brand matters, one can set $\gamma_1 = 0$. But one can also consider the parameter restriction $\gamma_1 = -\gamma_2$, resulting in $\sigma^2_{l,k,t} = \sigma^2 \big(\frac{M_{k,t-1}}{M_{l,t-1}}\big)^{\gamma_2}$, so that the variance directly depends on the size of the receiving brand relative to the size of the supplying brand.

In Appendix 4.A, it is shown that if the elements $e_{l,k,t}$ of E_t are independently $N(0, \sigma^2 M^{\gamma_1}_{l,t-1} M^{\gamma_2}_{k,t-1})$ distributed *before* imposing the error summation restrictions (4.9), the restriction in (4.9) implies that our final model reads as

$$\tilde{M}_t = \begin{pmatrix} M_{1,t} \\ \vdots \\ M_{J-1,t} \end{pmatrix} \sim N(\, \mu_t, \sigma^2 V_t \,), \tag{4.12}$$

where

$$\mu_t = \begin{pmatrix} \sum_{l=1}^{J} \tilde{\lambda}_{l,1,t} M_{l,t-1} \\ \vdots \\ \sum_{l=1}^{J} \tilde{\lambda}_{l,J-1,t} M_{l,t-1} \end{pmatrix}, \tag{4.13}$$

and

$$V_t = \sum_{l=1}^{J} M^{2+\gamma_1}_{l,t-1} \left[-\frac{1}{\sum_{j=1}^{J} M^{\gamma_2}_{j,t-1}} \begin{pmatrix} M^{\gamma_2}_{1,t-1} \\ \vdots \\ M^{\gamma_2}_{J-1,t-1} \end{pmatrix} \big(M^{\gamma_2}_{1,t-1} \cdots M^{\gamma_2}_{J-1,t-1} \big) + \operatorname{diag}\Big(M^{\gamma_2}_{1,t-1}, \ldots, M^{\gamma_2}_{J-1,t-1} \Big) \right]. \tag{4.14}$$

It is seen from (4.14) that the restriction in (4.9) induces that the market shares are negatively correlated. This is a desirable feature, as, by definition, a larger market share for one brand results in a smaller total market share for all other brands.

4.3.3 Model extension: inclusion of an outside good

In the discussion so far, we have defined $M_t = (M_{1,t}, \ldots, M_{J,t})'$ as a vector containing the market shares of the J brands. A straightforward and useful extension would be to redefine M_t such that it also includes the market share $M_{J+1,t}$ of an outside good

$J+1$, representing the share of customers who do not make a purchase at time t. This outside good would become the natural base brand, so that $\alpha_{l,J+1} = 0$, $l = 1, \ldots J+1$, for parameter identification. Furthermore, one can set $x_{J+1,t} = 0$, as by definition the outside good does not have its own marketing-mix.

As (i) the dimension of M_t increases from J to $J+1$, (ii) the no-purchase share equation $J+1$ is redundant, and (iii) $\exp(\alpha_{l,J+1} + x'_{J+1,t}\beta) = 1$, $l = 1, \ldots, J+1$, an extended model becomes

$$\tilde{M}_t = \begin{pmatrix} M_{1,t} \\ \vdots \\ M_{J,t} \end{pmatrix} \sim N(\,\mu_t, \sigma^2 V_t\,), \tag{4.15}$$

where

$$\mu_t = \begin{pmatrix} \sum_{l=1}^{J+1} \tilde{\lambda}_{l,1,t} M_{l,t-1} \\ \vdots \\ \sum_{l=1}^{J+1} \tilde{\lambda}_{l,J,t} M_{l,t-1} \end{pmatrix}, \tag{4.16}$$

$$V_t = \sum_{l=1}^{J+1} M_{l,t-1}^{2+\gamma_1} \left[-\frac{1}{\sum_{j=1}^{J+1} M_{j,t-1}^{\gamma_2}} \begin{pmatrix} M_{1,t-1}^{\gamma_2} \\ \vdots \\ M_{J,t-1}^{\gamma_2} \end{pmatrix} \left(M_{1,t-1}^{\gamma_2} \cdots M_{J,t-1}^{\gamma_2} \right) + \operatorname{diag}\left(M_{1,t-1}^{\gamma_2}, \ldots, M_{J,t-1}^{\gamma_2} \right) \right], \tag{4.17}$$

and

$$\tilde{\lambda}_{l,k,t} = \frac{\exp(\alpha_{l,k} + x'_{k,t}\beta)}{1 + \sum_{j=1}^{J} \exp(\alpha_{l,j} + x'_{j,t}\beta)}, \qquad l = 1, \ldots, J+1, \qquad k = 1, \ldots, J. \tag{4.18}$$

We note that all intercept parameters $\alpha_{l,k}$ in (4.18) are identified.

Inclusion of the share of no-purchase in the model allows for a richer analysis, as the analysis would not only consider how category sales are distributed across brands, but also how total category demand changes over time. For example, it might well be that absolute sales of a brand decrease at time t, while the corresponding market share increases. A model without an outside good would only focus on the increased market share (the secondary demand effect), whereas the sales decrease (the primary demand effect) might be managerially more relevant.

Van Heerde *et al.* (2003) put forward a unit sales decomposition for the primary and secondary demand effects of a price promotion. Using this decomposition, they find that the primary demand effect is larger than was previously believed, see, for example, Gupta (1988). This is a second argument to keep track of the no-purchase share.

Third, if the no-purchase option is not accounted for, share-switching from, say, brand A to brand B should not be related to the number of customers actually switching from

A to B. For example, if some of the customers of brand A at time t do not make a purchase at time $t+1$, the share of A would decrease relative to the share of B, implying a transfer of market share from A to B. Hence, although the notion of share-switching is still valid if the model does not include an outside good, its interpretation is little more difficult. Additionally, a model which does not allow for an outside good implicitly assumes that (in the absence of marketing activity effects) the share-switching dynamics across brands are fixed over time. This condition does not necessarily hold, as households can be expected to have heterogeneous interpurchase times, so that different households may buy in different weeks.

An important issue is whether purchase timing is related to brand-switching. For example, consider a market for a product category with only two brands, A and B. Furthermore, suppose there exists substantial heterogeneity in purchase timing across households, and the market can be split up into two groups of households. The first group is completely loyal to brand A, whereas the second group considers both brands. If there is no clear relationship between a household's loyalty and purchase frequency, the effect of not including an outside good on the share-switching estimates should be reasonably small. But now suppose that the group of potential switchers consists of frequent purchasers only and the group of loyal households consists of infrequent purchasers only. In that particular case, share-switching between brand A and brand B cannot be considered fixed over time, as the overall tendency to purchase A or B differs between current purchasers and current non-purchasers. A relatively large group of non-purchasers would indicate that very few households loyal to brand A made a purchase at time t, implying relatively much growth potential for brand A at time $t+1$, compared to the potential for brand B. Again, such a transfer of market share would not be induced by customers actually switching from brand B to brand A at time $t+1$, but would be induced by the interaction with the no-purchase option.

4.4 Interpretation

In this section, we discuss how our share-switching model can provide insight into the competitive structure. The well-known market share attraction model is often used to compute own and cross price elasticities to infer what might happen to the own brand market share if the own price is lowered or when competitors lower their prices. Our model can deliver such elasticities as well. Additionally, our model can be used to infer the magnitudes of share-switching across brands. Traditional market share models, such as the attraction model, do not allow for this.

To derive the price elasticities, we first introduce some additional notation. We recall that $\tilde{\lambda}_{l,k,t}$ is interpreted as the *fraction* of the share of brand l moving to brand k at time t. Hence, $\tilde{\lambda}_{l,k,t}$ corresponds to share-switching *conditional* on the supplying brand l, that is, the share-switching fractions away from l sum to unity, or $\sum_{k=1}^{J} \tilde{\lambda}_{l,k,t} = 1$. As an additional measure, we define the share-switching *portion* $\bar{\lambda}_{l,k,t} = M_{l,t-1}\tilde{\lambda}_{l,k,t}$, which relates

switching from brand l to brand k to the total market. Clearly, $\sum_{l=1}^{J}\sum_{k=1}^{J}\bar{\lambda}_{l,k,t}=1$. The share-switching portion $\bar{\lambda}_{l,k,t}$ is an *unconditional* measure, which is in contrast to the share-switching fraction $\tilde{\lambda}_{l,k,t}$.

We define the share-switching price elasticity $\delta_{l\to k,j,t}$ as

$$\delta_{l\to k,j,t}=\frac{\partial\bar{\lambda}_{l,k,t}}{\partial P_{j,t}}\,\frac{P_{j,t}}{\bar{\lambda}_{l,k,t}}=\frac{\partial\tilde{\lambda}_{l,k,t}}{\partial P_{j,t}}\,\frac{P_{j,t}}{\tilde{\lambda}_{l,k,t}}\,, \tag{4.19}$$

where $P_{j,t}$ denotes the price of brand j at time t. The interpretation of (4.19) is that a one percent price increase of brand j at time t results in $\delta_{l\to k,j,t}$ percent additional share-switching from brand l to brand k. We note that $\delta_{l\to k,j,t}$ is not necessarily zero if $j\neq l,k$, that is, when the price adjustment concerns a brand different from the two brands l and k for which share-switching is considered. For example, if brand j lowers its price, it might gain some share from brand l which would have gone to brand k otherwise.

If price is one of the marketing-mix variables included in (4.11), with response parameter β_p, then it immediately follows from the logit structure of (4.11) that

$$\delta_{l\to k,j,t}=\begin{cases}\beta_p\,P_{k,t}\,(1-\tilde{\lambda}_{l,k,t}) & \text{if } k=j\\ -\beta_p\,P_{j,t}\,\tilde{\lambda}_{l,j,t} & \text{if } k\neq j\end{cases}\,, \tag{4.20}$$

see, for example, Ben-Akiva and Lerman (1985). Although prices are considered in first-differences in (4.11), the $P_{j,t}$ in (4.19) and (4.20) are price levels.

The share-switching price elasticities above determine the reaction of market shares to price changes, as market shares are eventually the result of share-switching. For our model, it can be shown that the elasticity of the expected market share $E(M_{k,t})=\sum_{l=1}^{J}\bar{\lambda}_{l,k,t}$ with respect to the price $P_{j,t}$ is given by

$$\delta_{k,j,t}=\frac{\partial E(M_{k,t})}{\partial P_{j,t}}\,\frac{P_{j,t}}{E(M_{k,t})}=\frac{\sum_{l=1}^{J}\bar{\lambda}_{l,k,t}\,\delta_{l\to k,j,t}}{\sum_{i=1}^{J}\bar{\lambda}_{i,k,t}}\,. \tag{4.21}$$

This market-share price elasticity can also be written as

$$\delta_{k,j,t}=\sum_{l=1}^{J}\eta_{l\to k,j,t}\qquad\text{with}\qquad\eta_{l\to k,j,t}=\frac{\bar{\lambda}_{l,k,t}}{\sum_{i=1}^{J}\bar{\lambda}_{i,k,t}}\,\delta_{l\to k,j,t}, \tag{4.22}$$

where $\eta_{l\to k,j,t}$ is the contribution of share-switching from brand l to brand k in the overall effect of a price adjustment of brand j on the market share of brand k. So, (4.22) provides a decomposition of the reaction of market share into components attributable to changes in share-switching from each of the brands towards the brand whose market share is affected.

The decomposition (4.22) provides interesting insights which cannot be obtained from market share attraction models. For example, a decrease of the own price can be expected to result in an increased market share by (i) retaining a larger part of the own current share (own customers have a smaller incentive to switch away) and (ii) taking over additional

share from competing brands (customers of competing brands have a larger incentive to switch). The part of the increased market share attributable to retaining more of the own share is $\eta_{k\to k,j,t}/\delta_{k,j,t}$ with $j = k$, whereas the part attributable to gaining more of the competitive share is given by $1 - (\eta_{k\to k,j,t}/\delta_{k,j,t})$. In other words, the decomposition can be used to analyze which part of the reaction of the own market share to an own price reduction is caused by increased customer retention, and which part is not. We note that each component $\eta_{l\to k,j,t}$ in the decomposition (4.22) consists of the share-switching price elasticity $\delta_{l\to k,j,t}$ multiplied by a weight proportional to the corresponding share-switching portion $\bar{\lambda}_{l,k,t}$. Hence, contributions are determined by both the price-sensitivity and the relative magnitude of corresponding share-switching.

The price elasticities defined by (4.21) amount to a generalization of a result established by Bucklin *et al.* (1998b). This result is a relationship between price elasticities and aggregate brand-switching, which has been derived on the basis of the conditional logit model. Basically, it states that[1]

$$\delta_{k,j,t} = \begin{cases} \beta_p \, P_{k,t} \, (1 - \tilde{\lambda}_{k,k,t}) & \text{if } k = j \\ -\beta_p \, P_{j,t} \, \tilde{\lambda}_{k,j,t} & \text{if } k \neq j \end{cases}, \tag{4.23}$$

which, using (4.20), boils down to

$$\delta_{k,j,t} = \delta_{k\to k,j,t}. \tag{4.24}$$

In terms of our model, this relationship states that the reaction of market share to a price adjustment is only caused by the effect on customer retention, and not by the effect on share-switching from competing brands. This is more restrictive than our price elasticity specification.

4.5 Parameter estimation

The parameters of our share-switching model can be estimated using maximum likelihood [ML]. The parameter estimates result from maximization of the log-likelihood function, which, in the absence of an outside good, is given by

$$\begin{aligned} \ln \mathcal{L} = &-\frac{(T-1)(J-1)}{2} \left[\ln(2\pi) + \ln(\sigma^2)\right] + \frac{1}{2} \sum_{t=2}^{T} \log\det(V_t^{-1}) \\ &-\frac{1}{2\sigma^2} \sum_{t=2}^{T} (\tilde{M}_t - \mu_t)' \, V_t^{-1} (\tilde{M}_t - \mu_t), \end{aligned} \tag{4.25}$$

where μ_t and V_t are defined by (4.13) and (4.14), respectively. We note that the inverse of V_t is given by $V_t^{-1} = \left(\sum_{l=1}^{J} M_{l,t-1}^{2+\gamma_1} \right)^{-1} \left[M_{J,t-1}^{-\gamma_2} \, \iota_{J-1} \, \iota_{J-1}' + \mathrm{diag}(M_{1,t-1}^{-\gamma_2}, \ldots, M_{J-1,t-1}^{-\gamma_2}) \right]$,

[1] Bucklin *et al.* (1998b) actually consider the logarithm of price in the conditional logit model, so that the price term disappears in their result.

see Appendix 4.A, and that it is straightforward to adjust (4.25) for inclusion of an outside good in the model. Standard errors are obtained by taking the square roots of the diagonal elements of the estimated covariance matrix, which, in turn, can be computed as minus the inverse of the Hessian of (4.25) evaluated for the optimal parameter values. Numerical techniques, such as the BFGS algorithm or the Newton-Raphson algorithm, have to be used to get the ML parameter estimates.

4.6 Empirical analysis

In this section, we apply our model to weekly store-level scanner data, drawn from the ERIM data base (GSB, University of Chicago) for catsup and peanut butter in the Sioux Falls SD market (henceforth referred to as market 1) and in the Springfield MO market (henceforth referred to as market 2). The considered period consists of 124 weeks from 1985 to 1988. The last 20 weeks are used for out-of-sample forecasting. For each product category, the number of brands is restricted to four by considering the three largest brands and a "rest brand". This Rest brand represents several smaller brands as well as private labels. We allow for marketing-mix effects by including price and 0/1 display variables. Table 4.1 contains a few summary statistics. We note that the prices of the Rest brands have been computed by taking the ratio of corresponding aggregate sales in dollars and aggregate sales in units. It can be seen that, in general, the price of the Rest brand is lower than the prices of the other three brands. Furthermore, the table shows that for both catsup and peanut butter the Rest brand has a larger market share in the Springfield area than in the Sioux Falls area.

4.6.1 Parameter estimates

In the subsequent discussion of our share-switching model, we focus on the heteroscedasticity specification $\sigma^2_{l,k,t} = \sigma^2 M^{\gamma_2}_{k,t-1}$, so that the model is given by (4.11)−(4.14) after setting $\gamma_1 = 0$. However, we have also considered other specifications. For the unrestricted variance $\sigma^2_{l,k,t} = \sigma^2 M^{\gamma_1}_{l,t-1} M^{\gamma_2}_{k,t-1}$, we find that γ_1 is not significant at the 10% level for three of the four data sets. This parameter is only significant at the 5% level for the Springfield catsup market. On the other hand, γ_2 is significant at 1% for three of the four category-markets, and it is still significant at 5% for the remaining data set. So, it seems that the effect of the supplying brand l on the variance is dominated by the effect of the receiving brand k. Even after setting $\gamma_2 = 0$, the variance parameter γ_1 turns out to be insignificant at the 10% level for two of the four data sets. Hence, we set $\gamma_1 = 0$ for all four category-markets, and we continue with $\sigma^2_{l,k,t} = \sigma^2 M^{\gamma_2}_{k,t-1}$.

In the empirical applications, we focus on how *given* category sales are distributed across brands, and we do not deal with category expansion effects. Although inclusion of an outside good would certainly have its benefits, as discussed earlier, it might be difficult to construct this outside good requiring additional data information such as store traffic,

Table 4.1: A summary of the data, that is, the average values and the standard deviations (given in parentheses) of the market shares, prices and 0/1 display variables.

	share		price		display	
catsup 1						
Heinz	0.51	(0.22)	1.16	(0.10)	0.41	(0.49)
Hunts	0.26	(0.19)	1.09	(0.09)	0.54	(0.50)
Del Monte	0.17	(0.16)	1.09	(0.09)	0.40	(0.49)
Rest	0.07	(0.08)	0.87	(0.05)	0.09	(0.28)
catsup 2						
Heinz	0.37	(0.16)	1.34	(0.07)	0.68	(0.47)
Hunts	0.22	(0.14)	1.34	(0.11)	0.75	(0.43)
Del Monte	0.10	(0.09)	1.37	(0.09)	0.56	(0.50)
Rest	0.30	(0.11)	0.84	(0.07)	0.88	(0.33)
pbutter 1						
Skippy	0.31	(0.19)	1.68	(0.12)	0.39	(0.49)
Jif	0.25	(0.13)	1.72	(0.11)	0.17	(0.38)
Peter Pan	0.25	(0.16)	1.68	(0.16)	0.35	(0.48)
Rest	0.19	(0.11)	1.36	(0.12)	0.19	(0.39)
pbutter 2						
Skippy	0.10	(0.06)	1.78	(0.21)	0.10	(0.31)
Jif	0.16	(0.07)	1.81	(0.20)	0.10	(0.30)
Peter Pan	0.36	(0.16)	1.70	(0.25)	0.67	(0.47)
Rest	0.37	(0.13)	1.23	(0.13)	0.66	(0.47)

see, for example, Chintagunta (2000). So, we illustrate our model for the simplest case possible, minimizing data requirements, but we emphasize that our modeling approach would extend immediately to a situation dealing with category expansion.

Table 4.2 reports the estimates of the response parameters β_1 and β_2 for price and display and the estimates of the two variance parameters σ^2 and γ_2. The effect of price on market share is significant at the 1% level and has the expected sign for both catsup and peanut butter in the two considered areas. Display has a significant effect (at the 5% level) with the expected sign for the two catsup markets, but the effect is not significant for the two peanut butter markets. Furthermore, the heteroscedasticity parameter γ_2 is positive and significant for all four category-market combinations. The sign of γ_2 is as expected.

Table 4.2: Estimates of the marketing-mix response parameters and the variance parameters. The estimated standard errors are given in parentheses.

	catsup 1	catsup 2	pbutter 1	pbutter 2
β_1 (price)	-0.860^{***}	-1.099^{***}	-0.686^{***}	-0.406^{***}
	(0.110)	(0.145)	(0.079)	(0.040)
β_2 (display)	0.452^{***}	0.383^{**}	-0.046	0.129
	(0.108)	(0.166)	(0.148)	(0.108)
σ^2	0.291^{***}	0.171^{***}	0.103^{***}	0.102^{***}
	(0.058)	(0.040)	(0.026)	(0.022)
γ_2	0.870^{***}	1.043^{***}	0.419^{***}	0.987^{***}
	(0.081)	(0.115)	(0.145)	(0.111)

** significant at 5%.
*** significant at 1%.

4.6.2 Model performance

To see how well our model performs in terms of in-sample fit and out-of-sample forecasting, we compare it with three popular alternatives. We emphasize here that these models do not have the same attractive interpretation opportunities as our model has.

The first competing model is the full effects attraction model in which the attraction of brand k at time t is defined by

$$A_{k,t} = \exp\left(\alpha_k + \sum_{s=1}^{S}\sum_{j=1}^{J} \beta_{(s)j,k} \ln(x_{(s)j,t}) + \varepsilon_{k,t}\right). \tag{4.26}$$

Here, α_k is a brand-specific intercept, $x_{(s)j,t}$ is the value of marketing-mix variable s for brand j at time t (in our case, $s = 1$ corresponds to price and $s = 2$ corresponds to display), $\beta_{(s)j,k}$ is the involved response parameter, and $\varepsilon_{k,t}$ is a disturbance term. The $\varepsilon_{k,t}$ are distributed independently as $N(0, \sigma^2_{k,t})$. We note that, in this full effects specification, the attraction of a brand may be affected by the marketing-mix variables of each of the considered brands.

Bell *et al.* (1975) show that under certain axioms the market share of a brand is proportional to its attraction, so that

$$M_{k,t} = \frac{A_{k,t}}{\sum_{i=1}^{J} A_{i,t}}. \tag{4.27}$$

The full effects attraction model is given by (4.26) and (4.27). We note that in case the variables $x_{(s)j,t}$ are considered in levels instead of logarithms, the model above is often referred to as an MNL model for market shares.

The second model we consider is the differential effects specification in which competition is restricted in the sense that the attraction of a brand is only affected by its own marketing-mix, that is, $\beta_{(s)j,k} = 0$ if $j \neq k$. The differential effects attraction model is defined by (4.27) and

$$A_{k,t} = \exp\left(\alpha_k + \sum_{s=1}^{S} \beta_{(s)k} \ln(x_{(s)k,t}) + \varepsilon_{k,t}\right). \tag{4.28}$$

As our share-switching model is dynamic by definition, we include lagged market shares in (4.26) and (4.28) in the same way as price and display are included. The literature often considers one common lagged market share parameter for all brands, see, for example, Brodie and De Kluyver (1984), Chen *et al.* (1994), Kumar and Heath (1990) and Naert and Weverbergh (1981). However, we find that this does not improve the predictions. For comparability with our share-switching model, we consider the MNL specifications of the two attraction models described above, that is, lagged market shares are in logarithms but prices and displays are in levels.

Finally, the third model we consider is a first-order vector autoregression, or VARX(1) model, with price and display included as exogenous variables. In the model, the market shares, prices and displays of the first $J-1$ brands are considered in deviation from the base brand J. Furthermore, the market shares are taken in logarithms, so that the dependent variables amount to log-ratios of market shares.

We estimate the parameters of the two attraction models by applying a linearizing transformation such that these models are also defined in terms of $J-1$ log-ratios of market shares, see, for example, Fok *et al.* (2002). For the VARX(1) model, we apply Ordinary Least Squares estimation to each of the $J-1$ equations separately. This is equivalent to maximum likelihood estimation, see, for example, Hamilton (1994). For all three competing models, we generate appropriate market share predictions using the simulation approach described in Fok *et al.* (2002).

Table 4.3 shows the in-sample and out-of-sample values of the Root Mean Squared Prediction Error [RMSPE] for the considered models. In general, the table indicates that our model does not perform best within sample, but it does deliver more accurate out-of-sample forecasts. To be more specific, for three of the four category-markets our model is outperformed in-sample by all three competitors. The in-sample fit is only comparable to the fit of the differential effects model for the Sioux Falls catsup market. A partial explanation for the in-sample result might be the parsimonious parameterization of our model. It only contains 16 parameters, just like the differential effects model, whereas the VARX(1) model contains 30 parameters and the full effects model even needs 43 parameters.

