Regensburger
DISKUSSIONSBEITRÄGE
zur Wirtschaftswissenschaft

University of Regensburg Working Papers in Business,
Economics and Management Information Systems

A Model of Heterogeneous Multicategory Choice for Market Basket Analysis

Katrin Dippold[*], Harald Hruschka[**]

April 2011

Nr. 456

JEL Classification: C11, C35, C52, M31

Key Words: Marketing, market basket analysis, finite mixture model, variable selection, multivariate logistic regression, pseudo likelihood estimation, maximum likelihood approximation, multicategory purchase incidence models

[*] Katrin Dippold is a research specialist at GfK HealthCare, 90419 Nuremberg, Germany
[**] Harald Hruschka holds the Marketing Chair at the Department of Business Administration at the University of Regensburg, 93040 Regensburg, Germany, Phone: +49 941 943 2279, E-mail: harald.hruschka@wiwi.uni-regensburg.de

A Model of Heterogeneous Multicategory Choice for Market Basket Analysis

Abstract

Based on market basket data, multicategory purchase incidence models analyze demand interdependencies between product categories. We propose a finite mixture multivariate logit model to derive segment-specific intercategory effects of market basket purchase. Under the assumption that only a fraction of intercategory effects are significant, we exclude irrelevant effects by variable selection. This leads to a detailed description of consumers' shopping behavior that varies over segments not only w.r.t. parameters' values but also w.r.t. included interaction effects. We find that a homogeneous model would overestimate the intensity