Table 4.3: In-sample fit and out-of-sample forecasting performance, reflected by the Root Mean Squared Prediction Error.

		catsup 1	catsup 2	pbutter 1	pbutter 2
in-sample	our model	0.142	0.098	0.116	0.078
	full effects	0.132	0.079	0.102	0.063
	diff. effects	0.142	0.085	0.109	0.073
	VARX(1)	0.135	0.082	0.106	0.066
out-of-sample	our model	0.147	0.103	0.120	0.084
	full effects	0.162	0.124	0.150	0.096
	diff. effects	0.135	0.118	0.158	0.092
	VARX(1)	0.139	0.111	0.148	0.101

On the other hand, Table 4.3 also shows that our model generates the most accurate out-of-sample forecasts for three of the four data sets. Actually, our share-switching model is the only model for which the in-sample and out-of-sample RMSPE values match closely. The other three models combine small RMSPE values in-sample with large out-of-sample forecasting errors. Now we have investigated the in-sample and out-of-sample performance of our model relative to three competing models, we can focus on the additional insights that our model can provide.

4.6.3 Price elasticity analysis

Table 4.4 contains the estimated market-share price elasticities $\delta_{k,j,t}$ for our model. Furthermore, it provides estimates of $\eta_{k\to k,j,t}/\delta_{k,j,t}$, indicating which proportions of the price elasticities can be attributed to changes in customer retention.

The retention effect

It can be seen from the 16 diagonal elements in the right-hand part of Table 4.4 that increased customer retention accounts for 29% to 50% of the total increase in own market share when the own price is lowered. On average, this is about 42%. So, for the four considered markets, we find that the majority of the gain in market share is caused by gaining additional share from competing brands, and not by increasing own customer retention. Similarly, the off-diagonal elements in the right-hand part of Table 4.4 show that the part of the decrease in own market share when a competing brand lowers its price, attributable to decreased own customer retention, is at most 50% in 39 out of 48 cases. On average, this retention effect is about 40%. So, also for a competitive price adjustment, the retention effect does not seem to be dominating.

Table 4.4: The left-hand part shows the estimated market-share price elasticities $\delta_{k,j,t}$. The index j indicates the brand whose price is changed, and the index k indicates the brand whose market share is affected. Correspondingly, the right-hand part shows the estimated shares $\eta_{k\to k,j,t}/\delta_{k,j,t}$ of changed customer retention for brand k in the price elasticities $\delta_{k,j,t}$. The reported values amount to averages over time.

		k				k			
		price elasticities				shares of retention effect			
	catsup 1								
	Heinz	−4.44	4.25	4.38	5.13	0.50	0.37	0.34	0.39
j	Hunts	2.03	−6.03	2.27	2.09	0.50	0.38	0.36	0.16
	Del Monte	1.42	1.55	−6.73	0.91	0.53	0.38	0.35	0.00
	Rest	0.43	0.36	0.23	−6.24	0.43	0.52	0.30	0.29
	catsup 2								
	Heinz	−4.13	2.86	3.84	2.21	0.47	0.44	0.47	0.42
j	Hunts	1.65	−6.63	2.76	2.07	0.47	0.46	0.39	0.48
	Del Monte	0.77	0.93	−8.45	0.32	0.41	0.50	0.40	0.62
	Rest	1.05	1.72	0.85	−2.99	0.49	0.45	0.05	0.47
	pbutter 1								
	Skippy	−5.90	2.79	1.78	2.12	0.43	0.54	0.22	0.48
j	Jif	2.69	−6.65	2.11	2.12	0.36	0.43	0.45	0.46
	Peter Pan	1.67	2.04	−5.58	2.28	0.55	0.36	0.41	0.32
	Rest	1.18	1.20	1.36	−5.25	0.36	0.30	0.52	0.42
	pbutter 2								
	Skippy	−4.49	0.72	0.45	0.43	0.37	0.41	0.36	0.58
j	Jif	1.25	−4.44	0.93	0.83	0.42	0.38	0.40	0.57
	Peter Pan	1.53	1.79	−2.79	1.42	0.42	0.46	0.46	0.40
	Rest	1.08	1.18	1.05	−1.99	0.25	0.27	0.52	0.48

Asymmetric price competition

Table 4.1 shows that for all four category-market combinations the prices of the three national brands are comparable, whereas the price of the Rest brand is substantially lower. It therefore seems reasonable to assume that the perceived quality of brands summarized in the Rest group is also lower than the perceived quality of the national brands, as otherwise Rest should have an enormous market share. A finding, which is due to Blattberg and Wisniewski (1989), predicts that price promotions are more effective for higher-price higher-quality [HPHQ] brands than for lower-price lower-quality [LPLQ] brands. This means that HPHQ brands can gain more share from LPLQ brands by offering a price reduction than vice versa. Kamakura and Russell (1989) and Allenby and Rossi (1991) provide additional empirical support for this asymmetric price competition result, and Blattberg *et al.* (1995) make it an empirical generalization. This empirical generalization is refined by Sethuraman *et al.* (1999) who state that the asymmetry holds in terms of cross price elasticities but tends to disappear when absolute cross price effects are considered.

The literature offers several explanations for asymmetric price competition. Blattberg and Wisniewski (1989) argue that the asymmetry can result from a U-shaped heterogeneity distribution for the quality preferences concerning any two brands with regular prices such that the indifference point is located towards the lower quality end of this distribution. This implies that customers of LPLQ brands are more price sensitive than customers of HPHQ brands. Allenby and Rossi (1991) propose an alternative explanation that any price reduction induces a positive income effect, stimulating switching from low-quality brands to high-quality brands. They formalize this through rotating utility indifference curves. Hardie *et al.* (1993) consider reference effects and loss aversion concerning the prices and qualities of brands. They explain asymmetric price competition from the notion that losses in terms of price or quality, incurred by brand-switching, are weighted more heavily than resulting gains. Bronnenberg and Wathieu (1996) conditionally support the result of Blattberg and Wisniewski. They state that it holds if and only if the quality gap between an HPHQ brand and an LPLQ brand is sufficiently large compared to the price gap. If this condition does not hold, asymmetric price competition is reversed.

The market-share price elasticities in Table 4.4 indicate that asymmetric price competition holds for both catsup and peanut butter in the Sioux Falls market. For example, in the Sioux Falls peanut butter market, Rest loses 2.12% of its share when Skippy lowers its price by one percent, whereas this is only 1.18% in the opposite direction. However, it can also be seen from the table that the asymmetric price competition effect is sometimes reversed in the Springfield market. For catsup, a one percent price reduction of Del Monte turns out to be less effective than a one percent price reduction of the Rest brand. For peanut butter, the asymmetric price competition effect of Blattberg and Wisniewski is reversed for Skippy and Jif relative to the cheaper Rest brand. The explanation of Bronnenberg and Wathieu (1996) would be that, in the Springfield market, the perceived

difference in quality between these national brands and the Rest brand is insufficient to account for the price difference. We note from Table 4.1 that the price gap between the national brands and the Rest brand is larger in the Springfield market than in the Sioux Falls market, and that in the Springfield market the Rest brands also have larger market shares. This indicates that the Rest brands offer "good value" in the Springfield market, which is consistent with the Bronnenberg-Wathieu finding.

4.6.4 Share-switching analysis

Our model can be used to analyze share-switching patterns. Also this provides insights which cannot be obtained from market share attraction models. Table 4.5 shows the share-switching estimates, averaged over the in-sample period, for the considered markets. It reports the row-conditional share-switching fractions $\tilde{\lambda}_{l,k,t}$ and the absolute share-switching portions $\bar{\lambda}_{l,k,t} = M_{l,t-1}\tilde{\lambda}_{l,k,t}$.

A casual inspection of the share-switching fractions $\tilde{\lambda}_{l,k,t}$ in Table 4.5 indicates that Heinz has the most loyal customer base in the two catsup markets, whereas Peter Pan can be regarded as having the most loyal customer base in the two peanut butter markets. It can also be seen that, compared to the Sioux Falls market, customers in the Springfield market are more loyal to the Rest brand, representing several smaller brands and private labels, than they are loyal to the more expensive national brands. This holds for both catsup and peanut butter. Furthermore, there appears to be little share-switching between Del Monte and the Rest brand in both catsup markets.

The share-switching portions $\bar{\lambda}_{l,k,t}$ reveal some strong asymmetries. For example, in the Sioux Falls peanut butter market, on average 4.4% of the total market switches from Skippy to Peter Pan, whereas on average only 1.0% switches from Peter Pan to Skippy. In both peanut butter markets, we find that Skippy loses much more share to Peter Pan than vice versa, Peter Pan loses much more share to the Rest brand than vice versa, and the Rest brand loses much more share to Jif than vice versa. However, for all brands in all four markets, the average total share coming in (the sum of column elements) and the average total share going out (the sum of corresponding row elements) are approximately equal. This is a necessary condition for stationarity of market shares. Hence, although we find asymmetries in share-switching, our model seems to support the finding of Lal and Padmanabhan (1995) that market shares are mostly stationary, see also Dekimpe and Hanssens (1995) who consider stochastic trends.

4.7 Conclusions

In this chapter, we have proposed a model for market shares which can be used to infer share-switching across brands from store-level scanner data. Such insights cannot be obtained from market share attraction models. A second contribution of our model is that it allows for a decomposition of own and cross price elasticities into components attributable

Table 4.5: Share-switching estimates (in percents), that is, the row-conditional share-switching fractions $\tilde{\lambda}_{l,k,t}$ from brand l to brand k and the absolute share-switching portions $\bar{\lambda}_{l,k,t} = M_{l,t-1}\tilde{\lambda}_{l,k,t}$ from brand l to brand k. The reported values amount to averages over time.

		k				k			
		share-switching fractions				share-switching portions			
	catsup 1								
	Heinz	69.2	15.7	11.6	3.5	33.3	8.6	6.3	1.8
l	Hunts	27.0	56.7	11.3	4.9	7.9	13.4	3.3	1.3
	Del Monte	27.7	17.4	53.3	1.5	5.6	3.5	8.3	0.3
	Rest	51.8	8.7	0.0	39.6	3.4	0.5	0.0	2.5
	catsup 2								
	Heinz	84.9	6.3	1.9	6.9	31.8	2.9	0.9	3.1
l	Hunts	11.2	73.9	3.8	11.1	3.3	16.3	1.0	2.7
	Del Monte	18.3	9.8	71.6	0.3	1.7	0.9	5.8	0.0
	Rest	7.8	8.6	1.6	82.1	2.3	2.7	0.5	24.3
	pbutter 1								
	Skippy	67.7	11.7	13.6	7.0	18.5	3.7	4.4	2.5
l	Jif	21.1	62.6	10.2	6.2	5.9	15.4	2.9	1.7
	Peter Pan	3.5	11.9	71.9	12.7	1.0	3.7	17.5	3.8
	Rest	13.5	13.4	9.5	63.7	2.9	3.0	1.9	11.3
	pbutter 2								
	Skippy	60.5	13.1	17.5	9.0	6.0	1.3	1.9	0.9
l	Jif	7.4	59.2	22.4	11.0	1.3	9.7	4.1	1.9
	Peter Pan	2.9	6.8	74.5	15.8	1.2	2.9	26.3	6.6
	Rest	4.7	9.0	11.3	75.0	1.8	3.4	4.4	26.4

to changes in share-switching from each of the brands towards the brand whose market share is affected. This decomposition can, for example, be used to investigate which part of the reaction of the own market share to an own price reduction is caused by increased own customer retention, and which part is not.

We have illustrated our model for store-level scanner data concerning catsup and peanut butter in the Sioux Falls SD and Springfield MO markets. The elasticity decompositions indicate that the part of the reaction of own market share to an own or a competitive price reduction, attributable to changed own customer retention, is usually less than 50%.

We have also investigated whether the price elasticities obtained from our model are consistent with the finding of Blattberg and Wisniewski (1989) that higher-price higher-quality brands draw more share from lower-price lower-quality brands by offering a price reduction than vice versa. This result is supported for both catsup and peanut butter in the Sioux Falls market, but it is not supported in the Springfield market. An explanation for the disparity, consistent with the considered data, is that in the Springfield market the perceived quality gap between the expensive national brands and the cheaper Rest brand is not always sufficient to account for the price difference. This explanation would be in accordance with a result in Bronnenberg and Wathieu (1996).

Finally, we have inferred share-switching across brands. The share-switching estimates indicate that Heinz has the most loyal customer base in the two catsup markets, whereas Peter Pan can be regarded as having the most loyal customer base in the two peanut butter markets. Although the estimated share-switching portions reveal some strong asymmetries across brands, the resulting market shares seem balanced in the sense that for each brand the total share coming in and the total share going out are approximately equal.

Our share-switching model and the empirical analysis can be extended in several directions. First, although we have discussed how an outside good can be incorporated in the model to deal with changes in total category demand, we have not included it in the empirical part of this chapter. It would however be interesting to estimate the interactions between the no-purchase share and the market shares of each of the brands. The analysis would also be enriched by considering both primary and secondary demand effects.

Another interesting extension would be to relate our model to the aggregate logit approach for market shares. Aggregate logit models, in the tradition of Berry *et al.* (1995), average out individual brand choice probabilities over a heterogeneity distribution for the model parameters to obtain the market shares. This allows for more flexible model specifications. However, as closed-form expressions are usually not available, the model parameters would have to be simulated for each hypothetical household. Additionally, one would have to determine the number of drawings required to obtain reasonably stable estimation results. Nonetheless, it would be very interesting to combine state dependence and unobserved heterogeneity in an aggregate model for market shares.

Finally, we have treated the marketing-mix variables in our model as being exogenous, while in fact they might well be endogenous. For example, the price of a brand might be correlated with its shelf location. If the latter is not observed by the researcher, it can be represented by an unobserved error term in the model, implying potential correlation between the price variable and the error term. Villas-Boas and Winer (1999) demonstrate that not accounting for such endogeneity might result in seriously biased parameter estimates. Berry *et al.* (1995) deal with endogeneity of marketing-variables by explicitly modeling price setting behavior of firms, whereas Villas-Boas and Winer (1999) and Chintagunta (2000) adopt an instrumental variables approach. Chintagunta provides an explicit blueprint for the implementation of an aggregate logit model that allows for both unobserved heterogeneity and price endogeneity.

4.A Derivation of the distribution for market shares

In this appendix, we derive the distribution of the market shares $M_t = \tilde{\Lambda}_t' M_{t-1} + E_t' M_{t-1}$ after imposing the restrictions $\sum_{k=1}^{J} e_{l,k,t} = 0$, $l = 1, \ldots, J$ on the error matrix E_t. Before imposing these restrictions, the share-switching errors $e_{l,k,t}$ are assumed to be independently $N(0, \sigma_{l,k,t}^2)$ distributed. We provide a further specification of this heteroscedasticity structure at the end of the appendix.

We first derive the distribution of $(e_{l,1,t}, \ldots, e_{l,J,t})'$, given $\sum_{k=1}^{J} e_{l,k,t} = 0$, $l = 1, \ldots, J$. Without imposing the summation restriction, the density of $(e_{l,1,t}, \ldots, e_{l,J,t})'$ is given by

$$f(e_{l,1,t}, \ldots, e_{l,J,t}) \propto \exp\left(-\frac{1}{2} \sum_{k=1}^{J} \frac{e_{l,k,t}^2}{\sigma_{l,k,t}^2} \right), \tag{4.29}$$

where $\propto$ denotes "is proportional to". To impose that $\sum_{k=1}^{J} e_{l,k,t} = 0$, we make a transformation of variables from $(e_{l,1,t}, \ldots, e_{l,J-1,t}, e_{l,J,t})'$ to $(e_{l,1,t}, \ldots, e_{l,J-1,t}, z)'$ with $z = \sum_{k=1}^{J} e_{l,k,t}$. Note that $e_{l,J,t} = z - \sum_{k=1}^{J-1} e_{l,k,t}$. As the Jacobian matrix of this transformation has determinant one, the transformed density becomes

$$f(e_{l,1,t}, \ldots, e_{l,J-1,t}, z) \propto \exp\left(-\frac{1}{2} \left[\sum_{k=1}^{J-1} \frac{e_{l,k,t}^2}{\sigma_{l,k,t}^2} + \frac{(z - \sum_{k=1}^{J-1} e_{l,k,t})^2}{\sigma_{l,J,t}^2} \right] \right). \tag{4.30}$$

Now, imposing the restriction $z = \sum_{k=1}^{J} e_{l,k,t} = 0$ gives

$$\begin{aligned} f(e_{l,1,t}, \ldots, e_{l,J-1,t}, 0) &\propto \exp\left(-\frac{1}{2} \left[\sum_{k=1}^{J-1} \frac{e_{l,k,t}^2}{\sigma_{l,k,t}^2} + \frac{(\sum_{k=1}^{J-1} e_{l,k,t})^2}{\sigma_{l,J,t}^2} \right] \right) \\ &= \exp\left(-\frac{1}{2} \tilde{e}_{l,t}' \, \Gamma_{l,t} \, \tilde{e}_{l,t} \right) \end{aligned} \tag{4.31}$$

with $\tilde{e}_{l,t} = (e_{l,1,t}, \ldots, e_{l,J-1,t})'$ and $\Gamma_{l,t} = \frac{1}{\sigma_{l,J,t}^2} \, \iota_{J-1} \, \iota_{J-1}' + \mathrm{diag}(\frac{1}{\sigma_{l,1,t}^2}, \ldots, \frac{1}{\sigma_{l,J-1,t}^2})$, where the diag operator transforms a vector into a diagonal matrix with the vector elements on the diagonal. It follows from (4.31) that $(e_{l,1,t}, \ldots, e_{l,J-1,t})' \sim N(0, \Omega_{l,t})$ with

$$\begin{aligned} \Omega_{l,t} &= \Gamma_{l,t}^{-1} \\ &= -\frac{1}{\sum_{j=1}^{J} \sigma_{l,j,t}^2} \begin{pmatrix} \sigma_{l,1,t}^2 \\ \vdots \\ \sigma_{l,J-1,t}^2 \end{pmatrix} \begin{pmatrix} \sigma_{l,1,t}^2 & \cdots & \sigma_{l,J-1,t}^2 \end{pmatrix} + \mathrm{diag}\left(\sigma_{l,1,t}^2, \ldots, \sigma_{l,J-1,t}^2 \right), \end{aligned} \tag{4.32}$$

$l = 1, \ldots, J$. We note that it is easily verified that $\Omega_{l,t} \, \Gamma_{l,t} = I_{J-1}$, the $(J-1) \times (J-1)$ identity matrix, so that $\Omega_{l,t}$ is indeed the inverse of $\Gamma_{l,t}$. Next, as $M_t = \tilde{\Lambda}_t' M_{t-1} + E_t' M_{t-1}$, so that

$$\begin{pmatrix} M_{1,t} \\ \vdots \\ M_{J-1,t} \end{pmatrix} = \sum_{l=1}^{J} M_{l,t-1} \begin{pmatrix} \tilde{\lambda}_{l,1,t} \\ \vdots \\ \tilde{\lambda}_{l,J-1,t} \end{pmatrix} + \sum_{l=1}^{J} M_{l,t-1} \begin{pmatrix} e_{l,1,t} \\ \vdots \\ e_{l,J-1,t} \end{pmatrix}, \tag{4.33}$$

and $(e_{l,1,t}, \cdots, e_{l,J-1,t})'$, $l = 1 \ldots, J$, are independently distributed, it follows that

$$\tilde{M}_t = \begin{pmatrix} M_{1,t} \\ \vdots \\ M_{J-1,t} \end{pmatrix} \sim N(\,\mu_t, \tilde{V}_t\,), \tag{4.34}$$

where

$$\mu_t = \begin{pmatrix} \sum_{l=1}^{J} \tilde{\lambda}_{l,1,t} M_{l,t-1} \\ \vdots \\ \sum_{l=1}^{J} \tilde{\lambda}_{l,J-1,t} M_{l,t-1} \end{pmatrix}, \tag{4.35}$$

and

$$\tilde{V}_t = \sum_{l=1}^{J} M_{l,t-1}^2 \, \Omega_{l,t}\,. \tag{4.36}$$

The general distribution of the market shares after imposing the summation restrictions $\sum_{k=1}^{J} e_{l,k,t} = 0$, $l = 1, \ldots, J$ is given by (4.34)−(4.36). It is now straightforward to impose a specific heteroscedasticity structure $\sigma_{l,k,t}^2$ on the model by substituting it into (4.36) via (4.32). By imposing that $\sigma_{l,k,t}^2 = \sigma^2 M_{l,t-1}^{\gamma_1} M_{k,t-1}^{\gamma_2}$, we obtain

$$\begin{aligned} \tilde{V}_t &= \sigma^2 \sum_{l=1}^{J} M_{l,t-1}^{2+\gamma_1} \\ &\quad \left[-\frac{1}{\sum_{j=1}^{J} M_{j,t-1}^{\gamma_2}} \begin{pmatrix} M_{1,t-1}^{\gamma_2} \\ \vdots \\ M_{J-1,t-1}^{\gamma_2} \end{pmatrix} \left(M_{1,t-1}^{\gamma_2} \cdots M_{J-1,t-1}^{\gamma_2} \right) + \operatorname{diag}\left(M_{1,t-1}^{\gamma_2}, \ldots, M_{J-1,t-1}^{\gamma_2} \right) \right] \\ &\equiv \sigma^2 \, V_t. \end{aligned} \tag{4.37}$$

We note that by analogy with $\Omega_{l,t} = \Gamma_{l,t}^{-1}$, the inverse of V_t is given by $V_t^{-1} = \left(\sum_{l=1}^{J} M_{l,t-1}^{2+\gamma_1} \right)^{-1}$ $\left[M_{J,t-1}^{-\gamma_2} \, \iota_{J-1} \, \iota_{J-1}' + \operatorname{diag}(M_{1,t-1}^{-\gamma_2}, \ldots, M_{J-1,t-1}^{-\gamma_2}) \right]$.

Part II

Monte Carlo integration methods

Chapter 5

A radial-based Monte Carlo methodology

5.1 Introduction

In the first part of this thesis, we have discussed three new models which can be used to analyze purchase behavior of households and to analyze market shares of brands. In the empirical applications, we have obtained parameter estimates such that the observed data are "most likely" and we have computed the corresponding (asymptotic) standard errors. Hence, we have treated the model parameters as unknown constants which have to be recovered, and the resulting parameter estimates are functions of the data. We have actually adopted the classical approach of making inference.

In the classical framework, inferential procedures such as testing for parameter significance are based on assuming a "true" data generating process (by imposing parameter restrictions) from which an infinitely large hypothetical data set can be obtained. A test statistic is constructed from the *observed* data, and its realization is compared to the corresponding distribution resulting from the *hypothetical* data set. If the realized value is "unlikely" for the assumed "true" data generating process, the underlying hypothesis is rejected. In the classical approach, the concept of "probability" is *objective* in the sense that "probability" is defined as the fraction of occurrences when a process is repeated infinitely often.

A second approach to inference which we have not considered so far is the Bayesian approach. In this Bayesian framework, the parameters of a model are considered to be random variables themselves, so that the concept of "true parameter values" on which any classical analysis is conditioned does not have a Bayesian counterpart. Instead, a Bayesian believes that the model parameters can take on a wide range of values. The beliefs that are present on the parameter outcomes are summarized in a so-called prior density. This density should be constructed before observing the data. An important implication of the probability concept above is that probabilities are *subjective*, as they

are states of the mind which may differ from one person to another. We refer to Poirier (1995) for an extensive comparison of the classical and the Bayesian approach to inference.

The prior density only reflects prior expectations about the parameters. When actual outcomes of the data generating process become available, these prior beliefs can be updated by incorporating the new information from the data. This way of "learning" is formalized by Bayes' theorem. Let $y = (y_1, \ldots, y_N)$ summarize the available data and let θ denote the vector of model parameters. As both are stochastic, the joint density of the data y and the parameters θ can be written as the product of a marginal density and a conditional density in two ways, that is,

$$p(y, \theta) = p(y)\, p(\theta|y) = p(\theta)\, p(y|\theta). \tag{5.1}$$

Rearranging this identity results in Bayes' theorem, that is,

$$p(\theta|y) = \frac{p(\theta)\, p(y|\theta)}{p(y)}\,. \tag{5.2}$$

However, as the denominator in (5.2) does not depend on θ and only serves as a scaling constant, it is convenient to represent Bayes' theorem by

$$p(\theta|y) \propto p(\theta)\, p(y|\theta), \tag{5.3}$$

where $\propto$ denotes "is proportional to". Bayes' theorem (5.3) relates the prior density $p(\theta)$, containing only prior beliefs about the parameters θ, to the posterior density $p(\theta|y)$ which also accounts for the observed data y. The link between the prior density and the posterior density is the likelihood function $p(y|\theta)$. A consequence of (5.3) is that learning from the data is entirely done through this likelihood function.

The posterior density $p(\theta|y)$ contains all what is known about the parameters θ after observing the data y. However, the format of the information in $p(\theta|y)$ is not very convenient. For example, if θ is ten-dimensional, $p(\theta|y)$ is a function of ten variables of which the shape cannot be plotted. For sensible interpretation, one is usually interested in the posterior means, variances and correlations. Furthermore, one might be interested in the posterior probability that θ belongs to a region D of the parameter space, for example, to obtain (smoothed) histograms of the marginal densities, or for testing purposes. In all cases, interest is focused on the integral

$$E[g(\theta)|y] = \int g(\theta)\, p(\theta|y)\, d\theta, \tag{5.4}$$

where g is the relevant function. We note that the probability that $\theta \in D$ results from (5.4) by setting $g(\theta) = I\{\theta \in D\}$, where $I\{\cdot\}$ is the 0/1 indicator function. If the posterior density is only known up to a scaling constant, so that $p(\theta|y)$ does not integrate to unity, (5.4) should be replaced by

$$E[g(\theta)|y] = \frac{\int g(\theta)\, p(\theta|y)\, d\theta}{\int p(\theta|y)\, d\theta}\,. \tag{5.5}$$

In this chapter, we discuss several methods to compute (5.4) and (5.5), and we also put forward a new methodology which extends current methods. In the overview, we consider both deterministic integration and Monte Carlo integration. The former approach means that the involved integrand is evaluated at many fixed points, after which the integral is approximated by a weighted sum of these function evaluations. The latter approach is based on the idea that $E[g(\theta)|y]$ can be approximated by the sample mean $\frac{1}{n}\sum_{i=1}^{n} g(\theta_i)$, where $\theta_1, \ldots, \theta_n$ are drawings obtained from the posterior distribution. The difference between Monte Carlo integration and deterministic integration is nicely characterized by the equivalency relationship (5.4). Monte Carlo algorithms focus on the left-hand posterior expectation, whereas deterministic methods aim to explicitly evaluate the right-hand integral.

The outline of this chapter is as follows. In Section 5.2, we provide an overview of some relevant and frequently used integration methods. Most of the discussed algorithms serve as building blocks in the development of our proposed Monte Carlo approach. We discuss this Monte Carlo methodology, referred to as *adaptive radial-based direction sampling*, in Section 5.3. Section 5.4 contains some simple applications to illustrate the performance of the algorithms. Finally, we make some concluding remarks in Section 5.5.

5.2 Review of some integration methods

In this section, we review some deterministic and Monte Carlo integration methods. The focus is on the algorithms themselves, and we do not pay much attention to their empirical application. Bauwens *et al.* (1999) provide a Bayesian treatment of time series models. Koop (2003) explains how various Monte Carlo methods can be applied to make inference in a broad class of econometric models including linear regression models with panel data and qualitative and limited dependent variable models.

5.2.1 Deterministic integration

For ease of exposition, we focus on computing the integral (5.4), which is assumed to be one-dimensional with bounds $\underline{\theta}$ and $\overline{\theta}$. Deterministic integration rules evaluate the integrand $f(\theta) \equiv g(\theta)\, p(\theta|y)$ at many fixed points, and use an appropriately weighted sum of the evaluations as an approximation to the value of the integral. Hence, deterministic integration methods are characterized by the relationship

$$\int f(\theta)\, d\theta \approx \sum_{i=1}^{n} w_i f(\theta_i), \tag{5.6}$$

where $\theta_1, \ldots, \theta_n$ are the evaluation points and $w_1, \ldots, w_n$ are the corresponding weights. We note that most deterministic integration rules can be extended to multi-dimensional integrals in a straightforward way. However, as the number of required function evaluations increases exponentially with the dimension of the integration problem, such inte-

gration rules become soon infeasible. As a result, they are rarely used when the number of variables exceeds three or four. We refer to Stoer and Bulirsch (1993) and Cheney and Kincaid (1994) for extensive textbook discussions on deterministic integration methods.

The trapezoid integration rule approximates the integral (5.4) by replacing the integrand $f(\theta)$ by a linear interpolating function. This results in the approximation

$$\int_{\underline{\theta}}^{\overline{\theta}} f(\theta)\, d\theta \approx \frac{\overline{\theta} - \underline{\theta}}{2} f(\underline{\theta}) + \frac{\overline{\theta} - \underline{\theta}}{2} f(\overline{\theta}). \tag{5.7}$$

Similarly, Simpson's rule replaces the integrand $f(\theta)$ by a quadratic interpolating function with equidistant interpolation points. This gives rise to the approximation

$$\int_{\underline{\theta}}^{\overline{\theta}} f(\theta)\, d\theta \approx \frac{\overline{\theta} - \underline{\theta}}{6} f(\underline{\theta}) + \frac{4(\overline{\theta} - \underline{\theta})}{6} f\left(\frac{\underline{\theta} + \overline{\theta}}{2}\right) + \frac{\overline{\theta} - \underline{\theta}}{6} f(\overline{\theta}). \tag{5.8}$$

Alternatively, Gaussian integration rules which do not need evaluation of the integrand at the bounds $\underline{\theta}$ and $\overline{\theta}$ can be used, see Stoer and Bulirsch (1993) and Cheney and Kincaid (1994). However, finite integration bounds, as required for the trapezoid rule and Simpson's rule, can always be obtained by an appropriate transformation of variables, see, for example, Bauwens *et al.* (1999, p. 70). To decrease the approximation error, deterministic integration rules are usually applied to many small subintervals of $[\underline{\theta}, \overline{\theta}]$.

An important issue that remains is how many evaluation points are needed to achieve a predetermined level of accuracy. Furthermore, one wants to obtain this numerical accuracy using a minimal number of function evaluations, as evaluations are costly in terms of computing time. For the trapezoid rule and Simpson's rule, an iterative procedure can be applied such that evaluations points are added in subintervals of $[\underline{\theta}, \overline{\theta}]$ as long as this results in a substantial improvement. Hence, evaluation points are only added in those regions where they are needed. We first consider the entire interval $[\underline{\theta}, \overline{\theta}]$. Let A_1 be the integral approximation resulting from either the trapezoid rule or Simpson's rule without splitting up $[\underline{\theta}, \overline{\theta}]$ into smaller subintervals. Moreover, let A_2 be the approximation when $[\underline{\theta}, \overline{\theta}]$ is split up into two equally large parts for which the subintegrals are evaluated separately. Cheney and Kincaid (1994) explain why $|A_1 - A_2| < 15\,\varepsilon$ can be used as a criterion to determine whether intervals have to be split up further to achieve an accuracy level of ε for the Simpson's rule. Similarly, $|A_1 - A_2| < 3\,\varepsilon$ can be used as a criterion for the trapezoid rule. If the stopping condition is met, the integral is approximated by A_2. In the alternative case, at least one further partition has to be made. Subintervals are split up until they all satisfy their corresponding stopping criteria, where the tolerance level for each subinterval is taken proportionally to the length of that subinterval, so that the total accuracy level remains ε.

5.2.2 Independence sampling: direct methods

In this subsection, we turn to Monte Carlo integration. Basically, a Monte Carlo method consists of two steps. First, a sample $\theta_1, \ldots, \theta_n$ is collected from the posterior distribution

with density (or density kernel) $p(\theta|y)$. Subsequently, this sample is used to estimate the expectation $E[g(\theta)|y]$ by the corresponding sample mean $\frac{1}{n}\sum_{i=1}^{n} g(\theta_i)$. For example, the probability $\Pr(\theta \in D|y) = E[\, I\{\theta \in D\}\,|y]$ is estimated by the number of drawing falling into the region D, that is, $\sum_{i=1}^{n} I\{\theta_i \in D\}$, divided by the sample size n.

In the ideal case, the sample $\theta_1, \ldots, \theta_n$ can be obtained directly from the posterior. However, even when the posterior distribution is non-standard, so that it is not straightforward to collect a sample from this posterior, direct sampling methods are useful. They can serve as building blocks for more involved algorithms. For example, any sampling algorithm is based on collecting drawings from the uniform $U(0,1)$ distribution, so that suitable methods to generate these "random numbers" are of utmost importance.

Uniform sampling

The most commonly used method to sample from the uniform distribution is the linear congruential random number generator [LCRNG]. This generator creates a sequence of "random numbers" $u_1, \ldots, u_n$ using the recursion

$$u_i = (a\, u_{i-1} + b) \bmod M, \qquad i = 1, \ldots n, \tag{5.9}$$

where mod M gives the remainder after division by M. The multiplier a and the modulus M are strictly positive integers, while the increment b is also allowed to be zero. The initial value u_0 of the sequence is called the seed. To map u_1, ...,u_n to the unit interval, these values are divided by M. We note that the recursion (5.9) is completely deterministic, so that the generated "random numbers" are actually not random at all. For properly chosen a, b and M, it only seems *as if* they are random.

An immediate consequence of the first-order recursion (5.9) is that the process repeats itself as soon as a realized value occurs for the second time. As there are only M possible values to encounter, the process must renew itself in at most M steps, that is, the maximum period is M. However, if a, b and M are not chosen with care, the actual period might be much shorter. Furthermore, even if the process has full period M, the random number generator is not necessarily a good one.

In practice, multiplicative LCRNGs are frequently considered. These arise from (5.9) by setting $b = 0$, so that the increment is turned off. Two popular multiplicative LCRNGs are the Lewis-Goodman-Miller generator, obtained by setting $a = 16{,}807$ and $M = 2^{31}-1$, and the Payne-Rabung-Bogyo generator, obtained by setting $a = 630{,}360{,}016$ and $M = 2^{31} - 1$. We refer to Law and Kelton (1991) for details.

Inversion method

The inversion method is another direct sampling method. It directly translates uniform $U(0,1)$ drawings into drawings from the (univariate) distribution of interest. The underlying idea is simple. If the random variable X follows a distribution with cumulative

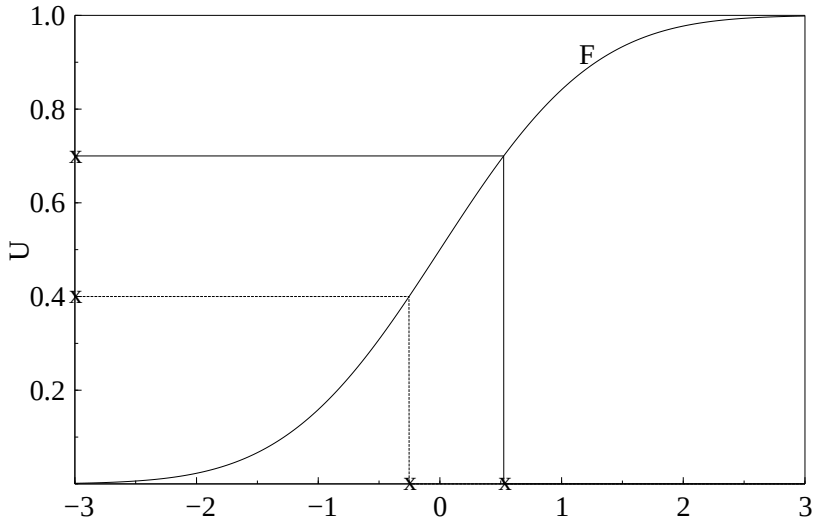

Figure 5.1: Illustration of the inversion method for the standard normal distribution.

distribution function [CDF] denoted by F, then the corresponding CDF value $U = F(X)$ is uniformly distributed, as

$$\Pr(U \leq u) = \Pr(F(X) \leq u) = \Pr(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u \tag{5.10}$$

with F^{-1} denoting the inverse CDF. Based on this result, the inversion method consists of two steps. First, a uniform sample $u_1, \ldots, u_n$ is collected. Next, this sample is transformed into realizations $x_1 = F^{-1}(u_1), \ldots, x_n = F^{-1}(u_n)$ from the distribution of interest. Figure 5.1 illustrates the inversion method for the standard normal distribution. Clearly, as the standard normal CDF is steepest around 0, that region is "hit" most frequently, so that most drawings have values relatively close to 0. On the other hand, not many drawings fall into regions far away from 0, as these regions are difficult to "hit". This mechanism induces that drawings are assigned to regions in accordance with their probability mass. We note that the inversion method is particularly suitable to sample from (univariate) truncated distributions. For example, if a distribution is truncated to the left of some value a and to the right of some value b, then all drawings should fall into the region (a,b). This is easily achieved by sampling $u_1 \ldots, u_n$ uniformly from the interval $(F(a), F(b))$, instead of sampling them from the interval $(0, 1)$. All that has to be done is redefining

$$u_i \equiv F(a) + [\, F(b) - F(a) \,]\, u_i, \qquad i = 1, \ldots n. \tag{5.11}$$

Although it is desirable that the inverse CDF F^{-1} has a closed form expression, this is not required. It is not even necessary that the CDF itself has a closed form expression. However, in such situations one has to resort to a numerical approximation. For example, an approximate CDF can be constructed by evaluating the probability density function (or a density kernel) at many points to build a grid, and using linear interpolation. As the resulting approximation is piecewise linear, inversion is straightforward.

5.2.3 Independence sampling: indirect methods

If it is impossible to draw directly from the distribution of interest, hereafter referred to as the target distribution, indirect methods may be considered. Such methods aim to collect a representative sample for the target distribution by considering an alternative "candidate" distribution. This candidate distribution should be easy to sample from, and for a good performance it should provide a reasonably accurate approximation to the original target distribution. Indirect sampling methods involve a correction mechanism to account for the difference between the target density and the candidate density. In this subsection, we discuss two indirect sampling approaches resulting in independent drawings, so that the Law of Large Numbers [LLN] and the Central Limit Theorem [CLT] still apply.

Rejection sampling

The first indirect method we discuss is rejection sampling. By adopting this approach, one collects a sample from the candidate distribution, and decides for each drawing whether it is accepted or rejected. If a drawing is accepted, it is included in the sample representing the target distribution. Rejection means that the drawing is thrown away. Note that the rejection step is the correction mechanism which is employed in rejection sampling.

To apply the rejection method to a target density p, one first needs to specify an appropriate candidate density q. For example, one might consider a normal or Student-t density. Next, a constant $c > 0$ has to be found such that

$$p(x) \leq c\, q(x) \tag{5.12}$$

for all x, so that the graph of the kernel $c\,q$ of the candidate density is entirely located above the graph of the target density p. We note that (5.12) implies that p is allowed to be a kernel of the target density, as the constant c can always adjust to p. However, the candidate density q should be such that the ratio $\frac{p(x)}{q(x)}$ is bounded for all x, so that c is finite. Basically, the rejection method consists of uniformly drawing points below the graph of $c\,q$, and accepting the points falling below the graph of p. The remaining points are rejected. This idea is illustrated by Figure 5.2 for a bimodal target density. The following rejection algorithm collects a sample of size n from the target distribution with density p, that is,

Initialize the algorithm:

The set of accepted drawings S is empty: $S = \emptyset$.

The number of accepted drawings i is zero: $i = 0$.

Do while $i < n$:

Obtain y from candidate distribution with density q.

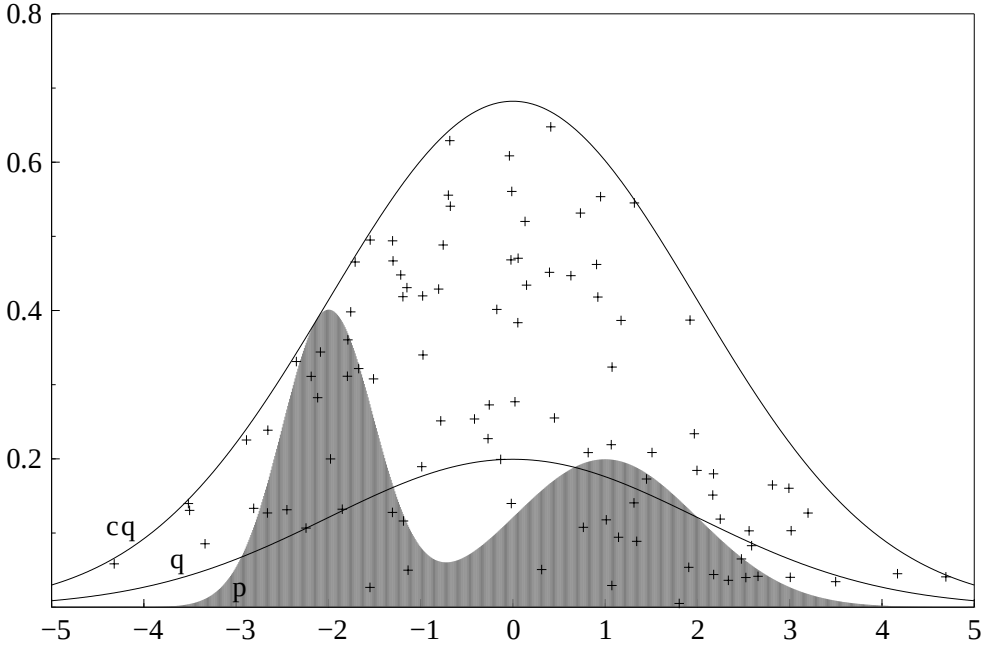

Figure 5.2: Illustration of rejection sampling.

Obtain u from uniform distribution $U(0,1)$.

If $u < \frac{p(y)}{c\,q(y)}$ then accept y:

Add y to the set of accepted drawings: $S = S \cup \{y\}$.

Update the number of accepted drawings: $i = i + 1$.

Return S.

To minimize the fraction of rejected drawings, one should take the constant c as small as possible, that is,

$$c = \max_x \frac{p(x)}{q(x)}\,. \tag{5.13}$$

Clearly, the optimal c is small if variation in the ratio $\frac{p(x)}{q(x)}$ is small. This explains that a candidate density providing a good approximation to the target density is desirable. We note that a natural application of rejection sampling arises when truncated distributions are considered. The candidate density is the density without truncation. Candidate drawings are accepted if they are feasible, and are rejected otherwise.

Importance sampling

Importance sampling is another indirect approach to obtain an estimate for $E[g(X)]$, where X is a random variable from the target distribution. It dates back to Hammersley and Handscomb (1964), and it was introduced in econometrics by Kloek and Van Dijk (1978). The method is related to rejection sampling. The rejection method either accepts

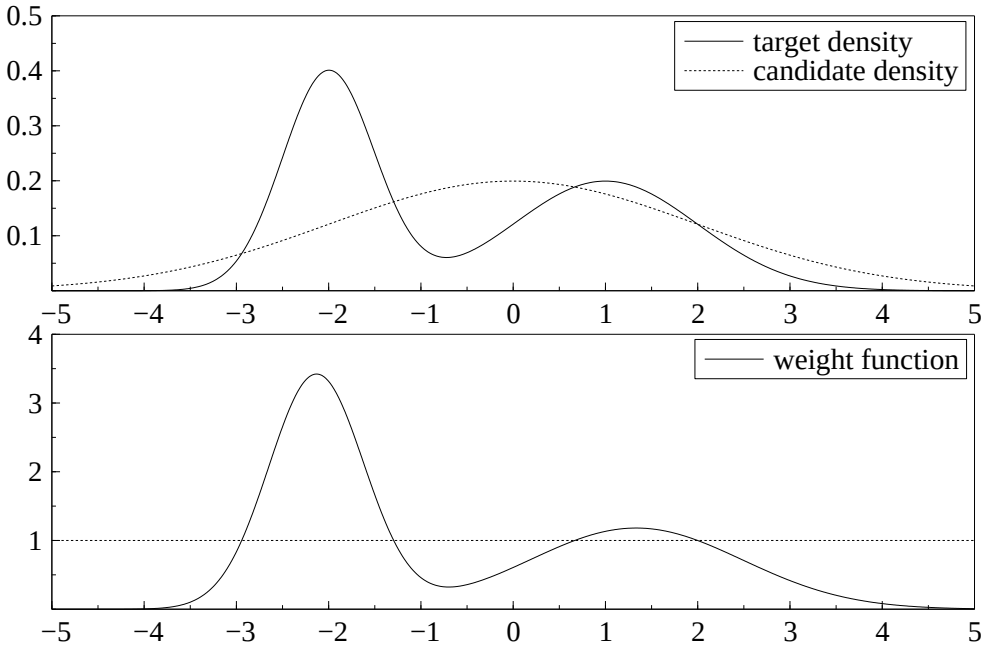

Figure 5.3: Illustration of importance sampling. The weight function reflects the importance of drawings from the candidate density.

or rejects candidate drawings, that is, drawings either receive full weight or they do not get any weight at all. Importance sampling is based on this notion of assigning weights to drawings. However, in contrast to the rejection method, these weights are not based on an all-or-nothing situation. Instead, they can take on any possible value, representing the relative importance of drawings. If q is the candidate density, or importance function, and p is a kernel of the target density, then importance sampling is based on the relationship

$$E[g(X)] = \frac{\int g(x)\,p(x)\,dx}{\int p(x)\,dx} = \frac{\int g(x)\,w(x)\,q(x)\,dx}{\int w(x)\,q(x)\,dx} = \frac{E[w(Y)\,g(Y)]}{E[w(Y)]}\,, \tag{5.14}$$

where Y is a random variable from the candidate distribution, and $w(y) = \frac{p(y)}{q(y)}$ is the weight function which should be bounded. It follows from (5.14) that a consistent estimate of $E[g(X)]$ is given by the weighted mean

$$\hat{g}_w = \frac{\sum_{i=1}^{n} w(y_i)\,g(y_i)}{\sum_{j=1}^{n} w(y_j)}\,, \tag{5.15}$$

where $y_1, \ldots, y_n$ are realizations from the candidate distribution and $w(y_1), \ldots, w(y_n)$ are the corresponding weights. Figure 5.3 provides a graphical illustration of the method. Points for which the graph of the target density is located above the graph of the candidate density are not sampled often enough. To correct for this, such drawings are assigned relatively large weights (weights larger than one). The reverse holds in the opposite case. We note that although importance sampling can be used to estimate characteristics of

the target density (such as the mean), it does not provide a sample according to this density. Drawings are generated from the *candidate* distribution. Hence, in a strict sense, importance sampling is not a sampling method, but a pure integration method. To obtain a sample representation of the target distribution, one may draw $x_1, \ldots, x_{\underline{n}}$ with replacement from the sample $y_1, \ldots, y_n$, $\underline{n} \ll n$, with probabilities proportional to the importance weights $w_1, \ldots, w_n$. This is the sampling/importance resampling approach of Rubin (1987).

The performance of the importance sampling algorithm is strongly affected by the choice of the candidate distribution. If the importance function q is inappropriate, the weight function $w(x) = \frac{p(x)}{q(x)}$ would vary a lot and it might happen that only a few drawings with extreme weights almost completely determine the estimate $\hat{g}_w$. This estimate would be unstable. In particular, a situation such that the tails of the target density are fatter than the tails of the candidate density is concerning, as this would imply that the weight function might even tend to infinity. In such a case, $E[g(X)]$ would not even exist, see (5.14). It is for this reason that a fat-tailed Student-t importance function is usually preferred to a normal candidate density, see Kloek and Van Dijk (1978) and Van Dijk and Kloek (1980), among others. The regularity conditions and numerical accuracy of importance sampling are discussed by Geweke (1989).

5.2.4 Dependence sampling

Another approach to sample from non-standard distributions is the Markov Chain Monte Carlo [MCMC] approach. An MCMC method aims to collect a sample representative for the target distribution by construction of a Markov chain converging to that distribution. After a sufficiently long burn-in period, so that the influence of the initialization conditions has become negligible, drawings from the Markov chain are regarded as drawings from the target distribution itself. However, as Markov chain sampling naturally induces correlation, the resulting drawings are not independent, so that the Law of Large Numbers [LLN] and the Central Limit Theorem [CLT] no longer apply. We refer to Ross (1997) for an introductory textbook discussion on Markov chain theory.

Metropolis-Hastings algorithm

The Metropolis-Hastings [MH] algorithm, introduced by Metropolis *et al.* (1953) and generalized by Hastings (1970), is an MCMC approach which can be used to draw from the posterior distribution for a wide variety of models. It has similarities with rejection sampling, as a rejection mechanism is involved. However, rejected drawings are dealt with in a different way. An excellent exposition on the MH algorithm is Chib and Greenberg (1995) in which theory and intuition as well as application of the algorithm are discussed. An important survey on the broader class of MCMC methods is Tierney (1994).

The MH algorithm constructs a Markov chain mimicking the target density p after a burn-in period. This is done as follows. Given the current state x, a candidate state y is

drawn from a transition density $q(x,\cdot) \equiv q(\cdot|x)$ from which sampling is straightforward. The state y is accepted with probability $\alpha(x,y)$, depending on both the current state x and the candidate state y. Acceptance implies that the chain moves to y. Rejection means that the move is not made, that is, the next state is again x. This procedure defines one iteration of the MH algorithm, and it is repeated many times. The appropriate acceptance probability turns out to be defined by

$$\alpha(x,y) = \min\left\{ \frac{p(y)\,q(y,x)}{p(x)\,q(x,y)}\,,1 \right\}, \tag{5.16}$$

see Chib and Greenberg (1995) for an explanation. In sum, the MH algorithm constructs a Markov chain of length n as follows, that is,

Initialize the algorithm:

Choose a feasible initial state x_0.

Do for $i = 1,\ldots,n$:

Obtain y from candidate transition density $q(x_{i-1},\cdot)$.

Obtain u from uniform distribution $U(0,1)$.

Compute transition probability $\alpha(x_{i-1},y)$, defined by (5.16).

If $u < \alpha(x_{i-1},y)$ then accept transition:

$x_i = y$.

Else reject transition:

$x_i = x_{i-1}$.

Return $x_1,\ldots,x_n$.

We note that the rejection step, determining whether the chain moves or stays, is the correction mechanism employed in the MH approach. For example, if a state x would not be visited often enough, there is a large probability that the chain settles down for a while whenever such a state occurs, as subsequent candidate states y are assigned low acceptance probabilities $\alpha(x,y)$. The opposite holds if a state x would occur too frequently.

Several approaches can be adopted to specify the candidate transition density q. Two popular specifications are such that the resulting Markov chain is either an independence chain or a random walk chain. An independence chain has the property that the candidate state y is drawn independently of the current state x, that is,

$$q(x,y) = q(y). \tag{5.17}$$

Typical choices for q are normal or Student-t densities. It follows from (5.16) and (5.17) that the acceptance probability in an independence chain is given by

$$\alpha(x,y) = \min\left\{ \frac{p(y)\,q(x)}{p(x)\,q(y)}\,,1 \right\} = \min\left\{ \frac{w(y)}{w(x)}\,,1 \right\}, \tag{5.18}$$

that is, the minimum of a ratio of importance weights and one. The interpretation of (5.18) is that a transition from x to y resulting in a larger importance weight is always made, whereas a transition resulting in a smaller importance weight is not always performed. We note that (5.18) establishes a link with importance sampling. The second specification of q that we discuss corresponds to a random walk chain. For such a chain, the transition step $y - x$ is drawn instead of the state y itself, that is,

$$q(x, y) = q(y - x). \tag{5.19}$$

Typical choices for $q(y-x)$ are normal or Student-t densities centered around 0, so that the expectation of the next state $y = x + (y - x)$ is the current state x. As a third possibility, we mention that if the transition density is symmetric, that is, $q(x, y) = q(y, x)$ for all x, y, the acceptance probability $\alpha(x, y)$ reduces to

$$\alpha(x, y) = \min\left\{ \frac{p(y)}{p(x)}, 1 \right\}, \tag{5.20}$$

as in the original Metropolis algorithm, see Metropolis *et al.* (1953). The acceptance probability (5.20) has a similar interpretation as (5.18). A transition from x to y implying an increase in the target density is always made, whereas a transition implying a decrease is not always performed. As for any indirect sampling method, the choice of the candidate density q strongly affects the quality of the resulting sample. Other practical implementation issues concern the length of the burn-in period for "convergence" of the Markov chain, and the number of effective drawings needed to obtain accurate estimates. Cowles and Carlin (1996) and Brooks and Roberts (1998) provide summaries of convergence diagnostics for MCMC methods.

Gibbs sampling

The MH algorithm is a general MCMC approach. A more problem-specific method within the MCMC class is the Gibbs sampling algorithm of Geman and Geman (1984). The Gibbs sampler is based on decomposing the multi-dimensional random variable X into k components $X^1, \ldots, X^k$, which are not necessarily univariate. It constructs a Markov chain, converging to the target distribution, by iteratively drawing the k components of X conditional on the values of all other components. For many seemingly intractable target densities, it is possible to derive a set of conditional densities for which sampling is straightforward. The Gibbs sampler exploits this notion, as it precisely considers these conditional densities. Its usefulness is, for example, demonstrated by Gelfand *et al.* (1990), Gelfand and Smith (1990) and Smith and Roberts (1993). Casella and George (1992) provide a tutorial on Gibbs sampling using simple examples to explain how and why the method works.

For implementation of the Gibbs sampler, it is needed that all full conditional distributions can be sampled from. These conditional distributions are described by the

densities $p(x^j|x^{-j})$, $j = 1, \ldots, k$, where $x^{-j} = (x^1, \ldots, x^{j-1}, x^{j+1}, \ldots x^k)$ denotes the set of $k-1$ components excluding the j-th component. The Gibbs sampling algorithm collects n drawings $x_i = (x_i^1, \ldots, x_i^k)$, $i = 1, \ldots, n$, as follows. The components x_i^j, $i = 1, \ldots, n$, $j = 1, \ldots, k$, are augmented into a single sequence $x_1^1, \ldots, x_1^k, x_2^1, \ldots, x_2^k, \ldots, x_n^1, \ldots, x_n^k$, and the elements of this Gibbs sequence are generated such that

$$x_i^j \text{ is obtained from } p(x^j|x_{i-1}^{-j}), \qquad i = 1, \ldots, n, \quad j = 1, \ldots, k,$$

where $x_{i-1}^{-j} = (x_i^1, \ldots, x_i^{j-1}, x_{i-1}^{j+1}, \ldots x_{i-1}^k)$ denotes all components except x^j at their *most recent* values. The complete algorithm is as follows, that is,

Initialize the algorithm:

Choose a feasible initial state $x_0 = (x_0^1, \ldots, x_0^k)$.

Do for drawing $i = 1, \ldots, n$:

Do for component $j = 1, \ldots, k$:

Obtain x_i^j from conditional target density $p(x^j|x_{i-1}^{-j})$.

Return $x_1, \ldots, x_n$.

We note that the Gibbs sampler may be regarded as a special case of the MH algorithm in which rejections do not occur, see, for example, Chib and Greenberg (1995) and Gelman *et al.* (1995, p. 328).

Figure 5.4 illustrates how the Gibbs sampler works for two 2-dimensional target distributions featuring correlation and bimodality. Clearly, as each time one of the two components (either x^1 or x^2) is fixed while the other component is sampled from its conditional distribution, the sampling paths move in orthogonal directions parallel to the coordinate axes. The horizontal position is updated given the current vertical position, and the vertical position is updated given the current horizontal position. The graph shows sampling paths after 10 iterations and after 1000 iterations, and it indicates that the orthogonal movement may cause the Gibbs sampler to break down. First, the two left-hand graphs demonstrate that high correlation results in a slowly moving sampling path, so that the Gibbs sampler might be stuck in a local region for quite a long time. This problem increases when the correlation between the two components becomes higher. Second, the two right-hand graphs demonstrate that if the target density has two modes located far away from each other, "mode hopping" seldom occurs. This essentially induces the same problem as high correlation, that is, the Gibbs sampler might again be stuck in a local region for a long time. As a result, an enormous number of drawings would be needed to obtain a representative coverage of the target density. A reparameterization of the sampling problem may be an effective way to deal with such high correlations, see Gilks and Roberts (1996).

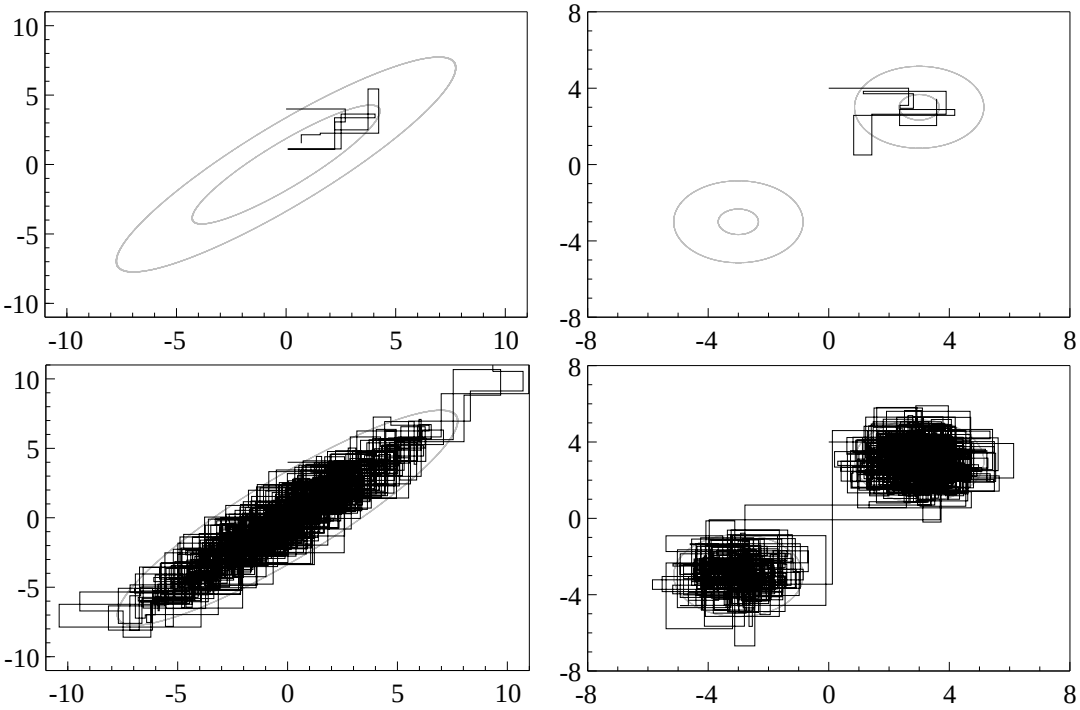

Figure 5.4: Illustration of the Gibbs sampler for a correlated target density (left) and a bimodal target density (right). The generated sample paths are shown for 10 iterations (above) and 1000 iterations (below).

For application of the Gibbs sampler, it is desirable but not necessary that all k conditional distributions can be *directly* sampled from. If a "difficult" conditional density is one-dimensional, an approximate cumulative distribution function [CDF] can be constructed by building a density grid and using linear interpolation. Subsequently, the inversion method can be applied to the piecewise linear approximation. This griddy Gibbs sampling approach is proposed by Ritter and Tanner (1992). Alternatively, an MH acceptance-rejection step might be incorporated in the Gibbs algorithm to sample from the (not necessarily univariate) "difficult" conditional density. Although this latter method is actually a special case of the MH algorithm, it is usually called the MH-within-Gibbs approach.

Data augmentation

For many models containing latent variables (such as probit choice models with unobserved random utilities), the parameters θ have a non-standard posterior distribution. Moreover, evaluation of the likelihood function, and hence evaluation of the posterior density, might be complicated and computationally intensive. However, standard distributions would arise if the latent data, denoted by z, would be known. For example, the binary probit model (which can be used to predict purchase incidence) would reduce to

a linear regression model with known variance. Hence, "observing" z would greatly facilitate the sampling procedure. The data augmentation algorithm of Tanner and Wong (1987) is a useful extension of the Gibbs sampler which is based on this notion. It extends the sampling space, as both the parameters θ and the latent data z are sampled. In the algorithm, z is drawn conditional on θ, and θ is drawn conditional on z. So, the latent data are imputed using the current parameter values, and subsequently the parameters are sampled *as if* the latent data are observed. By repeating this procedure many times, a sampling sequence is constructed involving both θ and z. Disregarding z, this process results in a Markov chain for the parameters θ converging to the posterior distribution. Data augmentation for the conditional probit model is discussed by Albert and Chib (1993), McCulloch and Rossi (1994) and McCulloch *et al.* (2000). Wei and Tanner (1990) and Chib (1992) consider data augmentation for the censored regression model.

Gibbs sampling with auxiliary variables

Gibbs sampling with auxiliary variables is a sampling approach developed by Damien *et al.* (1999). Similar to data augmentation, latent variables are incorporated in the sampling process to facilitate drawing from the set of full conditional distributions. However, contrary to data augmentation, the latent variables are not "missing data" from the model. Instead, they are introduced in an artificial way. The method of Damien *et al.* may be interpreted as a reversion of the independence chain MH algorithm. We recall that the MH algorithm first draws a candidate state y, given the current state x_{i-1}, and subsequently considers a uniform drawing $u \in (0, 1)$ to determine whether the candidate state is accepted. The sampling approach of Damien *et al.* turns this around, that is, first an auxiliary drawing u from the uniform distribution is obtained and subsequently the state y is sampled inside the acceptance region determined by u. The gain of this reversion is that the state y is accepted by definition. However, the price to pay is that sampling inside the acceptance region amounts to drawing from a truncated distribution. We refer to Damien *et al.* (1999) for details.

5.3 Adaptive radial-based direction sampling

In this section, we introduce the class of adaptive radial-based direction sampling methods.

5.3.1 Introduction and motivation

In the overview, we have discussed several methods to draw from the posterior, or target, density. For rejection sampling, importance sampling, and the MH algorithm, we have noted that it is important to find a reasonably accurate candidate density. An inaccurate candidate distribution may give rise to many rejected drawings, or many drawings with negligible weights, implying that an enormous number of drawings is needed to obtain reliable estimates. Even worse, complete regions of the target density with substantial

density mass may be missed, so that the resulting sample cannot be representative at all. A partial solution to find a suitable candidate density would be to update it sequentially. This can be done by iteratively collecting a sample via the current candidate density and using the information contained in the sample to improve the approximation to the target density. Updating can either be done for a fixed number of rounds, or until a convergence criterion has been met. Kloek and Van Dijk (1978), Van Dijk and Kloek (1980) and Oh and Berger (1992) consider parametric updating of the candidate density, whereas Givens and Raftery (1996) apply kernel density estimation.

We have also discussed the Gibbs sampling algorithm as a powerful tool to sample from the target distribution. In particular, Gibbs sampling with data augmentation is a natural and efficient way to draw the parameters of models containing latent variables. For such models, we do not expect our proposed Monte Carlo methodology to outperform the Gibbs sampler. However, there are some drawbacks too. First, for convenient implementation of the Gibbs algorithm, it is necessary that sampling is straightforward for most of the conditional densities. This is not always the case, and obtaining an appropriate sampling scheme might require a large analytical effort. For "difficult" conditional densities, one may resort to construction of an approximate grid, or alternatively, one may incorporate an MH acceptance-rejection step. A second complication is that movement of the Gibbs sampler may be very slow if the target density features strong correlations or multi-modality, as movement is orthogonal and parallel to the coordinate axes. We have illustrated this problem in Subsection 5.2.4. In principle, an appropriate reparameterization, or orthogonalization, of the sampling problem should reduce the burden, see Gilks and Roberts (1996). However, no guarantees can be given that this reparameterization approach should work well in practice. This has led Gilks *et al.* (1994) to develop a new class of MCMC methods, called "adaptive direction sampling" [ADS], in which the sampling directions are not restricted anymore. For example, the hit-and-run algorithm, developed by Belisle *et al.* (1993) and generalized by Chen and Schmeiser (1996), and the snooker algorithm of Gilks *et al.* (1994) are contained in the ADS class. Convergence of ADS is discussed in Roberts and Gilks (1994). However, the practical performance of the ADS algorithms turns out to be somewhat disappointing.

We propose the class of adaptive radial-based direction sampling [ARDS] methods to sample from target distributions which are possibly multi-modal, skew, and feature strong correlation. For such complicated densities, it might be difficult to come up with a suitable candidate density, and also the Gibbs sampler may break down. The ARDS algorithms are based on a composite transformation of the m-dimensional parameter space. This transformation contains (i) an orthogonalizing step, and (ii) a transformation to radial coordinates consisting of a distance measure and an $(m-1)$-dimensional vector of directions. The aim of the orthogonalizing step is to rescale the sampling space such that most of the density mass of the target distribution is located around the origin. The radial transformation amounts to a generalization of the well-known polar transformation. The "direction" can be regarded as defining a line through the origin, whereas the "distance"

can be interpreted as defining a specific point on this line. In the ARDS approach, directions are obtained by applying either an MH or an importance sampling step. Next, the one-dimensional distance is sampled, *given the obtained direction*, by constructing a density grid of the target distribution and subsequently applying the (numerical) inversion method. The resulting direction-distance combination can be transformed to a drawing in the original space. The idea behind ARDS is that suitable lines, or sampling directions, can be obtained by employing a correction mechanism as in the MH algorithm or in importance sampling, and that drawing from the *target density* along such lines amounts to a one-dimensional sampling/integration problem, which is computationally feasible. As the target density is perfectly followed along the given lines, and additionally, all lines cross the "center" of the target density, the ARDS approach should result in relatively accurate estimates. In the ARDS methodology, the orthogonalizing transformation is sequentially updated by using the estimates of the first and second moments of the target distribution obtained in successive sampling rounds. Basically, this is the parametric candidate updating approach adopted by Kloek and Van Dijk (1978), Van Dijk and Kloek (1980) and Oh and Berger (1992).

The advantages of the ARDS methodology are threefold. First, the algorithms are quite parsimonious in their use of information on the shape of the target density, that is, they can often be applied successfully without a preliminary study of the shape of the posterior density. Chen *et al.* (2000) refer to such algorithms, freeing the researcher from computational details, as black-box sampling. Second, the algorithms are flexible and quite robust, that is, they can handle a large variety of target distributions featuring multi-modality, strong correlation, extreme skewness and fat tails. Third, the algorithms can handle multiple linear inequality conditions on the parameter space without any additional complications in implementation. In practice, one may want to impose such parameter restrictions, for example, for identification purposes or because of theoretical considerations. The ARDS approach extends earlier methods like the method of Box and Muller (1958), the adaptive direction sampling [ADS] algorithms proposed by Gilks *et al.* (1994), the mixed integration method by Van Dijk *et al.* (1985), and the spherical-radial integration method by Monahan and Genz (1997). We provide details later on.

5.3.2 The radial transformation

As the radial transformation is at the heart of the ARDS algorithms, we first describe the transformation from Cartesian coordinates to radial coordinates. In our notation, $y = (y_1, \ldots, y_m)$ denotes the Cartesian coordinates of a point, and (ρ, η) denotes the corresponding radial coordinates. Here, $\eta = (\eta_1, \ldots, \eta_{m-1})$ indicates the direction of the point relative to the origin, and ρ is related to the Euclidean distance.

In a general form, the radial transformation may be characterized by its inverse transformation

$$\begin{aligned} y_1 &= \rho\, h_1(\eta_1, \ldots, \eta_{m-1}), \\ &\vdots \\ y_{m-1} &= \rho\, h_{m-1}(\eta_1, \ldots, \eta_{m-1}), \\ y_m &= \rho\, h_m(\eta_1, \ldots, \eta_{m-1}), \end{aligned} \tag{5.21}$$

where $m \geq 2$ is the dimension, and $h(\eta) = (h_1(\eta), \ldots, h_m(\eta))$ are differentiable functions. The radial transformation maps the Cartesian coordinates y to a position $h(\eta)$ on the unit circle (the direction) and a stretching factor ρ which further determines the position of y given $h(\eta)$ (the distance). It can be shown that the Jacobian of the general radial transformation is given by

$$J_y(\rho, \eta) = \det\begin{pmatrix} \frac{\partial y(\rho,\eta)}{\partial \eta'} & \frac{\partial y(\rho,\eta)}{\partial \rho} \end{pmatrix} = \rho^{m-1} \det\begin{pmatrix} \frac{\partial h(\eta)}{\partial \eta'} & h(\eta) \end{pmatrix} \equiv J_y(\rho)\, J_y(\eta). \tag{5.22}$$

The form of this Jacobian has two important implications which are used in ARDS. First, it turns out that implementation of the ARDS algorithms is only based on the Jacobian factor $J_y(\rho)$, and does not depend on $J_y(\eta)$. The factor $J_y(\eta)$ drops out of the calculations, because of the multiplicative structure of (5.22). Second, as $J_y(\rho) = \rho^{m-1}$ is invariant with respect to the functions $h_1, \ldots, h_m$, our approach can be applied to any transformation satisfying (5.21).

The polar transformation

A well-known special case of the general transformation described above is the two-dimensional polar transformation with inverse transformation

$$y_1 = \rho \cos(\eta), \tag{5.23}$$
$$y_2 = \rho \sin(\eta). \tag{5.24}$$

The standard polar transformation from $y \in R^2$ to $(\rho, \eta) \in R^+ \times (0, 2\pi)$, which is one-to-one with (5.23)−(5.24), is given by

$$\rho = \sqrt{y_1^2 + y_2^2}, \tag{5.25}$$
$$\eta = \operatorname{sgn}(y_2) \arccos(y_1/\rho). \tag{5.26}$$

The left panel of Figure 5.5 illustrates the relationship between orthogonal coordinates and standard polar coordinates. A property of the standard polar transformation is that $\rho \in R^+$, implying that the direction η only defines one half of a line, and not a full line. However, in ARDS, we draw from the target density along unbounded lines to take into account as much of the shape of this target density as possible. The signed polar

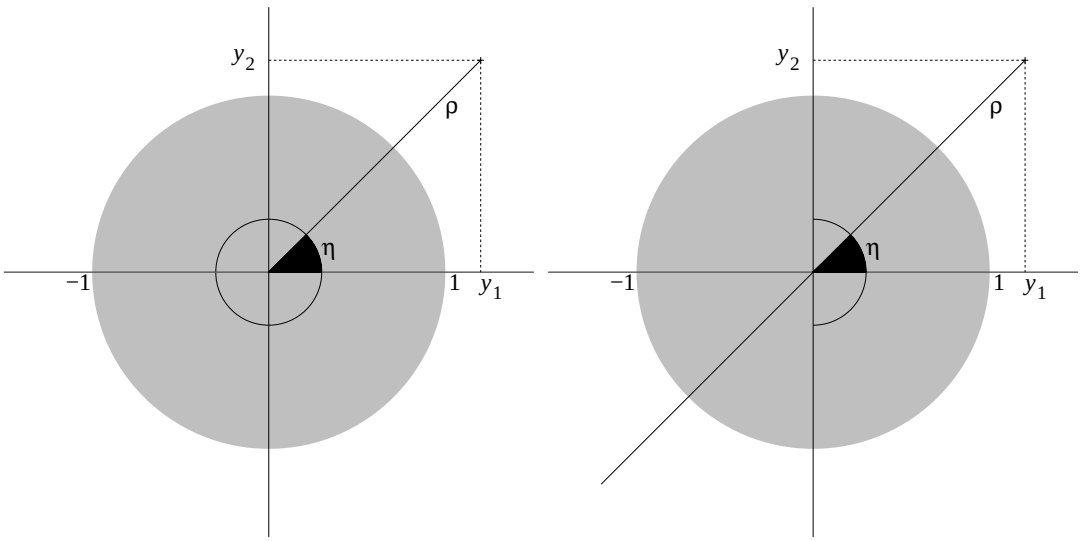

Figure 5.5: The relationship between Cartesian coordinates and polar coordinates in the two-dimensional case: standard polar coordinates in the left panel and signed polar coordinates in the right panel.

transformation from $y \in R^2$ to $(\rho, \eta) \in R \times (-\pi/2, \pi/2)$ is a variant of the standard polar transformation such that η does define a complete line. It is given by

$$\rho = \text{sgn}(y_1)\sqrt{y_1^2 + y_2^2}, \tag{5.27}$$

$$\eta = \arcsin(y_2/\rho), \tag{5.28}$$

with inverse transformation again defined by (5.23)−(5.24). The relationship between orthogonal coordinates and signed polar coordinates is shown in the right panel of Figure 5.5.

An efficient radial transformation

A drawback of the polar transformation is that it is possible but not straightforward to generalize it to more than two dimensions, see Muirhead (1982, Theorems 1.5.5 and 2.1.3). Furthermore, the polar transformation is computationally not very efficient. We therefore propose a transformation which satisfies (5.21), is easy to generalize to more than two dimensions, and is more efficient than the polar transformation. For m dimensions, the transformation from $(y_1, \ldots, y_m) \in R^m$ to $(\rho, \eta) = (\rho, \eta_1, \ldots, \eta_{m-1}) \in R \times \{\eta \in R^{m-1} : \eta'\eta < 1\}$ is given by

$$\rho = \text{sgn}(y_m)\sqrt{y'y}, \tag{5.29}$$

$$\eta_j = \frac{y_j}{\rho}, \qquad j = 1, \ldots, m-1, \tag{5.30}$$

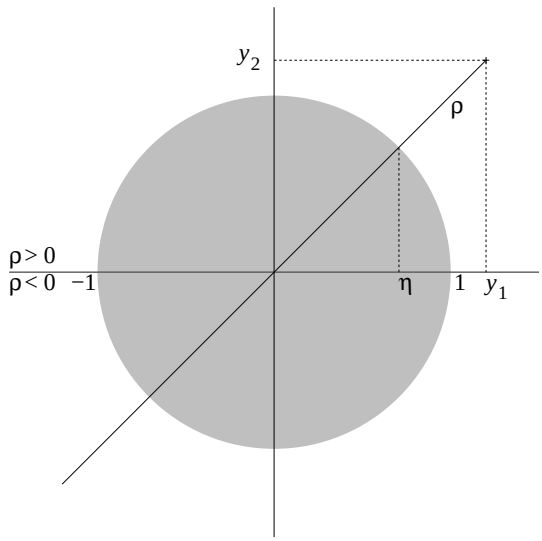

Figure 5.6: The relationship between Cartesian coordinates and radial coordinates in the two-dimensional case.

with inverse transformation

$$y_j = \rho\,\eta_j, \qquad j = 1, \ldots m-1, \tag{5.31}$$

$$y_m = \rho\sqrt{1-\eta'\eta}\,. \tag{5.32}$$

The Jacobian of this transformation is given by

$$J_y(\rho,\eta) = J_y(\rho)\,J_y(\eta) = \rho^{m-1}(1-\eta'\eta)^{-1/2}. \tag{5.33}$$

Basically, y is transformed to $m-1$ Cartesian coordinates η on the unit circle and a stretching factor ρ. This is illustrated in Figure 5.6 for $m = 2$ dimensions. We note that the sign of ρ determines whether y is located above or below the y_1 axis.

5.3.3 Adaptive radial-based Metropolis-Hastings sampling

We start by defining the adaptive radial-based Metropolis-Hastings sampling algorithm [ARMHS], which is based on a candidate generating density that is taken to be multivariate normal with parameters μ and Σ. However, in fact, any elliptically contoured candidate distribution can be considered without affecting the sampling results. Any density within the elliptical class is of the form $c_m(\det(\Sigma))^{-1/2} f((x-\mu)'\Sigma^{-1}(x-\mu))$, where c_m is a normalizing constant, see, for example, Muirhead (1982, Section 1.5). Elliptical distributions are symmetric, unimodal, and their kurtosis is determined by the function f. A fundamental property is that they are preserved under affine transformations, although the mean and the covariance matrix might change. This property is essential for the derivation of the ARDS algorithms. The normal density would arise

when $f(u) = \exp(-u/2)$. The m-variate Student-t density with ν degrees of freedom corresponds to $f(u) = (1 + u/\nu)^{-(\nu+m)/2}$. As drawing from a normal distribution is more efficient than drawing from a Student-t density or any other elliptically contoured density, the normal distribution is used in the implementation of ARMHS and any other ARDS algorithm. We emphasize that, although for expository purpose the subsequent discussion is focussed on convenient normal candidate drawings, these drawings may be obtained from any member of the class of elliptical distributions.

ARMHS is related to the independence chain MH algorithm. It considers drawings from a normal $N(\mu, \Sigma)$ candidate density, where hopefully μ and Σ provide good approximations to the unknown mean and covariance matrix of the target distribution. However, in contrast to the standard MH algorithm, these drawings are not immediately used for construction of a Markov chain in the original parameter space. Instead, a composite transformation is made. For expository purpose, we treat this transformation explicitly in two steps.

The first step in the composite transformation concerns a location-scale transformation of a realization x into a realization y. The aim of this orthogonalizing step is to rescale the sampling space such that most of the density mass of the target distribution is located around the origin. This rescaling corresponds to the affine transformation[1]

$$y = y(x|\,\mu, \Sigma) = \Sigma^{-1/2}(x - \mu), \tag{5.34}$$

with inverse transformation

$$x = x(y|\,\mu, \Sigma) = \mu + \Sigma^{1/2} y, \tag{5.35}$$

and Jacobian

$$J_x(y) = \det(\Sigma^{1/2}). \tag{5.36}$$

The second step is the radial transformation from y to (ρ, η), defined by (5.29) and (5.30), with inverse transformation given by (5.31) and (5.32), and Jacobian (5.33).

Combining the two transformations results in the composite transformation, denoted by

$$\begin{pmatrix} \rho \\ \eta \end{pmatrix} = \begin{pmatrix} \rho(x|\,\mu, \Sigma) \\ \eta(x|\,\mu, \Sigma) \end{pmatrix} \equiv \begin{pmatrix} \rho(\,y(x|\,\mu, \Sigma)\,) \\ \eta(\,y(x|\,\mu, \Sigma)\,) \end{pmatrix}, \tag{5.37}$$

with inverse transformation

$$x = x(\rho, \eta|\,\mu, \Sigma) \equiv x(\,y(\rho, \eta)\,|\,\mu, \Sigma), \tag{5.38}$$

and Jacobian

$$J_x(\rho, \eta) = J_y(\rho, \eta)\, J_x(y) = J_y(\rho)\, J_y(\eta) \det(\Sigma^{1/2}). \tag{5.39}$$

Applying the two transformations to a candidate realization x_i^* from $N(\mu, \Sigma)$ provides a one-dimensional distance ρ_i^* and an $(m-1)$-dimensional direction η_i^*. Discarding the

[1] $\Sigma^{1/2}$ denotes the Cholesky decomposition of Σ, and $\Sigma^{-1/2}$ denotes the inverse matrix of $\Sigma^{1/2}$.

distance ρ_i^*, the candidate direction η_i^* is either accepted or rejected in an MH step, that is, the direction becomes

$$\eta_i = \begin{cases} \eta_i^* & \text{with probability } \alpha(\eta_{i-1}, \eta_i^*) \\ \eta_{i-1} & \text{with probability } 1 - \alpha(\eta_{i-1}, \eta_i^*) \end{cases} \tag{5.40}$$

for an appropriate acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$, which is derived below. An iteration of ARMHS is completed by drawing from the target distribution on the line defined by the direction η_i. This can be done as follows. First, one draws a distance ρ_i from the transformed target density $p(\rho|\eta_i)$, given the direction η_i, by constructing a density grid and using the (numerical) inversion method. Next, ρ_i and η_i are transformed to the original space by inverting the composite transformation applied earlier. In sum, an iteration of ARMHS consists of the following steps, that is,

1. Generate x_i^* from $N(\mu, \Sigma)$.
2. Transform x_i^* to y_i^* using (5.34).
3. Transform y_i^* to ρ_i^* and η_i^* using (5.29) and (5.30).
4. Apply MH step to η_i, see (5.40).
5. Generate ρ_i from $p(\rho|\eta_i)$ by constructing a grid and using the inversion method.
6. Transform ρ_i and η_i to y_i using (5.31) and (5.32).
7. Transform y_i to x_i using (5.35).

We note that the first two steps can be combined by immediately drawing y_i^* from $N(0, I_m)$, where I_m denotes the $m \times m$ identity matrix.

Step 4 of an ARMHS iteration requires the acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$, and step 5 requires the distribution of the distance ρ conditional on the direction η_i. It is shown in Appendix 5.A that, for all elliptically contoured candidate distributions with mean μ and covariance matrix Σ,

$$\alpha(\eta_{i-1}, \eta_i^*) = \min\left\{ \frac{I(\eta_i^*)}{I(\eta_{i-1})}, 1 \right\}, \tag{5.41}$$

where

$$I(\eta) = \int_{-\infty}^{\infty} \kappa(\rho|\eta)\, d\rho, \tag{5.42}$$

and where $\kappa(\rho|\eta)$ is a kernel of the conditional density $p(\rho|\eta)$, defined by

$$p(\rho|\eta) \propto \kappa(\rho|\eta) = p(\, x(\rho, \eta|\, \mu, \Sigma)\,)\, |J_y(\rho)|. \tag{5.43}$$

Under very mild regularity conditions, the sampled Markov chain converges in distribution to the target distribution, see Bauwens *et al.* (2004).

An interesting property of the acceptance probability (5.41) is that it does not depend on the functional form of the candidate density provided that this candidate density is elliptically contoured. However, it does depend on the generated direction η and hence on the mean and covariance matrix of the candidate distribution. To compute (5.41), requiring evaluation of the one-dimensional integral $I(\eta)$, we use the iterative Simpson's procedure discussed in Subsection 5.2.1. As the density of ρ conditional on η is proportional to the integrand of $I(\eta)$, the function evaluations required in the Simpson's procedure can also be used to construct a grid for $p(\rho|\eta)$. Hence, sampling of the distance ρ conditional on the direction η, that is, step 5 in the ARMHS iteration, is straightforward and cheap in terms of computing time. To capitalize on the obtained density grid, one may consider several drawings of ρ for each drawing of η. This amounts to collecting several drawings from the target density along the considered line. Although such drawings would be dependent, the approach still provides consistent estimates, see Geweke (1999, p. 44).

The integral $I(\eta)$ has infinite integration bounds. However, we use finite bounds for its numerical evaluation. Of course, it is important to take these bounds such that practically all density mass of ρ given η is included, but for the sake of efficiency it is also desirable that the integration interval is as small as possible. To obtain bounds for ρ, we impose minimum and maximum values for each element of x in the original space. It is often possible to find sensible bounds by either theory and/or common sense. More generally, by imposing bounds on the values of x we impose linear inequality conditions on the original parameter space. These restrictions are of the form $c_j' x \leq b_j$, where j indicates the number of the restriction. For given η, the conditions imposed on the original space translate to bounds $\rho_{\min}$ and $\rho_{\max}$ for ρ through the relationships $\rho_{\min} = \max_j\{\rho_j : \rho_j < 0\}$ and $\rho_{\max} = \min_j\{\rho_j : \rho_j > 0\}$, where $\rho_j = \frac{b_j - c_j'\mu}{c_j'(\tilde{x}-\mu)}$ with $\tilde{x} = x(\rho = 1, \eta|\, \mu, \Sigma)$. As additional linear restrictions reduce the integration interval for $I(\eta)$, making evaluation of $I(\eta)$ more efficient, they do not put a burden on the algorithm, but they might result in an efficiency gain. We note that μ should be in the feasible region, and that one may want to split up the integration interval $(\rho_{\min}, \rho_{\max})$ into two subintervals $(\rho_{\min}, 0)$ and $(0, \rho_{\max})$ and apply deterministic integration to these subintervals separately. The Jacobian $J_y(\rho) = \rho^{m-1}$ implies that the integrand of $I(\eta)$, and hence $p(\rho|\eta)$, has value 0 at $\rho = 0$. Taking this given point into account turns out to be rewarding.

For implementation of ARMHS, the mean μ and the covariance matrix Σ of the normal candidate distribution have to be specified. Heuristically, the quality of the sample should improve if μ and Σ are close to, rather than far from, the target mean and covariance matrix, respectively. ARMHS employs an adaptive updating approach. Given a generated sample $x_1, x_2, \ldots, x_n$ from a previous run of the algorithm, μ and Σ are replaced by their

Monte Carlo estimates, which are given by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{5.44}$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})(x_i - \hat{\mu})', \tag{5.45}$$

respectively. Using these estimates, one can proceed with a new sampling round. This process can be repeated any number of times. We note that information coming from a "wrong" sample may have a misleading effect and may worsen convergence. Hence, convergence should be monitored by usual tools, see Van Dijk and Kloek (1980) and Oh and Berger (1992). However, as only the direction η, and not the distance ρ, is generated from the candidate distribution, the risk of collecting a "wrong" sample is limited. ARMHS is quite robust, as the distance ρ conditional on the direction η is immediately obtained from the target distribution, that is, sampling on a given line mimics the target density. To monitor convergence over sampling rounds, we find the Mahalanobis distance particularly useful. It is defined as $\text{Mah}_j = (\hat{\mu}^{(j)} - \hat{\mu}^{(j-1)})'[\hat{\Sigma}^{(j)}]^{-1}(\hat{\mu}^{(j)} - \hat{\mu}^{(j-1)})$, where j indicates the sampling round. The Mahalanobis distance measures the extent to which the estimated posterior mean changes between successive sampling rounds, while taking into account parameter uncertainty and the underlying correlation structure.

Illustration

Figure 5.7 illustrates ARMHS for a bivariate bimodal target distribution. The upper two graphs display the target density in the original space. A point, representing a realization from the normal candidate distribution $N(\mu, \Sigma)$, is shown in the contour plot. If μ and Σ would coincide with the mean and the covariance matrix of the target distribution, the location-scale transformation would lead to the target density depicted in the middle graphs. The gain of the orthogonalizing transformation is that the density mass is better located around the origin in the sense that a line through the origin, defined by a direction η, "hits the density mass" more easily. As ARMHS precisely considers such lines, the location-scale transformation may lead to a substantial improvement for appropriate μ and Σ. The target density after applying the radial transformation is depicted in the bottom two graphs.

We have distinguished seven steps in an iteration of ARMHS. The visualization of these steps in Figure 5.7 is as follows. In step 1, the point in the upper contour plot is drawn from $N(\mu, \Sigma)$. This point is transformed in step 2 to the point in the middle contour plot. Step 3 provides the point in the bottom contour plot. The direction η, that is, the horizontal position of the point in the bottom contour plot, is either accepted or rejected in step 4. In this illustration, we accept η. Step 5 consists of drawing one (or several) distance(s) ρ on the vertical line through the point. Step 6 can be represented by the transformation of points generated on the line in the bottom contour plot to

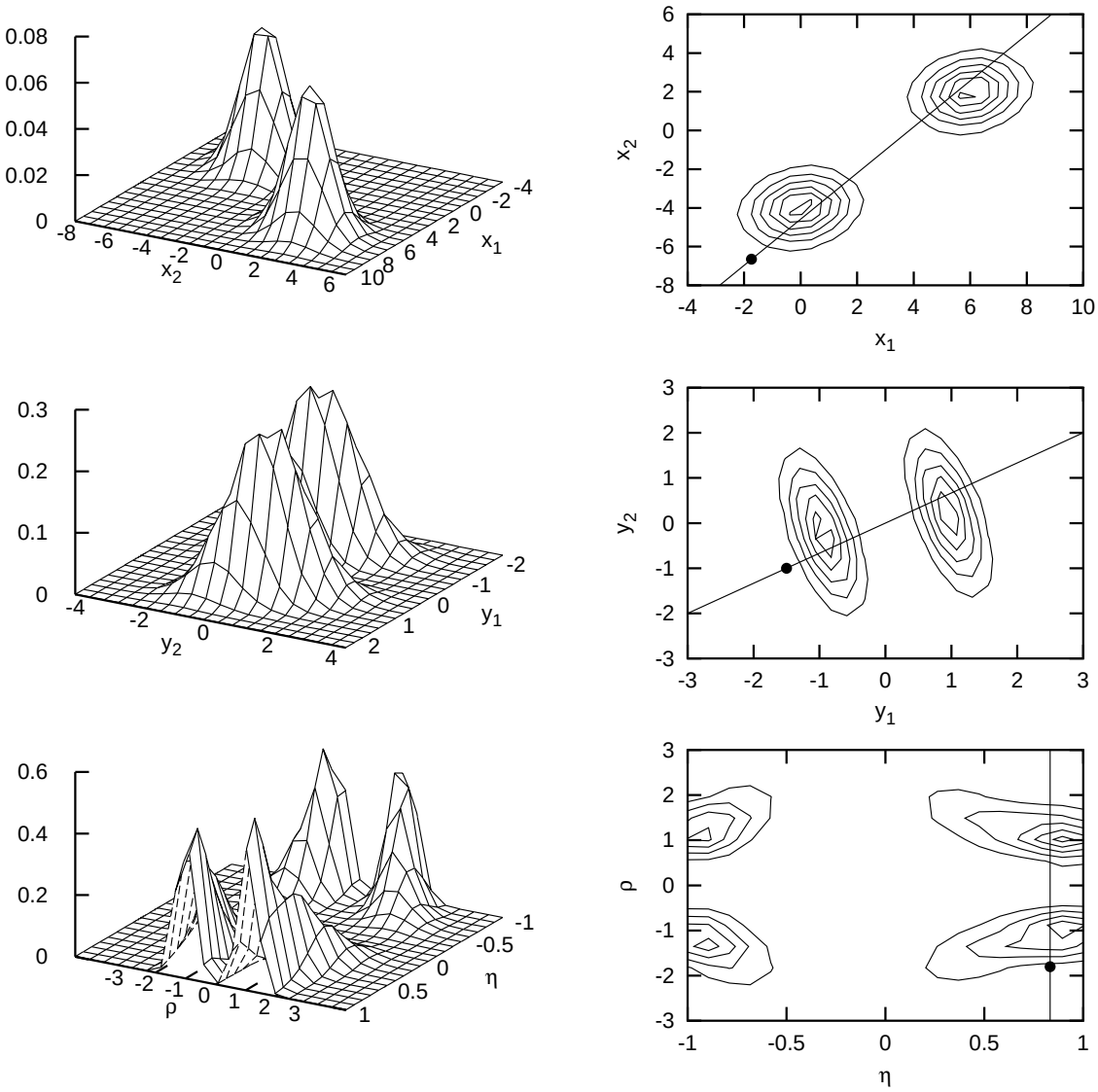

Figure 5.7: Adaptive radial-based direction sampling: target density in original space (above), target density after location-scale transformation (middle) and target density after radial transformation (below).

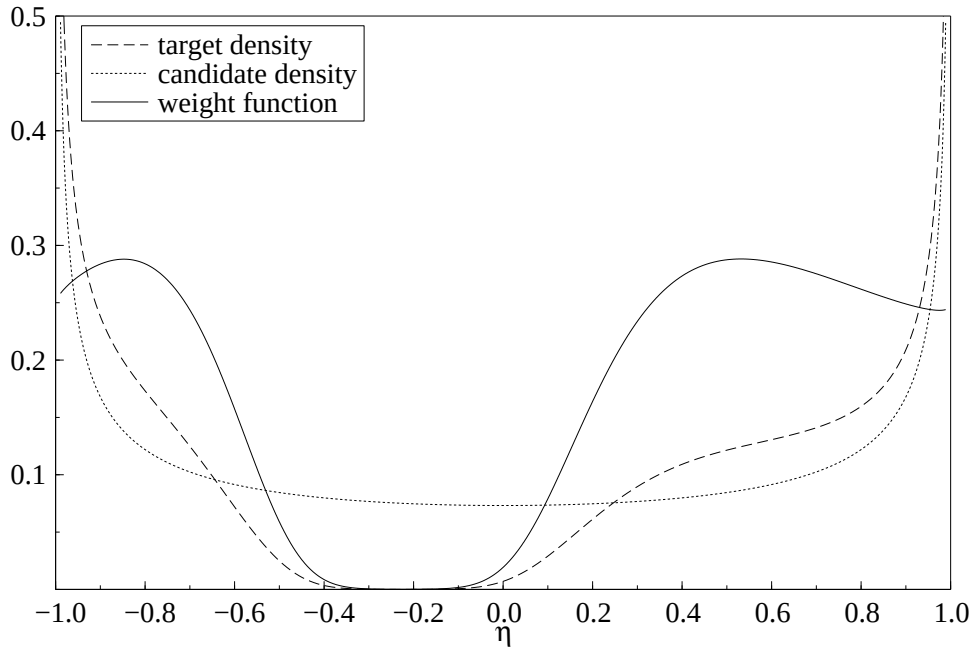

Figure 5.8: The marginal target density $p(\eta)$, the marginal candidate density $q(\eta)$, and the weight function $p(\eta)/q(\eta)$ (up to scaling constants). The original target density is the same as in the previous figure.

points generated on the line in the middle contour plot. Similarly, step 7 results in points generated on the line in the upper contour plot. We note that the MH acceptance-rejection mechanism employed in ARMHS takes into account the shape of the target density, so that most of the generated lines correspond to directions with "high" density.

Figure 5.8 shows the marginal target density $p(\eta)$ resulting from the bottom two graphs in Figure 5.7. This density looks ill-behaved. However, similar to the relationship between the standard independence chain MH algorithm and standard importance sampling, it is not directly the target density that matters in the MH step of ARMHS. What matters is the ratio of the target density to the candidate density, that is, $p(\eta)/q(\eta)$, where $q(\eta)$ is the marginal density resulting from the elliptically contoured candidate distribution. It is seen from Figure 5.8 that the weight function $p(\eta)/q(\eta)$ is well-behaved in the sense that variation in its values is only moderate. We further note from the bottom two graphs in Figure 5.7 that the transformed target density is bimodal with respect to the distance ρ. However, this is not a problem, as a grid for $p(\rho|\eta)$ can always be constructed easily.

5.3.4 Adaptive radial-based importance sampling

Adaptive radial-based importance sampling [ARIS] replaces the MH step of ARMHS for the direction η by an importance sampling step. In ARIS, every sampled direction η_i is kept, a distance ρ_i is sampled conditional on it, and the resulting radial coordinates are transformed to a drawing x_i in the original space, which is weighted according to the

appropriate importance weight

$$w(\eta_i) = \frac{p(\eta_i)}{q(\eta_i)} \propto I(\eta_i), \tag{5.46}$$

where $I(\eta)$ is defined by (5.42), see Appendix 5.A. In sum, an iteration of ARIS consists of the following steps, that is,

1. Generate x_i^* from $N(\mu, \Sigma)$.
2. Transform x_i^* to y_i^* using (5.34).
3. Transform y_i^* to ρ_i^* and η_i^* using (5.29) and (5.30).
4. Apply importance sampling step to η_i: $\eta_i = \eta_i^*$ with weight $w(\eta_i)$ defined by (5.46).
5. Generate ρ_i from $p(\rho|\eta_i)$ by constructing a grid and using the inversion method.
6. Transform ρ_i and η_i to y_i using (5.31) and (5.32).
7. Transform y_i to x_i using (5.35).

Again, the first two steps can be combined by immediately drawing y_i^* from $N(0, I_m)$, where I_m denotes the $m \times m$ identity matrix.

An interpretation of ARIS is that one samples from the target distribution on lines with directions being derived from the elliptically contoured candidate distribution. Each line receives a weight, indicating the importance of the underlying direction. The weight of a line is carried over to any realization on that line. Alternatively, one may interpret ARIS as being just a special case of standard importance sampling. A realization x in the original space is a function of a realization (ρ, η) in the transformed space, see (5.38), implying that the importance weight of (ρ, η) is also the importance weight of x. Taken together, step 1 to step 5 of an ARIS iteration can be regarded as providing a realization (ρ_i, η_i) from a candidate distribution with density

$$q_{\text{imp}}(\rho, \eta) = q(\eta)\, p(\rho|\eta), \tag{5.47}$$

and providing the corresponding importance weight

$$w(\rho, \eta) = \frac{p(\rho, \eta)}{q_{\text{imp}}(\rho, \eta)} = \frac{p(\eta)\, p(\rho|\eta)}{q(\eta)\, p(\rho|\eta)} = \frac{p(\eta)}{q(\eta)}\,, \tag{5.48}$$

which coincides with (5.46). We note that the importance function q_{imp} takes into account the shape of the target density through $p(\rho|\eta)$. This is not the case in standard importance sampling with importance function $q(\eta)\, q(\rho|\eta)$ in the transformed space, explaining that the importance function of ARIS is more accurate than the importance function in standard importance sampling with a normal or Student-t candidate density. As ARIS

can be interpreted as a special case of importance sampling, its convergence properties follow immediately, see Geweke (1989) for details.

Similar to ARMHS, the parameters μ and Σ of the location-scale transformation can be updated by replacing them by their Monte Carlo estimates. These estimates are given by

$$\hat{\mu}_w = \frac{\sum_{i=1}^{n} w(\eta_i)\, x_i}{\sum_{i=1}^{n} w(\eta_i)}\,, \tag{5.49}$$

$$\hat{\Sigma}_w = \frac{\sum_{i=1}^{n} w(\eta_i)\, (x_i - \hat{\mu}_w)(x_i - \hat{\mu}_w)'}{\sum_{i=1}^{n} w(\eta_i)}\,, \tag{5.50}$$

where $x_1, x_2, \ldots, x_n$ is the collected sample, and $w(\eta_1), w(\eta_2), \ldots, w(\eta_n)$ are the corresponding importance weights.

5.3.5 A classification of related methods

The ARDS methodology is related to several algorithms put forward in the literature. First, it can be regarded as an extension of the well-known method of Box and Muller (1958) to generate two normal random variables. The latter method is based on the two-dimensional standard polar transformation, whereas the ARDS algorithms can be implemented for a broad class of radial transformations. Furthermore, the ARDS methods generate directions η using an MH or an importance sampling step, and generate distances ρ from a very accurate numerical approximation to the model-specific target distribution. Basically, the Box-Muller method generates a direction η and a distance ρ for a bivariate normal distribution, and transforms the resulting direction-distance combination to the original space. The employed polar transformation implies that η is uniformly distributed between 0 and 2π and that ρ^2 has a χ^2-distribution with two degrees of freedom, or equivalently, an exponential distribution with mean 2. In the Box-Muller algorithm, drawings from the exponential distribution are obtained through the inversion method, see, for example, Law and Kelton (1991, p. 491).

It is also interesting to compare ARDS with the class of adaptive direction sampling [ADS] algorithms proposed by Gilks *et al.* (1994). In ADS, only information on the shape of the target density is used, and no candidate density is considered. ADS was developed as a flexible alternative to the Gibbs sampler, which is restricted by its orthogonal sample paths. Another difference is that, when using ARDS, one generates distances ρ from a numerically very accurate approximation to the target distribution. This step is not spelled out in ADS. As ARMHS and ARIS are members of the MH and importance sampling classes, the convergence properties of these methods are well established. This is not so transparent for the ADS methods, see Gilks *et al.* (1994) and Roberts and Gilks (1994).

The proposed ARDS class contains several algorithms. So far, we have discussed ARMHS and ARIS. However, one may also define an adaptive radial-based rejection

sampling variant [ARRES]. In this algorithm, the sampled direction η_i is accepted if $u < \frac{w(\eta_i)}{c}$, and it is rejected otherwise. Here, u is uniformly drawn in $(0,1)$ and c is a constant such that $w(\eta) \leq c$ for all $\eta \in \{\eta \in R^{m-1} : \eta'\eta < 1\}$. The Monte Carlo updates (5.44) and (5.45) for μ and Σ apply to the accepted transformed drawings.

As another variant, one can replace sampling of the distances ρ by evaluating the one-dimensional integral of interest after transforming the original integral to radial coordinates and conditioning on the generated directions η. One can combine such deterministic integration with respect to ρ with rejection sampling, importance sampling or MH sampling with respect to η, and evaluate posterior moments. For the case of importance sampling, this has been done in the so-called mixed integration method of Van Dijk *et al.* (1985). A related method is described by Monahan and Genz (1997) who use the term spherical-radial integration. These methods can be regarded as special cases of the ARDS class in which the step of generating random drawings of ρ is reduced to evaluating only a one-dimensional integral. A limitation of deterministic integration with respect to ρ is that one has to compute a different one-dimensional integral for each moment of the target distribution in which one is interested.

5.4 Applications

In this section, we use a set of models to illustrate the versatility of the radial-based algorithms. We compare ARMHS and ARIS to the (independence chain) MH algorithm and to importance sampling. In all examples, the MH method and importance sampling are based on a Student-t candidate density with 5 degrees of freedom. We start with an artificial example in which the moments of the posterior distribution are known by construction. Next, we consider two empirical examples. The first empirical example involves a regression model with scale contamination to investigate a study from Justel and Peña (1996) concerning the oxidation of nitric acid in a plant. Second, we consider a mixture model as in Frühwirth-Schnatter (2001) to analyze economic growth in the USA. We note that in these two empirical examples the mixture process refers to the data space. However, such mixture processes may give rise to bimodality, extreme correlation and skewness in the parameter space.

An artificial example

To illustrate the performance of ARDS, we first consider an 8-dimensional trimodal mixture distribution featuring skewness, high correlation (varying from -0.95 to 0.90) and multimodality. This target distribution is given by

$$p_1\, N(\mu_1, \Sigma_1) + p_2\, N(\mu_2, \Sigma_2) + p_3\, N(\mu_3, \Sigma_3),$$

Table 5.1: Sampling results for the trimodal mixture distribution.

	bounds ARDS		initialization		ARMHS		ARIS		MH		IS		true	
	min.	max.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
x_1	−20.00	20.00	6.00	10.00	7.03	4.41	7.07	4.40	1.83	0.04	7.44	0.11	7.00	4.42
x_2	−20.00	20.00	6.00	10.00	7.54	3.38	7.52	3.43	2.57	0.04	8.10	0.15	7.50	3.40
x_3	−20.00	20.00	6.00	10.00	8.03	2.73	7.98	2.70	4.09	0.35	10.21	0.08	8.00	2.74
x_4	−20.00	20.00	6.00	10.00	8.54	2.70	8.47	2.73	6.01	0.39	11.21	0.40	8.50	2.74
x_5	−20.00	20.00	6.00	10.00	4.99	2.71	5.06	2.70	6.13	0.45	2.96	0.91	5.00	2.74
x_6	−20.00	20.00	6.00	10.00	5.44	2.77	5.49	2.72	8.91	0.62	2.14	0.61	5.50	2.74
x_7	−20.00	20.00	6.00	10.00	6.07	3.38	5.96	3.41	10.21	0.32	4.24	0.48	6.00	3.40
x_8	−20.00	20.00	6.00	10.00	6.48	4.39	6.44	4.46	12.01	1.18	5.14	0.85	6.50	4.42
drawings per iteration ($\eta \times \rho$)					5000×5		5000×5		250000		250000			
number of iterations					8		5		8		8			
average time per iteration (in s)					112.9		82.0		110.4		93.9			
Mahalanobis distance					0.01		0.01		1.57		10.68			
acceptance rate (in %)					27.8				0.1					
5% most influential weights (in %)							29.2				99.7			

where

$$\begin{aligned}
\mu_1 &= 1.5\,(1,2,3,4,5,6,7,8)',\\
\mu_2 &= 1.5\,(5,6,7,8,1,2,3,4)',\\
\mu_3 &= 1.5\,(8,7,6,5,4,3,2,1)',\\
\Sigma_1 &= \Sigma_2 = \Sigma_3 = I_8,\\
p_1 &= p_2 = p_3 = 1/3.
\end{aligned}$$

We estimate the first and second moments of this distribution using ARMHS, ARIS, MH and importance sampling. This is done in several sampling rounds. In our adaptive approach, additional sampling rounds are considered as long as the Mahalanobis distance is larger than 0.02. However, in the experiment we allow for at most 8 rounds. In any round, ARMHS and ARIS draw 5000 directions and 5 distances per direction, resulting in a sample of size 25000. To make the computing times comparable, the MH and importance sampling algorithms are allowed to collect a larger sample of size 250000. The mean for the initial candidate density is set at 6 for all eight components. Furthermore, the scale is taken sufficiently large so that the MH algorithm and importance sampling can initially cover the whole target density.

The estimates of the location and scaling parameters are reported in Table 5.1, together with the true parameter values. The table indicates that ARMHS and ARIS do a good job, whereas the MH algorithm and importance sampling fail. ARIS only needs 5 rounds to reach convergence (according to our definition), whereas the other three algorithms need the maximum number of rounds. However, after eight rounds also ARMHS has converged, whereas the other two algorithms clearly have not, see the reported Mahalanobis distances which concern the final sampling round. We note that the average computing times per sampling round are comparable for the four algorithms.

The acceptance rates for ARMHS and MH (reported for the final round) show a large difference in values. Furthermore, the moment estimates, obtained from importance sampling, are almost completely determined (99.7%) by only 5% of the drawings. In contrast, in ARIS the 5% most influential drawings only account for 29.2% of the total weight.

Finally, to see whether the favorable results are merely a coincidence, we repeat the experiment above ten times for different seeds of the random number generator. The results are robust. ARMHS succeeds 9 times and ARIS even succeeds for all 10 repetitions, whereas MH and importance sampling only have success rates of 30%, making the outcomes of the latter two methods unreliable.

Table 5.2: Sampling results for the scale contamination model.

	bounds		ARDS initialization		ARMHS		ARIS		MH		IS		large sample	
	min.	max.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
β_1	-10.00	10.00	0.00	2.00	0.81	0.19	0.81	0.19	0.81	0.20	0.81	0.20	0.81	0.19
β_2	-10.00	10.00	0.00	2.00	1.01	0.52	1.00	0.55	1.01	0.58	1.02	0.57	1.01	0.54
β_3	-10.00	10.00	0.00	2.00	-0.61	0.09	-0.61	0.10	-0.61	0.10	-0.61	0.10	-0.61	0.10
σ	0.00	10.00	5.00	2.00	2.84	1.32	2.77	1.38	3.06	1.38	3.09	1.39	2.82	1.36
ω	1.00	10.00	5.00	2.00	3.40	2.35	3.48	2.39	3.60	2.48	3.70	2.58	3.48	2.40
p	0.00	1.00	0.50	0.20	0.47	0.34	0.51	0.34	0.43	0.34	0.43	0.34	0.49	0.34
drawings per iteration ($\eta \times \rho$)					5000×5		5000×5		250000		250000			
number of iterations					5		4		4		4			
average time per iteration (in s)					39.8		37.6		43.1		47.0			
Mahalanobis distance					0.02		0.00		0.01		0.01			
acceptance rate (in %)					58.2				11.4					
5% most influential weights (in %)							19.6				62.8			

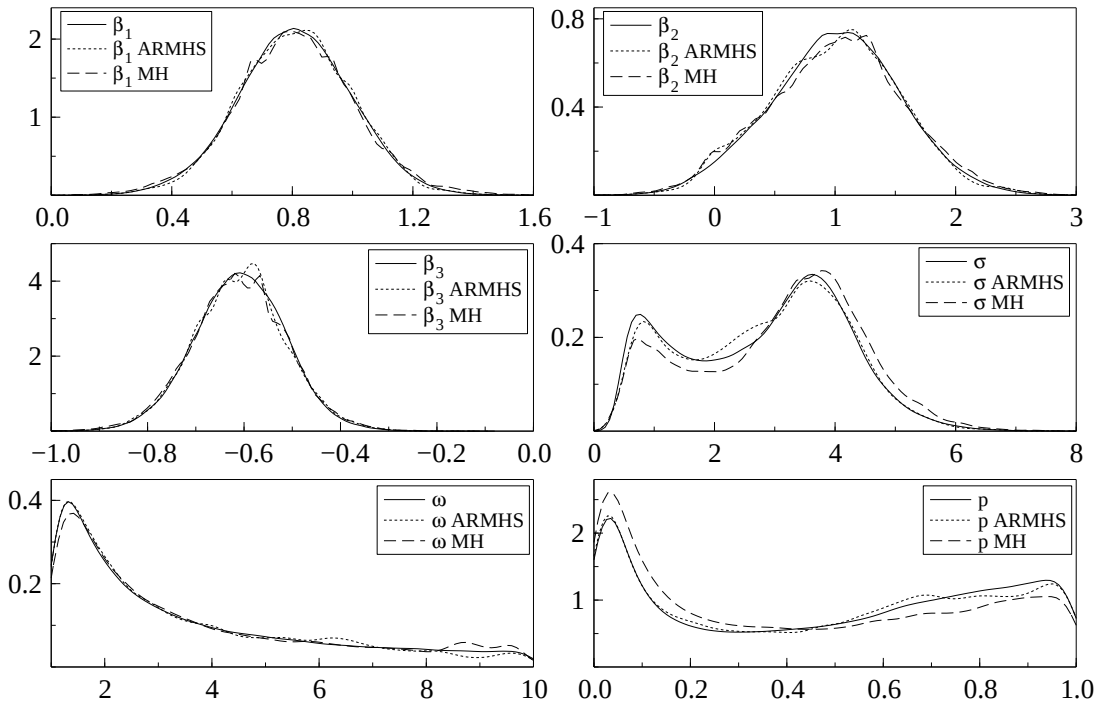

Figure 5.9: Marginal densities for the scale contamination model.

Scale contamination

Justel and Peña (1996) investigate a data set from Brownlee (1965, p. 491−500) concerning the oxidation of ammonia to nitric acid in a plant. The data set contains 21 daily observations on four variables. The stack loss rate y_t, that is, the proportion of ingoing ammonia to the plant that escapes unabsorbed, is related to the amount of air flow $x_{1,t}$ (representing the rate of operation of the plant), the temperature of the cooling water $x_{2,t}$, and the concentration of the circulating acid $x_{3,t}$. However, several observations might be classified as outliers. In a regression setting, it is sufficient to allow for scale contamination, as in the model

$$
\begin{aligned}
&y_t = \beta_1\, x_{1,t} + \beta_2\, x_{2,t} + \beta_3\, x_{3,t} + \varepsilon_t,\\
&\varepsilon_t \sim N(0, \sigma_t^2),\\
&\sigma_t = \begin{cases} \sigma & \text{with probability } 1-p\\ \omega\sigma & \text{with probability } p, \end{cases}
\end{aligned}
$$

For identification of the two variance regimes, we impose that $\omega > 1$. The prior for the parameter vector $(\beta_1, \beta_2, \beta_3, \sigma, \omega, p)'$ is chosen proportional to $[(1-p)\sigma + p\omega\sigma]^{-1}$ within the parameter bounds reported in Table 5.2. The prior density is zero outside these bounds.

Estimates of the mean and the covariance matrix of the six model parameters are obtained by considering drawings from ARMHS, ARIS, MH and importance sampling. The sampling setup is similar to the setup in the previous example. The parameter

estimates are reported in Table 5.2, together with the corresponding large sample values (computed from 250000 ARMHS drawings). All four methods trace back the response parameters β_1, β_2 and β_3 accurately. However, this is not the case for the remaining three parameters. In particular, MH and importance sampling underestimate the left mode of the bimodal marginal density of σ, resulting in overestimation of the mean of σ. This is also illustrated in Figure 5.9 in which the large sample marginal densities (250000 ARMHS drawings) and the estimated densities from 25000 ARMHS drawings and from 250000 MH drawings are shown. Furthermore, MH and importance sampling underestimate the right mode of the bimodal marginal density for the mixture probability p, resulting in underestimation of the mean of p. We note from Table 5.2 that the average computing times per sampling round are again comparable for the four considered algorithms, and that all algorithms have converged (in terms of our Mahalanobis distance definition) within 4 or 5 rounds.

The difference between the reported acceptance rates of ARMHS and MH is striking. The acceptance rate of the former method is about five times as high as the acceptance rate of the latter method. Furthermore, the distribution of the importance weights for ARIS is much more uniform than the weight distribution for importance sampling. In the final round of the importance sampling procedure, the 5% most influential drawings have 63% of the total weight, whereas this is only about 20% in ARIS. Again, this demonstrates the accuracy of the ARDS algoritms.

A mixture model for the growth rate of US GNP

As another illustration of our algorithms, we consider a mixture model with two AR(1) regimes for the growth of real gross national product [real GNP]. This model is similar to the model considered in Frühwirth-Schnatter (2001) where another recent sampling method is discussed. The model is given by

$$\begin{aligned} y_t &= \begin{cases} \beta_{11} + \beta_{12}\, y_{t-1} + \varepsilon_t & \text{with probability} \quad p \\ \beta_{21} + \beta_{22}\, y_{t-1} + \varepsilon_t & \text{with probability} \quad 1-p, \end{cases} \\ \varepsilon_t &\sim N(0, \sigma^2), \end{aligned} \tag{5.51}$$

where y_t denotes the quarterly growth rate. We investigate data concerning US real GNP (source: Economagic). The data consist of observations from the first quarter of 1959 to the last quarter of 2001. Figure 5.10 displays both the real GNP level and the quarterly growth rate (defined as 100 times the first difference of the logarithm). The priors for β_{11}, β_{12}, β_{21}, β_{22} and p are chosen uniform, and the prior for σ is taken proportional to $1/\sigma$. For identification, it is imposed that $\beta_{11} < \beta_{21}$. Again parameter bounds apply, see Table 5.3. The sampling setup is identical to the setup in the previously considered scale contamination example.

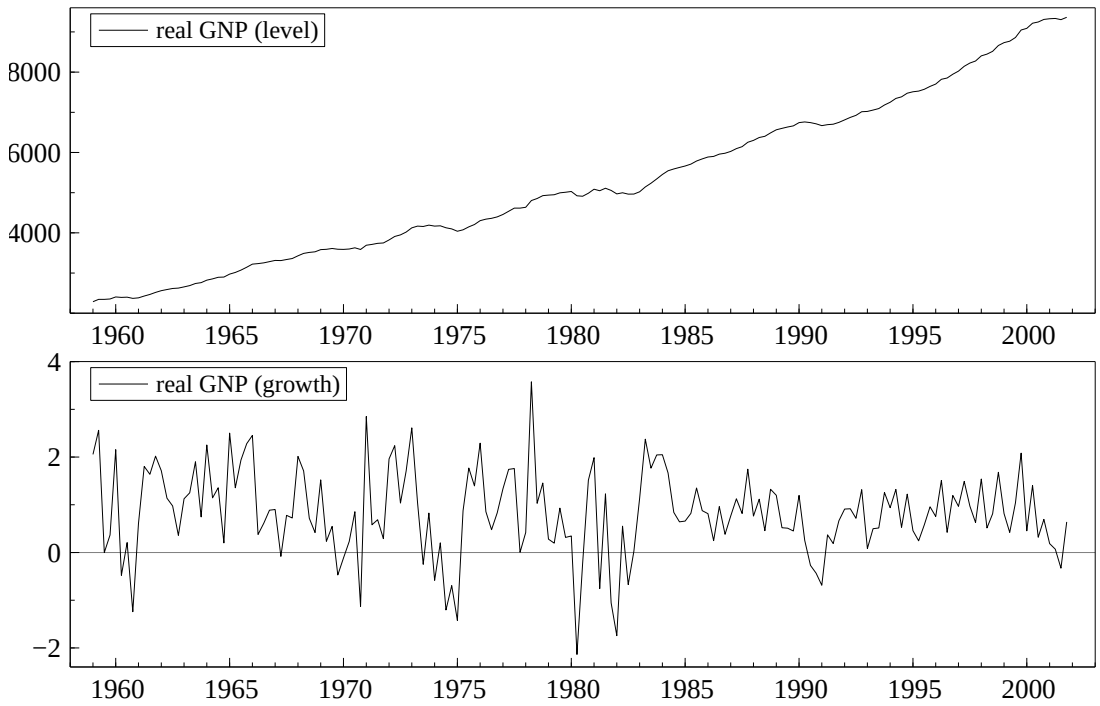

Figure 5.10: Real GNP of the United States in billions of dollars (above), and its quarterly growth rate in percents (below).

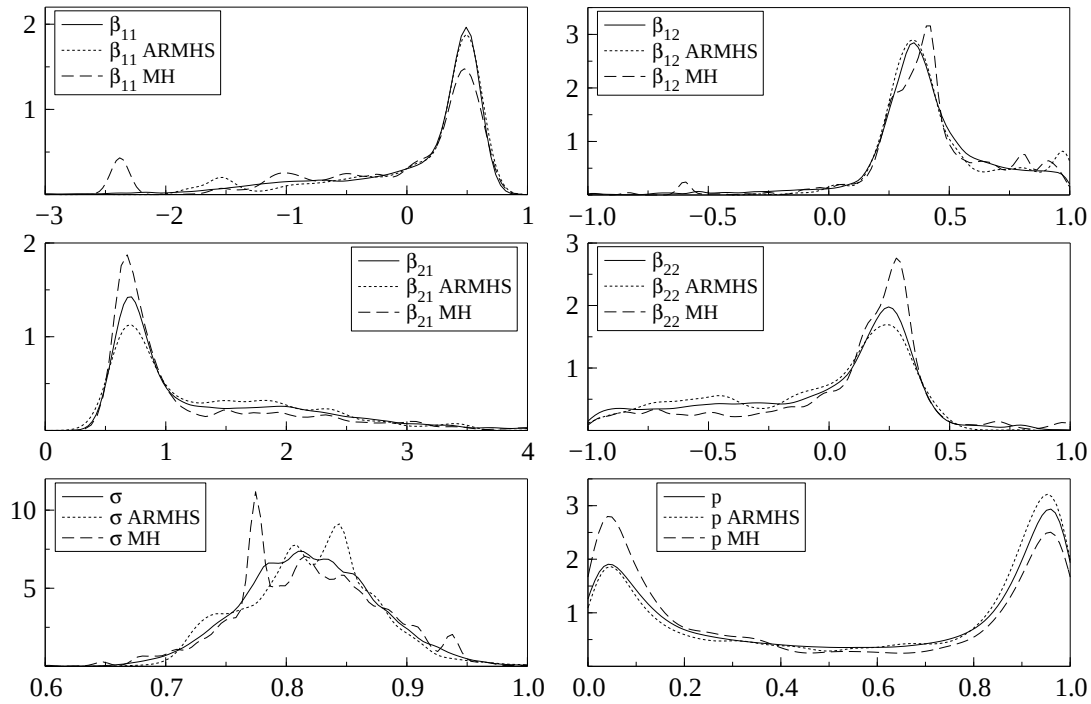

Figure 5.11: Marginal densities for the two-regime mixture model for US real GNP.

Table 5.3: Sampling results for the two-regime mixture model for US real GNP.

	bounds ARDS		initialization		ARMHS		ARIS		MH		IS		large sample	
	min.	max.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
β_{11}	-4.00	4.00	0.00	0.50	0.11	0.64	0.10	0.59	-0.14	0.88	0.01	0.72	0.07	0.70
β_{12}	-1.00	1.00	0.00	0.50	0.45	0.24	0.45	0.25	0.42	0.28	0.40	0.28	0.41	0.27
β_{21}	-4.00	4.00	1.00	0.50	1.32	0.74	1.27	0.78	1.22	0.83	1.28	0.85	1.30	0.79
β_{22}	-1.00	1.00	0.00	0.50	-0.07	0.39	-0.02	0.38	0.05	0.39	0.01	0.40	-0.04	0.41
σ	0.00	2.00	1.00	0.50	0.82	0.05	0.82	0.06	0.82	0.06	0.82	0.06	0.82	0.06
p	0.00	1.00	0.50	0.20	0.59	0.38	0.53	0.38	0.48	0.39	0.52	0.39	0.55	0.38
drawings per iteration ($\eta \times \rho$)					5000×5		5000×5		250000		250000			
number of iterations					8		5		8		8			
average time per iteration (in s)					81.4		82.9		89.6		85.1			
Mahalanobis distance					0.04		0.02		0.20		0.15			
acceptance rate (in %)					17.6				1.2					
5% most influential weights (in %)							57.9				99.7			

The parameter estimates are reported in Table 5.3, together with the corresponding large sample values (computed from 250000 ARMHS drawings). Furthermore, Figure 5.11 shows the large sample marginal densities, and the densities estimated from 25000 ARMHS drawings and from 250000 MH drawings. The large sample densities (obtained from 250000 ARMHS drawings) are smooth, while this is not always the case for the "small sample" densities. However, we note that the "small sample" MH densities have been constructed from the same number of drawings as the smooth large sample ARMHS densities. In general, the ARMHS, ARIS and importance sampling estimates are close to each other, but the MH estimates are sometimes quite different. In particular, this is illustrated in the density of the mixture probability p. MH overestimates the left mode and underestimates the right mode. The average computing times per sampling round are again comparable for the four considered algorithms, but only ARIS has "converged" within 8 rounds, although the Mahalanobis distance for ARMHS is also small. The acceptance rates (ARMHS and MH) and the total weight of the 5% most influential drawings (ARIS and importance sampling) again provide support for the ARDS algorithms.

5.5 Conclusions

In this chapter, we have provided an overview of some frequently used integration methods. These methods can be applied to get insight into the posterior density of model parameters. We have also paid attention to the limitations of current algorithms, and we have proposed a new methodology which extends current methods. In particular, we have focussed on extending the Metropolis-Hastings [MH] and importance sampling methods by applying a radial transformation to the parameter space of the posterior, or target, density. In our adaptive radial-based direction sampling [ARDS] algorithms, sampling does not take place in the m-dimensional parameter space directly, but in an $(m-1)$-dimensional subspace of directions. The final dimension, which corresponds to a distance measure, is drawn from the target density (conditional on the directions) using the inversion method. In this way, the shape of the posterior density is perfectly taken into account along the sampled directions. For a given number of drawings, this approach requires more function evaluations of the posterior density than a traditional MH or importance sampling algorithm. The usual type of tradeoff occurs, that is, with a more sophisticated algorithm one hopes to get "correct" results with less drawings than with a less sophisticated algorithm. It may also happen that a simple method cannot deliver reliable results. It would however be surprising when ARDS cannot provide good results while the simpler, computationally less intensive, methods can. This is confirmed by three illustrations. Moreover, one can use the ARDS algorithms as a preliminary step in which the posterior distribution is explored to prepare for a method more tailored to the problem being studied. Our methods can be characterized as black-box algorithms, see Chen *et al.* (2000). We emphasize that there is no claim that the ARDS algorithms are superior in theory to other kinds of algorithms. Such a claim would not make sense.

For any model/data combination, a sufficient research effort might result in a specific algorithm performing better than ARDS or other algorithms. However, this is not guaranteed, and the effort required may not be rewarding. An interesting extension of the discussed methodology would be to incorporate ARMHS in a Gibbs algorithm, where a subset of the parameters can be sampled directly, while the remaining parameters cannot.

5.A Derivations

In this appendix, we derive the conditional density $p(\rho|\eta)$, the weight function $w(\eta)$ and the acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$. First, it follows from (5.38) and (5.39) that the target density $p(x)$ in terms of ρ and η is given by

$$\begin{aligned} p(\rho,\eta) &= p(\, x(\rho,\eta|\,\mu,\Sigma)\,)\,|J_x(\rho,\eta)| \\ &= p(\, x(\rho,\eta|\,\mu,\Sigma)\,)\,|J_y(\rho)|\,J_y(\eta)\det(\Sigma^{1/2}), \end{aligned} \tag{5.52}$$

implying that

$$\begin{aligned} p(\rho|\eta) &\propto p(\rho,\eta) \\ &= p(\, x(\rho,\eta|\,\mu,\Sigma)\,)\,|J_y(\rho)|\,J_y(\eta)\det(\Sigma^{1/2}) \\ &\propto p(\, x(\rho,\eta|\,\mu,\Sigma)\,)\,|J_y(\rho)| \\ &\equiv \kappa(\rho|\,\eta), \end{aligned} \tag{5.53}$$

and implying that

$$\begin{aligned} p(\eta) &= \int_{-\infty}^{\infty} p(\rho,\eta)\,d\rho \\ &= \int_{-\infty}^{\infty} p(\, x(\rho,\eta|\,\mu,\Sigma)\,)\,|J_y(\rho)|\,J_y(\eta)\det(\Sigma^{1/2})\,d\rho \\ &\propto J_y(\eta)\int_{-\infty}^{\infty} p(\, x(\rho,\eta|\,\mu,\Sigma)\,)\,|J_y(\rho)|\,d\rho \\ &= J_y(\eta)\int_{-\infty}^{\infty} \kappa(\rho|\eta)\,d\rho. \end{aligned} \tag{5.54}$$

Similarly, it follows from (5.38) and (5.39) that the elliptically contoured candidate density $q(x) = c_m(\det(\Sigma))^{-1/2} f((x-\mu)'\Sigma^{-1}(x-\mu))$ in terms of ρ and η is given by

$$\begin{aligned} q(\rho,\eta) &= q(\, x(y(\rho,\eta)|\,\mu,\Sigma)\,)\,|J_x(\rho,\eta)| \\ &= c_m(\det(\Sigma))^{-1/2} f(\rho^2)\,|J_y(\rho)|\,J_y(\eta)\det(\Sigma^{1/2}) \\ &\propto f(\rho^2)\,|J_y(\rho)|\,J_y(\eta), \end{aligned} \tag{5.55}$$

so that

$$\begin{aligned} q(\eta) &= \int_{-\infty}^{\infty} q(\rho,\eta)\,d\rho \\ &\propto J_y(\eta)\int_{-\infty}^{\infty} f(\rho^2)\,|J_y(\rho)|\,d\rho \\ &\propto J_y(\eta). \end{aligned} \tag{5.56}$$

Note that (5.55) implies that ρ and η are independent. Finally, we obtain from (5.54) and (5.56) that

$$\begin{aligned} w(\eta) &= \frac{p(\eta)}{q(\eta)} \\ &\propto \int_{-\infty}^{\infty} \kappa(\rho|\eta)\, d\rho \\ &\equiv I(\eta) \end{aligned} \tag{5.57}$$

and

$$\begin{aligned} \alpha(\eta_{i-1}, \eta_i^*) &= \min\left\{ \frac{w(\eta_i^*)}{w(\eta_{i-1})}, 1 \right\} \\ &= \min\left\{ \frac{I(\eta_i^*)}{I(\eta_{i-1})}, 1 \right\}. \end{aligned} \tag{5.58}$$

Chapter 6

Summary and conclusions

In this thesis, we have discussed several models and methods with the general intention to "get more out of data". In the first part, we have considered the *modeling* component of this objective, where the focus is on marketing research, and in particular, on purchase behavior of households and on market shares of brands. We have introduced three new models to analyze household scanner panel data and store-level scanner data. In the corresponding analyses, we have adopted the classical approach and we have estimated the model parameters using maximum likelihood. Additionally, in the second part of the thesis, we have addressed the *algorithmic* component of Bayesian data analysis. We have discussed several sampling-based methods to investigate the posterior density of model parameters, and we have put forward a new methodology. In the current chapter, we conclude with a summary and we discuss our main findings.

In Chapter 2 and Chapter 3, we have presented two models for scanner panel data. Both models can be thought of as dealing with adjustment effects following a price promotion. Such adjustments imply that promotions do not only have an impact on current sales, but also affect future sales. Even if the deviation in sales relative to the non-promoted sales level is not permanent (as frequently reported in the literature), the existence of adjustment effects may have far-reaching implications for the profitability of promotions.

We have developed a utility-based model for purchase incidence, brand choice and purchase quantity in Chapter 2. The model assumes that purchase behavior of households is rational, and it accounts for the possibility that households consume their inventory stock faster when the inventory level is higher. Increased consumption after a price promotion (through a higher inventory level) would effectively increase demand in the product category, and would prevent that a large part of the current increase in sales comes at the expense of future sales. Hence, consumption acceleration amounts to a positive adjustment effect. It offers a possible explanation for the frequently reported lack of a postpromotion dip in store-level scanner data. For the yogurt category, we find that the speed of consumption strongly depends on the inventory level, which is consistent with other studies in the literature. However, a simulation study shows that the increase in consumption is not strong enough to prevent a small dip in sales immediately after a

price promotion. The dip disappears after one week, and it is about one percent of the current sales effect.

In Chapter 3, we have proposed a new reference price model for brand choice. This model takes into account that households may use prices experienced in the past to form comparison prices for the current brand choice decision. The unique feature of our model is that it allows that households may *not always* be able to recall, and hence use, past price information. Current brand choice models either do not account for reference effects, or they assume that households always have sufficient price knowledge to form a reference price. As internal reference price formation would imply a potentially strong negative adjustment after a price promotion, it is important to have a good understanding of how reference prices are formed from past price experience. This includes questions such as (i) what is the size of the reference effect?, (ii) what is the duration of the reference effect?, and (iii) how often are households able to construct a reference price? To address these questions, we have explicitly modeled the price recall process of households by assuming that unobserved price recall evolves over time according to a first-order Markov process in which the "forgetting state" is absorbing. The resulting reference price specification is flexible, allows for uncertainty in the underlying process, allows that households may forget past prices, and even allows that households cannot construct a reference price at all.

Our main findings for the catsup category are as follows. First, a price supported by a feature/display has a higher probability to be recalled than a price which is not supported by such promotional activity. Second, our model does not indicate a response asymmetry concerning perceived price gains and losses. This runs counter to an empirical generalization in the literature, see Kalyanaram and Winer (1995). However, our results also suggest that the impact of internal reference price formation on brand choice might be even larger than is currently believed. Third, we find that price recall by households is very limited. Nonetheless, a simulation shows that the initial gain in brand sales due to a temporary price reduction is largely offset by a decrease in subsequent sales.

In Chapter 4, we have discussed a Markov model for market shares. The proposed model can be used to infer share-switching across brands when only store-level data are available. This provides new insights which cannot be obtained from the frequently considered market share attraction model. Additionally, the model structure allows for a decomposition of own and cross price elasticities. This decomposition can be used to analyze which part of the reaction of the own market share to an own or a competitive price reduction is caused by changed customer retention, and which part is not.

Application of our model to four store-level scanner data sets has revealed some strong asymmetries in share-switching across brands. Nonetheless, the resulting market shares seem balanced in the sense that for each brand the total share coming in and the total share going out are approximately equal. Furthermore, we find that the retention effect is usually not the dominant factor in the reaction of market shares to price changes. Interestingly, the estimated price elasticities violate the frequently reproduced finding of

Blattberg and Wisniewski (1989) that higher-price higher-quality [HPHQ] brands draw more share from lower-price lower-quality [LPLQ] brands by offering a price reduction than vice versa. Still, our results appear to be consistent with a refinement suggested by Bronnenberg and Wathieu (1996). This extension states that it is necessary that the quality gap between an HPHQ brand and an LPLQ brand is sufficiently large to account for the price gap.

Finally, in Chapter 5, we have considered Monte Carlo techniques which can be used in Bayesian analysis to get insight into the posterior density of model parameters. The basic idea behind the Monte Carlo approach is that an intractable posterior density can be investigated by generating a sample from it, and using the drawings to make inference. We have provided an overview of some frequently applied methods, and we have put forward a new and flexible methodology. Our algorithms extend current methods and should often result in more accurate and more reliable samples. They can often be applied successfully without a preliminary study of the shape of the posterior density. The good performance of the methods has been illustrated by three examples. Nonetheless, more research would be needed to fully validate our claim that the proposed methodology can be used effectively in a broad area of applications.

Nederlandse samenvatting (Summary in Dutch)

In de afgelopen decennia is de beschikbaarheid van grote databases en de rekenkracht van computers, benodigd om deze data te analyseren, enorm toegenomen. Tegelijkertijd zijn nieuwe rekentechnieken en modellen ontwikkeld om data en rekenkracht te combineren. Dit alles heeft geleid tot nieuwe mogelijkheden om relevante onderzoeksvragen te beantwoorden. In dit proefschrift bouwen we voort op deze ontwikkelingen en introduceren we zowel nieuwe modellen als nieuwe rekentechnieken.

Kwantitatieve marketing modellen

Een vakgebied waarbinnen vaak met grote databases wordt gewerkt is de kwantitatieve marketing. In de afgelopen twintig jaar zijn veel wetenschappelijke artikelen verschenen over modellen en methoden om zogenaamde scanner panel data te analyseren met als doel nieuwe en betere inzichten te verkrijgen in het aankoopgedrag van huishoudens. Scanner panel data worden verzameld door de aankopen en winkelbezoeken van een groot aantal huishoudens te registreren gedurende een lange periode (doorgaans meerdere jaren). Er wordt beschreven wanneer huishoudens aankopen binnen een productcategorie hebben gedaan, welke merkkeuzes hierbij hoorden en wat de aankoophoeveelheden waren. Verder zijn voor ieder winkelbezoek van ieder huishouden in de database de prijzen en eventuele promoties van alle merken bekend. Door deze informatie te combineren kunnen aankoopbeslissingen van huishoudens worden gekoppeld aan de promotionele activiteiten van de verscheidene merken. Op die manier kunnen inzichten worden verkregen in bijvoorbeeld de effectiviteit van prijspromoties. In hoofdstuk 2 en hoofdstuk 3 van dit proefschrift worden twee nieuwe modellen besproken om scanner panel data te analyseren. In deze modellen wordt uitdrukkelijk rekening gehouden met de mogelijkheid dat promotionele activiteiten niet alleen het huidige verkoopniveau verhogen, maar ook nog kunnen doorwerken op toekomstige verkoopcijfers. Dergelijke dynamische effecten kunnen van grote invloed zijn op de winstgevendheid van prijspromoties.

Scanner panel data bevatten gedetailleerde informatie over het aankoopgedrag van huishoudens en bieden meer mogelijkheden voor analyse dan data op een hoger aggregatieniveau, zoals data op winkelniveau met de wekelijkse prijzen, promoties en verkoopcij-

fers van alle merken binnen een productcategorie. Het gebruik van data op het niveau van individuele huishoudens heeft echter ook een paar nadelen. Het eerste nadeel is kostentechnisch. Het is vaak duur (uitgedrukt in geld en/of inspanning) om scanner panel data te verkrijgen. Een tweede nadeel is de mogelijk beperkte representativiteit van dergelijke data. Dit zal afhangen van het aantal huishoudens in het panel en van de manier waarop de deelnemende huishoudens zijn geselecteerd. Data op winkelniveau zijn veel minder gedetailleerd dan data op het niveau van individuele huishoudens, maar hier staat tegenover dat ze representatiever zijn, ruimer voorradig zijn en goedkoper zijn om te verzamelen. Als aanvulling op de voorgaande twee hoofdstukken wordt in hoofdstuk 4 een model besproken voor data op winkelniveau. Ook dit model kan worden gebruikt om de effectiviteit van promoties te bepalen. We zullen de drie modellen (hoofdstukken 2, 3 en 4) hieronder kort bespreken.

Een model voor aankoopgedrag met flexibele consumptiesnelheid

Een aankoopbeslissing met betrekking tot een productcategorie, bijvoorbeeld yoghurt, kan worden opgedeeld in drie componenten. Ten eerste moet een huishouden bij winkelbezoek beslissen om het product wel of niet aan te schaffen. Deze beslissing zal bijvoorbeeld afhangen van de nog aanwezige voorraad thuis en van promotionele activiteiten binnen de productcategorie. Indien tot aankoop wordt overgegaan, dient het huishouden tevens een merkkeuze te maken en de aankoophoeveelheid te bepalen. Zowel merkkeuze als aankoophoeveelheid kunnen afhangen van de marketing-mix (prijzen, displays, etc.) van de verschillende merken. Verder kan de omvang van de aankoop ook worden beïnvloed door het voorraadniveau van het huishouden.

Een prijspromotie van een merk binnen een productcategorie kan van invloed zijn op zowel de wel-of-niet-kopen beslissing, merkkeuze als aankoophoeveelheid. Indien een huishouden door de promotionele activiteit wordt verleid een aankoop in de categorie te doen en/of een grotere hoeveelheid te kopen dan gebruikelijk, zal de promotie leiden tot een grotere voorraad bij het huishouden thuis. Een belangrijke vraag met betrekking tot de winstgevendheid van prijspromoties is of de toegenomen voorraad zal leiden tot een hogere consumptiesnelheid. Indien dit inderdaad het geval is, zullen prijskortingen de effectieve vraag naar het product vergroten. Echter, indien huishoudens hun consumptiesnelheid niet aanpassen, zal categoriepromotie slechts leiden tot eerdere en grotere aankopen ten koste van toekomstige aankopen. Met name voor bederfelijke producten die voor meerdere doeleinden kunnen worden gebruikt valt een voorraadafhankelijke consumptiesnelheid te verwachten. Toegenomen consumptie na prijspromotie is een mogelijke verklaring voor de vaak gerapporteerde afwezigheid van een zogenaamde postpromotiedip in scanner data op winkelniveau. Met andere woorden, toegenomen consumptie na een promotie zou ervoor kunnen zorgen dat het verkoopniveau onmiddellijk na de promotie niet of nauwelijks inzakt.

In hoofdstuk 2 modelleren we het aankoopgedrag van huishoudens. We integreren de drie aankoopbeslissingen (wel-of-niet-kopen, merkkeuze en aankoophoeveelheid) in één raamwerk waarin wordt uitgegaan van rationeel aankoopgedrag en waarin tevens wordt toegestaan dat huishoudens hun voorraad sneller consumeren wanneer het voorraadniveau hoger is. Wanneer we ons model toepassen op scanner panel data betreffende aankopen van yoghurt, vinden we dat de consumptiesnelheid sterk afhangt van het voorraadniveau. Een simulatiestudie toont echter aan dat de toename in consumptie niet sterk genoeg is om een kleine dip in verkopen onmiddellijk na een prijspromotie te voorkomen.

Een model voor merkkeuze met referentieprijzen

In hoofdstuk 3 concentreren we ons op de merkkeuzebeslissing van huishoudens. We introduceren een model dat rekening houdt met de mogelijkheid dat huishoudens in het verleden waargenomen prijzen gebruiken om referentiepunten te vormen waarmee de huidige merkprijzen worden vergeleken. Indien de huidige prijs van een merk lager ligt dan de geanticipeerde referentieprijs, zal dit worden ervaren als een meevaller. Hierdoor kan de aankoopkans van het betreffende merk toenemen. Het tegenovergestelde zal gelden wanneer de huidige prijs van het merk hoger ligt dan de corresponderende referentieprijs. De unieke eigenschap van het besproken model is dat het toestaat dat huishoudens niet altijd in staat zullen zijn om zich prijzen uit het verleden te herinneren en deze prijzen te gebruiken bij de merkkeuzebeslissing. Zulke beperkte prijskennis wordt gesuggereerd door verscheidene studies in de literatuur, die zijn gebaseerd op het observeren en interviewen van consumenten in supermarkten. Ons model kan worden gebruikt om inzicht te verkrijgen in de mate waarin prijzen uit het verleden worden herinnerd om referentieprijzen te vormen.

We passen ons referentieprijs model toe op scanner panel data betreffende aankopen van ketchup. De belangrijkste bevindingen zijn als volgt. Ten eerste heeft een prijs die wordt ondersteund door een “feature” of “display” een grotere kans later te worden herinnerd dan een prijs die niet door zulke promotionele activiteiten wordt ondersteund. Ten tweede suggereert ons model dat huishoudens niet gevoeliger zijn voor prijstegenvallers (met betrekking tot de referentieprijs) dan voor meevallers. Dit is een enigszins verrassend resultaat, omdat deze asymmetrie veelvuldig is gerapporteerd in de literatuur. Ook suggereren onze resultaten dat de invloed van referentieprijzen op merkkeuze mogelijk zelfs groter is dan tot op heden werd aangenomen. Dit is een belangrijke implicatie, omdat het onderschrijft dat veelvuldige prijskortingen een nadelige uitwerking hebben op toekomstige verkoopcijfers. Huishoudens zullen prijskortingen als normaal gaan ervaren en zullen het ontbreken van een prijspromotie als een tegenvaller gaan beschouwen. Dit kan het korte-termijn voordeel grotendeels teniet doen. In overeenstemming met de interviewstudies vinden we verder dat de prijskennis van huishoudens zeer beperkt is. Zo wordt de kans dat een huishouden zich prijzen uit het verleden kan herinneren en deze prijsinformatie gebruikt bij de merkkeuzebeslissing geschat op gemiddeld 31 procent. Tenslotte

toont een simulatiestudie aan dat de onmiddellijke toename in verkopen voortvloeiend uit een prijskorting inderdaad grotendeels teniet wordt gedaan door een negatief effect op toekomstige verkoopcijfers. Het netto effect van een prijskorting blijkt minder te zijn dan 40 procent van het onmiddellijke effect.

Een Markov model voor marktaandelen

In hoofdstuk 4 bespreken we een model dat de huidige marktaandelen van merken relateert aan de marktaandelen resulterend uit de vorige periode. Dit model kan worden gebruikt om inzicht te verkrijgen in de overgangen van marktaandelen van en naar ieder merk. Met andere woorden, het model kan worden gebruikt voor inferentie waarbij de vraag "welk merk pakt marktaandeel van welk merk?" centraal staat. Verder maakt het model een decompositie van prijselasticiteiten mogelijk. Deze decompositie kan worden gebruikt om de totale reactie van marktaandeel op een eigen of competitieve prijsverandering op te splitsen in een gedeelte toewijsbaar aan veranderde "customer retention" (eigen marktaandeel uit de vorige periode dat behouden blijft) en een resterend gedeelte. Dit alles leidt tot inzichten die niet kunnen worden verkregen aan de hand van het populaire attractiemodel voor marktaandelen.

We implementeren ons model voor vier scanner data sets op winkelniveau. Deze data betreffen de productcategieën ketchup en pindakaas op twee verschillende locaties in de Verenigde Staten. Voor drie van de vier data sets geeft ons model betere voorspellingen dan twee versies van het attractiemodel en een vector autoregressie. De geschatte overgangen van marktaandelen geven aan dat er een aantal sterke asymmetrieën bestaan tussen de beschouwde merken. Desalniettemin zijn de resulterende marktaandelen in balans in de zin dat voor ieder merk het binnenkomende marktaandeel en het uitgaande marktaandeel ongeveer gelijk aan elkaar zijn. Verder vinden we dat veranderde "customer retention" doorgaans niet de dominante factor is in de reactie van marktaandelen op prijsveranderingen. Een ander interessant resultaat is dat de geschatte prijselasticiteiten niet overeen komen met het in de literatuur veelvuldig bevestigde resultaat dat hoge-prijs hoge-kwaliteit [HPHK] merken meer marktaandeel kunnen overnemen van lage-prijs lage-kwaliteit [LPLK] merken via het geven van een prijskorting dan andersom. Onze prijselasticiteiten zijn echter wel consistent met een verfijning van dit asymmetrische prijscompetitie resultaat. De verfijning is dat het resultaat alleen geldt wanneer het verschil in kwaliteit tussen een HPHK merk en een LPLK merk "voldoende groot" is om het verschil in prijs te compenseren.

Monte Carlo integratie methoden

In het tweede gedeelte van dit proefschrift bespreken we een aantal rekentechnieken die in het bijzonder geschikt zijn voor Bayesiaanse analyse. Een belangrijk verschil tussen de klassieke benadering van econometrie/statistiek en de Bayesiaanse benadering ligt in

de interpretatie van de modelparameters. Vanuit klassiek oogpunt zijn de parameters van een model onbekende constanten met een zekere ware waarde. De parameterwaarden kunnen dan worden geschat met de methode van maximale aannemelijkheid, zoals we hebben gedaan in het eerste gedeelte van het proefschrift. De Bayesiaanse interpretatie is echter totaal verschillend. Vanuit een Bayesiaans perspectief zijn de modelparameters kansvariabelen waarbij een breed scala aan uitkomsten mogelijk is. Met andere woorden, een parameter heeft niet één ware waarde, maar wordt gekarakteriseerd door onzekerheid. Deze onzekerheid kan worden weergegeven door een kansverdeling. De kansdichtheid die wordt toegekend aan parameters is subjectief (kan van persoon tot persoon verschillen op basis van verschillende opvattingen en kennis) en kan worden bijgewerkt wanneer nieuwe gegevens (data) beschikbaar komen. Om tot bruikbare interpretatie van de resulterende kansverdeling te komen is men vaak geïnteresseerd in gemiddelden, spreidingsmaatstaven en kansen dat de parameters in een bepaald interval/gebied liggen. Monte Carlo methoden zijn zeer geschikt om dit soort grootheden uit te rekenen voor hoog-dimensionale en gecompliceerde kansdichtheden. Het basisidee achter deze simulatietechnieken is dat trekkingen uit een kansverdeling kunnen worden gebruikt om deze kansverdeling te analyseren. In hoofdstuk 5 bespreken we een aantal veelvuldig gebruikte Monte Carlo methoden. Verder introduceren we een nieuwe effectieve methodologie die voortbouwt op bestaande algoritmen. De voorgestelde techniek maakt gebruik van een transformatie naar radiaal-coördinaten.

Samenvattend, in dit proefschrift bespreken we nieuwe modellen, nieuwe resultaten, en nieuwe algoritmen. We hopen hiermee een waardevolle bijdrage te leveren aan zowel de econometrische literatuur als de marketing literatuur.

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.

Albert, J. H. and S. Chib (1993), Bayesian Analysis of Binary and Polychotomous Response Data, *Journal of the American Statistical Association*, **88**, 669–679.

Allenby, G. M. and P. E. Rossi (1991), Quality Perceptions and Asymmetric Switching between Brands, *Marketing Science*, **10**, 185–204.

Andrews, R. L. and I. S. Currim (2003), A Comparison of Segment Retention Criteria for Finite Mixture Logit Models, *Journal of Marketing Research*, **40**, 235–243.

Assuncão, J. L. and R. J. Meyer (1993), The Rational Effect of Price Promotions on Sales and Consumption, *Management Science*, **39**, 517–535.

Bauwens, L., C. S. Bos, H. K. Van Dijk, and R. D. Van Oest (2003), Explaining Adaptive Radial-Based Direction Sampling, in *Proceedings: Bayesian Statistical Science*, American Statistical Association, pp. 4353–4357.

Bauwens, L., C. S. Bos, H. K. Van Dijk, and R. D. Van Oest (2004), Adaptive Radial-Based Direction Sampling: Some Flexible and Robust Monte Carlo Integration Methods, *Journal of Econometrics*, **123**, 201–225.

Bauwens, L., M. Lubrano, and J. Richard (1999), *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press.

Belisle, C. J. P., H. E. Romeijn, and R. L. Smith (1993), Hit-and-Run Algorithms for Generating Multivariate Distributions, *Mathematics of Operations Research*, **18**, 255–266.

Bell, D. R., J. Chiang, and V. Padmanabhan (1999), The Decomposition of Promotional Response: An Empirical Generalization, *Marketing Science*, **18**, 504–526.

Bell, D. R., G. Iyer, and V. Padmanabhan (2002), Price Competition under Stockpiling and Flexible Consumption, *Journal of Marketing Research*, **39**, 292–303.

Bell, D. R., R. L. Keeney, and J. D. C. Little (1975), A Market Share Theorem, *Journal of Marketing Research*, **12**, 136–141.

Bell, D. R. and J. M. Lattin (2000), Looking for Loss Aversion in Scanner Panel Data: The Confounding Effect of Price Response Heterogeneity, *Marketing Science*, **19**, 185–200.

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

Berry, S., J. Levinsohn, and A. Pakes (1995), Automobile Prices in Market Equilibrium, *Econometrica*, **63**, 841–890.

Blattberg, R. C., R. A. Briesch, and E. J. Fox (1995), How Promotions Work, *Marketing Science*, **14**, G122–G132.

Blattberg, R. C. and K. J. Wisniewski (1989), Price-Induced Patterns of Competition, *Marketing Science*, **8**, 291–309.

Böckenholt, U. (1999), Mixed INAR(1) Poisson Regression Models: Analyzing Heterogeneity and Serial Dependencies in Longitudinal Count Data, *Journal of Econometrics*, **89**, 317–338.

Box, G. E. P. and M. E. Muller (1958), A Note on the Generation of Random Normal Deviates, *Annals of Mathematical Statistics*, **29**, 610–611.

Bozdogan, H. (1994), Mixture-Model Cluster Analysis using Model Selection Criteria and a New Informational Measure of Complexity, in H. Bozdogan (ed.), *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach*, Kluwer Academic Publishers.

Briesch, R. A., L. Krishnamurthi, T. Mazumdar, and S. P. Raj (1997), A Comparative Analysis of Reference Price Models, *Journal of Consumer Research*, **24**, 202–214.

Brodie, R. and C. A. De Kluyver (1984), Attraction versus Linear and Multiplicative Market Share Models: An Empirical Evaluation, *Journal of Marketing Research*, **11**, 194–201.

Bronnenberg, B. J. and L. Wathieu (1996), Asymmetric Promotion Effects and Brand Positioning, *Marketing Science*, **15**, 379–394.

Brooks, S. P. and G. O. Roberts (1998), Convergence Assessment Techniques for Markov Chain Monte Carlo, *Statistics and Computing*, **8**, 319–335.

Brownlee, K. A. (1965), *Statistical Theory and Methodology in Science and Engineering*, 2 edn., Wiley.

Bucklin, R. E. and S. Gupta (1992), Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach, *Journal of Marketing Research*, **29**, 201–215.

Bucklin, R. E., S. Gupta, and S. Siddarth (1998a), Determining Segmentation in Sales Response across Consumer Purchase Behaviors, *Journal of Marketing Research*, **35**, 189–197.

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., G. J. Russell, and V. Srinivasan (1998b), A Relationship Between Market Share Elasticities and Brand Switching Probabilities, *Journal of Marketing Research*, **35**, 99–113.

Casella, G. and E. I. George (1992), Explaining the Gibbs Sampler, *The American Statistician*, **46**, 167–174.

Chandon, P. and B. Wansink (2002), When Are Stockpiled Products Consumed Faster? A Convenience-Salience Framework of Postpurchase Consumption Incidence and Quantity, *Journal of Marketing Research*, **39**, 321–335.

Chang, K., S. Siddarth, and C. B. Weinberg (1999), The Impact of Heterogeneity in Purchase Timing and Price Responsiveness on Estimates of Sticker Shock Effects, *Marketing Science*, **18**, 178–192.

Che, H., P. B. Seetharaman, and K. Sudhir (2003), Pricing Behavior in Markets with State Dependence in Demand, working paper, Berkeley.

Chen, M. and B. W. Schmeiser (1996), General Hit-and-Run Monte Carlo Sampling for Evaluating Multidimensional Integrals, *Operations Research Letters*, **19**, 161–169.

Chen, M., Q. Shao, and J. G. Ibrahim (2000), *Monte Carlo Methods in Bayesian Computation*, Springer-Verlag.

Chen, Y., V. Kanetkar, and D. L. Weiss (1994), Forecasting Market Shares with Disaggregate or Pooled Data: A Comparison of Attraction Models, *International Journal of Forecasting*, **10**, 263–276.

Cheney, W. and D. Kincaid (1994), *Numerical Mathematics and Computing*, 3 edn., Brooks-Cole.

Chiang, J. (1991), A Simultaneous Approach to the Whether, What and How Much to Buy Questions, *Marketing Science*, **10**, 297–315.

Chib, S. (1992), Bayesian Inference in the Tobit Censored Regression Model, *Journal of Econometrics*, **51**, 79–99.

Chib, S. and E. Greenberg (1995), Understanding the Metropolis-Hastings Algorithm, *The American Statistician*, **49**, 327–335.

Chintagunta, P. K. (1993), Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households, *Marketing Science*, **12**, 184–208.

Chintagunta, P. K. (2000), A Flexible Aggregate Logit Demand Model, working paper, University of Chicago.

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.

Cooper, L. G. (1993), Market-Share Models, in J. Eliashberg and G. L. Lilien (eds.), *Handbooks in Operations Research and Management Science, Vol. 5*, Elsevier Science Publishers.

Cooper, L. G. and M. Nakanishi (1988), *Market Share Analysis: Evaluating Competitive Marketing Effectiveness*, Kluwer Academic Publishers.

Cowles, M. K. and B. P. Carlin (1996), Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review, *Journal of the American Statistical Association*, **91**, 883–904.

Damien, P., J. Wakefield, and S. Walker (1999), Gibbs Sampling for Bayesian Non-Conjugate and Hierarchical Models by using Auxiliary Variables, *Journal of the Royal Statistical Society B*, **61**, 331–344.

Dekimpe, M. G. and D. M. Hanssens (1995), Empirical Generalizations about Market Evolution and Stationarity, *Marketing Science*, **14**, G109–G121.

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, *Journal of the Royal Statistical Society B*, **39**, 1–38.

Dickson, P. R. and A. G. Sawyer (1990), The Price Knowledge and Search of Supermarket Shoppers, *Journal of Marketing*, **54**, 42–53.

Erdem, T., S. Imai, and M. P. Keane (2003), Brand and Quantity Choice Dynamics under Price Uncertainty, *Quantitative Marketing and Economics*, **1**, 5–64.

Erdem, T., G. E. Mayhew, and B. Sun (2001), Understanding Reference-Price Shoppers: A Within- and Cross-Category Analysis, *Journal of Marketing Research*, **38**, 445–457.

Fok, D., P. H. Franses, and R. Paap (2002), Econometric Analysis of the Market Share Attraction Model, in *Advances in Econometrics Vol. 16: Econometric Models in Marketing*, Elsevier Science Publishers.

Folkes, V. S., I. M. Martin, and K. Gupta (1993), When to Say When: Effects of Supply on Usage, *Journal of Consumer Research*, **20**, 467–477.

Frühwirth-Schnatter, S. (2001), Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching Models, *Journal of the American Statistical Association*, **96**, 194–209.

Gelfand, A. E., S. E. Hills, A. Racine-Poon, and A. F. M. Smith (1990), Illustration of Bayesian Inference in Normal Data Models using Gibbs Sampling, *Journal of the American Statistical Association*, **85**, 972–985.

Gelfand, A. E. and A. F. M. Smith (1990), Sampling-Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, **85**, 398–409.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995), *Bayesian Data Analysis*, Chapman and Hall.

Geman, S. and D. Geman (1984), Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Geweke, J. (1989), Bayesian Inference in Econometric Models using Monte Carlo Integration, *Econometrica*, **57**, 1317–1339.

Geweke, J. (1999), Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication, *Econometric Reviews*, **18**, 1–73.

Gilks, W. R. and G. O. Roberts (1996), Strategies for Improving MCMC, in W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (eds.), *Markov Chain Monte Carlo in Practice*, chap. 6, Chapman and Hall, pp. 89–114.

Gilks, W. R., G. O. Roberts, and E. I. George (1994), Adaptive Direction Sampling, *The Statistician*, **43**, 179–189.

Givens, G. H. and A. E. Raftery (1996), Local Adaptive Importance Sampling for Multivariate Densities with Strong Nonlinear Relationships, *Journal of the American Statistical Association*, **91**, 132–141.

Givon, M. (1984), Variety Seeking through Brand Switching, *Marketing Science*, **3**, 1–22.

Goldfeld, S. M. and R. E. Quandt (1973), A Markov Model for Switching Regressions, *Journal of Econometrics*, **1**, 3–16.

Greenleaf, E. A. (1995), The Impact of Reference Price Effects on the Profitability of Price Promotions, *Marketing Science*, **14**, 82–104.

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.

Gupta, S., P. K. Chintagunta, A. Kaul, and D. R. Wittink (1996), Do Household Scanner Data Provide Representative Inferences from Brand Choices: A Comparison with Store Data, *Journal of Marketing Research*, **33**, 383–398.

Hamilton, J. D. (1989), A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica*, **57**, 357–384.

Hamilton, J. D. (1990), Analysis of Time Series Subject to Changes in Regime, *Journal of Econometrics*, **45**, 39–70.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.

Hammersley, J. M. and D. C. Handscomb (1964), *Monte Carlo Methods*, 1 edn., Methuen, London.

Hanemann, W. M. (1984), Discrete/Continuous Models of Consumer Demand, *Econometrica*, **52**, 541–561.

Hardie, B. G. S., E. J. Johnson, and P. S. Fader (1993), Modeling Loss Aversion and Reference Dependence Effects on Brand Choice, *Marketing Science*, **12**, 378–394.

Hastings, W. K. (1970), Monte Carlo Sampling Methods using Markov Chains and Their Applications, *Biometrika*, **57**, 97–109.

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.

Helson, H. (1964), *Adaptation-Level Theory*, Harper and Row Publishers, New York.

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 and Economic Statistics*, **12**, 317–328.

Justel, A. and D. Peña (1996), Gibbs Sampling Will Fail in Outlier Problems with Strong Masking, *Journal of Computational and Graphical Statistics*, **5**, 176–189.

Kahneman, D. and A. Tversky (1979), Prospect Theory: An Analysis of Decision Under Risk, *Econometrica*, **47**, 263–291.

Kalyanaram, G. and J. D. C. Little (1994), An Empirical Analysis of Latitude of Price Acceptance in Consumer Package Goods, *Journal of Consumer Research*, **21**, 408–418.

Kalyanaram, G. and R. S. Winer (1995), Empirical Generalizations from Reference Price Research, *Marketing Science*, **14**, G161–G169.

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.

Kim, C.-J. (1994), Dynamic Linear Models with Markov-Switching, *Journal of Econometrics*, **60**, 1–22.

Kim, J., G. M. Allenby, and P. E. Rossi (2002), Modeling Consumer Demand for Variety, *Marketing Science*, **21**, 229–250.

Kloek, T. and H. K. Van Dijk (1978), Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo, *Econometrica*, **46**, 1–20.

Koop, G. (2003), *Bayesian Econometrics*, Wiley.

Kopalle, P. K., A. G. Rao, and J. L. Assuncão (1996), Asymmetric Reference Price Effects and Dynamic Pricing Policies, *Marketing Science*, **15**, 60–85.

Krishnamurthi, L., T. Mazumdar, and S. P. Raj (1992), Asymmetric Response to Price in Consumer Brand Choice and Purchase Quantity Decisions, *Journal of Consumer Research*, **19**, 387–400.

Krishnamurthi, L. and S. P. Raj (1988), A Model of Brand Choice and Purchase Quantity Price Sensitivities, *Marketing Science*, **7**, 1–20.

Kumar, V. and T. B. Heath (1990), A Comparative Study of Market Share Models using Disaggregate Data, *International Journal of Forecasting*, **6**, 163–174.

Lal, R. and V. Padmanabhan (1995), Competitive Response and Equilibria, *Marketing Science*, **14**, G101–G108.

Lattin, J. M. and R. E. Bucklin (1989), Reference Effects of Price and Promotion on Brand Choice Behavior, *Journal of Marketing Research*, **26**, 299–310.

Lattin, J. M. and L. McAlister (1985), Using a Variety-Seeking Model to Identify Substitute and Complementary Relationships among Competing Products, *Journal of Marketing Research*, **22**, 330–339.

Law, A. M. and W. D. Kelton (1991), *Simulation Modeling and Analysis*, 2 edn., McGraw-Hill.

Leeflang, P. S. H. (1974), *Mathematical Models in Marketing*, Stenfert Kroese, Leiden.

Leeflang, P. S. H., D. R. Wittink, M. Wedel, and P. A. Naert (2000), *Building Models for Marketing Decisions*, Kluwer Academic Publishers.

Mayhew, G. E. and R. S. Winer (1992), Analysis of Internal and External Reference Prices using Scanner Data, *Journal of Consumer Research*, **19**, 62–70.

Mazumdar, T. and P. Papatla (2000), An Investigation of Reference Price Segments, *Journal of Marketing Research*, **37**, 246–258.

McAlister, L. and E. Pessemier (1982), Variety Seeking Behavior: An Interdisciplinary Review, *Journal of Consumer Research*, **9**, 311–322.

McCulloch, R. E., N. G. Polson, and P. E. Rossi (2000), A Bayesian Analysis of the Multinomial Probit Model with Fully Identified Parameters, *Journal of Econometrics*, **99**, 173–193.

McCulloch, R. E. and P. E. Rossi (1994), An Exact Likelihood Analysis of the Multinomial Probit Model, *Journal of Econometrics*, **64**, 207–240.

McFadden, D. (1974), Conditional Logit Analysis of Qualitative Choice Behavior, in P. Zarembka (ed.), *Frontiers in Econometrics*, chap. 4, Academic Press.

McLachlan, G. J. and T. Krishnan (1997), *The EM Algorithm and Extensions*, Wiley.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of State Calculations by Fast Computing Machines, *The Journal of Chemical Physics*, **21**, 1087–1092.

Monahan, J. and A. Genz (1997), Spherical-Radial Integration Rules for Bayesian Computation, *Journal of the American Statistical Association*, **92**, 664–674.

Muirhead, R. (1982), *Aspects of Multivariate Statistical Theory*, Wiley.

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.

Neslin, S. A. and L. G. Schneider Stone (1996), Consumer Inventory Sensitivity and the Post-promotion Dip, *Marketing Letters*, **7**, 77–94.

Nijs, V. R., M. G. Dekimpe, and J. B. E. M. Steenkamp (2001), The Category-Demand Effects of Price Promotions, *Marketing Science*, **20**, 1–22.

Oh, M. S. and J. O. Berger (1992), Adaptive Importance Sampling in Monte Carlo Integration, *Journal of Statistical Computation and Simulation*, **41**, 143–168.

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.

Poirier, D. J. (1995), *Intermediate Statistics and Econometrics: A Comparative Approach*, The MIT Press.

Putler, D. S. (1992), Incorporating Reference Price Effects into a Theory of Consumer Choice, *Marketing Science*, **11**, 287–309.

Rajendran, K. N. and G. J. Tellis (1994), Contextual and Temporal Components of Reference Price, *Journal of Marketing*, **58**, 2–34.

Ramaswamy, V., E. W. Anderson, and W. S. DeSarbo (1994), A Disaggregate Negative Binomial Regression Procedure for Count Data Analysis, *Management Science*, **40**, 405–417.

Ritter, C. and M. A. Tanner (1992), Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler, *Journal of the American Statistical Association*, **87**, 861–868.

Roberts, G. O. and W. R. Gilks (1994), Convergence of Adaptive Direction Sampling, *Journal of Multivariate Analysis*, **49**, 287–298.

Romer, D. (1996), *Advanced Macroeconomics*, McGraw-Hill.

Ross, S. M. (1997), *Introduction to Probability Models*, 6 edn., Academic Press.

Rubin, D. B. (1987), Comment: A Noniterative Sampling/Importance Resampling Alternative to the Data Augmentation Algorithm for Creating a Few Imputations when Fractions of Missing Data are Modest: The SIR Algorithm, *Journal of the American Statistical Association*, **82**, 543–546.

Russell, G. J. and W. A. Kamakura (1994), Understanding Brand Competition using Micro and Macro Scanner Data, *Journal of Marketing Research*, **31**, 289–303.

Seetharaman, P. B. and P. K. Chintagunta (1998), A Model of Inertia and Variety-Seeking with Marketing Variables, *International Journal of Research in Marketing*, **15**, 1–17.

Sethuraman, R., V. Srinivasan, and D. Kim (1999), Asymmetric and Neighborhood Cross-Price Effects: Some Empirical Generalizations, *Marketing Science*, **18**, 23–41.

Smith, A. F. M. and G. O. Roberts (1993), Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods, *Journal of the Royal Statistical Society B*, **55**, 3–23.

Stoer, J. and R. Bulirsch (1993), *Introduction to Numerical Analysis*, 2 edn., Springer-Verlag.

Tanner, M. A. and W. H. Wong (1987), The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association*, **82**, 528–550.

Thaler, R. (1985), Mental Accounting and Consumer Choice, *Marketing Science*, **4**, 199–214.

Tierney, L. (1994), Markov Chains for Exploring Posterior Distributions, *The Annals of Statistics*, **22**, 1701–1762.

Van Dijk, H. K. and T. Kloek (1980), Further Experience in Bayesian Analysis using Monte Carlo Integration, *Journal of Econometrics*, **14**, 307–328.

Van Dijk, H. K., T. Kloek, and C. G. E. Boender (1985), Posterior Moments Computed by Mixed Integration, *Journal of Econometrics*, **29**, 3–18.

Van Heerde, H. J., S. Gupta, and D. R. Wittink (2003), Is 75% of the Sales Promotion Bump due to Brand Switching? No, Only 33% Is, *Journal of Marketing Research*, **40**, 481–491.

Van Oest, R. D. and P. H. Franses (2004), Which Brands Gain Share from Which Brands? Inference from Store-Level Scanner Data, *Quantitative Marketing and Economics*, forthcoming.

Van Oest, R. D., P. H. Franses, and R. Paap (2002a), A Dynamic Utility Maximization Model for Product Category Consumption, working paper TI 2002-097/4, Erasmus University Rotterdam.

Van Oest, R. D., L. F. Hoogerheide, and H. K. Van Dijk (2004), An Econometric Primer on Simulation-Based Bayesian Inference, in J. Geweke and H. K. Van Dijk (eds.), *Bayesian Models and Decision Making in Business and Economics: A Book of Readings*, Oxford University Press, to appear.

Van Oest, R. D. and R. Paap (2004), Analyzing the Effects of Past Prices on Reference Price Formation, Econometric Institute report 2004-36, Erasmus University Rotterdam.

Van Oest, R. D., R. Paap, and P. H. Franses (2002b), A Joint Framework for Category Purchase and Consumption Behavior, working paper TI 2002-124/4, Erasmus University Rotterdam.

Vanhuele, M. and X. Drèze (2002), Measuring the Price Knowledge Shoppers Bring to the Store, *Journal of Marketing*, **66**, 72–85.

Villas-Boas, J. M. and R. S. Winer (1999), Endogeneity in Brand Choice Models, *Management Science*, **45**, 1324–1338.

Wansink, B. and R. Deshpandé (1994), Out of Sight, Out of Mind: Pantry Stockpiling and Brand-Usage Frequency, *Marketing Letters*, **5**, 91–100.

Wedel, M., W. S. DeSarbo, J. R. Bult, and V. Ramaswamy (1993), A Latent Class Poisson Regression Model for Heterogeneous Count Data, *Journal of Applied Econometrics*, **8**, 397–411.

Wei, G. C. G. and M. A. Tanner (1990), Posterior Computations for Censored Regression Data, *Journal of the American Statistical Association*, **85**, 829–839.

Wheat, R. D. and D. G. Morrison (1990), Assessing Purchase Timing Models: Whether or Not Is Preferable to When, *Marketing Science*, **9**, 162–170.

Winer, R. S. (1986), A Reference Price Model of Brand Choice for Frequently Purchased Products, *Journal of Consumer Research*, **13**, 250–256.

Author index

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

297. J. LEVIN, *Essays in the economics of education.*

298. E. WIERSMA, *Non-financial performance measures: An empirical analysis of a change in a firm's performance measurement system.*

299. M. MEKONNEN AKALU, *Projects for shareholder value: A capital budgeting perspective.*

300. S. ROSSETTO, *Optimal timing of strategic financial decisions.*

301. P.W. VAN FOREEST, *Essays in financial economics.*

302. A. SIEGMANN, *Optimal financial decision making under loss averse preferences.*

303. A. VAN DER HORST, *Government interference in a dynamic economy.*

304. A.P. RUSSO, *The sustainable development of heritage cities and their regions: Analysis, policy, governance.*

305. G.A.W. GRIFFIOEN, *Technical analysis in financial markets.*

306. I.S. LAMMERS, *In conflict, een geschiedenis van kennismanagement.*

307. O.L. LISTES, *Stochastic programming approaches for strategic logistics problems.*

308. A.T. DE BLAEIJ, *The value of a statistical life in road safety.*

309. S.H.K. WUYTS, *Partner selection in business markets. A structural embeddedness perspective.*

310. H.C. DEKKER, *Control of inter-organizational relationships: The effects of appropriation concerns, coordination requirements and social embeddedness.*

311. I.V. OSSOKINA, *Environmental policy and environment-saving technologies. Economic aspects of policy making under uncertainty.*

312. D. BROUNEN, *Real estate securitization and corporate strategy: From bricks to bucks.*

313. J.M.P. DE KOK, *Human resource management within small and medium-sized enterprises.*

314. T. VERHAGEN, *Towards understanding online purchase behavior.*

315. R. HOEKSTRA, *Structural change of the physical economy. Decomposition analysis of physical and hybrid-units input-output tables.*

316. R.K. AIDIS, *By law and by custom: Factors affecting small and medium sized enterprises during the transition in Lithuania.*

317. S. MANZAN, *Essays in nonlinear economic dynamics.*

318. K. OLTMER, *Agricultural policy, land use and environmental effects: Studies in quantitative research synthesis.*

319. H.J. HORBEEK, *The elastic workfloor. About the implementation of internal flexibility arrangements.*

320. P. HOUWELING, *Empirical studies on credit markets.*

321. E. MENDYS, *Essays on network economics.*

322. J. NOAILLY, *Coevolutionary modeling for sustainable economic development.*

323. Y.V. VELD-MERKOULOVA, *Essays on futures markets and corporate spin-offs.*

324. F.J. WILLE, *Auditing using Bayesian decision analysis.*

325. D. HUISMAN, *Integrated and dynamic vehicle and crew scheduling.*

326. S. WANG, *Global climate change policies: An analysis of CDM policies with an adapted GTAP model.*

327. M.W. VAN GELDEREN, *Opportunity entry performance. Studies of entrepreneurship and small business.*

328. W. VAN WINDEN, *Essays on urban ICT policies.*

329. G.J. KULA, *Optimal retirement decision.*

330. R.J. IMESON, *Economic analysis and modeling of fisheries management in complex marine ecosystems.*

331. M. DEKKER, *Risk, resettlement and relations: Social security in rural Zimbabwe.*

332. A. MULATU, *Relative stringency of environmental regulation and international competitiveness.*

333. C.M. VAN VEELEN, *Survival of the fair: Modelling the evolution of altruism, fairness and morality.*

334. R. PHISALAPHONG, *The impact of economic integration programs on inward foreign direct investment.*

335. A.H. NÖTEBERG, *The medium matters: The impact of electronic communication media and evidence strength on belief revision during auditor-client inquiry.*

336. M. MASTROGIACOMO, *Retirement, expectations and realizations. Essays on the Netherlands and Italy.*

337. E. KENJOH, *Balancing work and family life in Japan and four European countries: Econometric analyses on mothers' employment and timing of maternity.*

338. A.H. BRUMMANS, *Adoption and diffusion of EDI in multilateral networks of organizations.*

339. K. STAAL, *Voting, public goods and violence.*

340. R.H.J. MOSCH, *The economic effects of trust. Theory and empirical evidence.*

341. F. ESCHENBACH, *The impact of banks and asset markets on economic growth and fiscal stability.*

342. D. LI, *On extreme value approximation to tails of distribution functions.*

343. S. VAN DER HOOG, *Disequilibrium dynamics.*

344. B. BRYS, *Tax-arbitrage in the Netherlands evaluation of the capital income tax reform of January 1, 2001.*

345. E. PRUZHANSKY, *Topics in game theory.*

346. P.D.M.L. CARDOSO, *The future of old-age pensions: Its implosion and explosion.*

347. C.J.H. BOSSINK, *To go or not to go...? International relocation willingness of dual-career couples.*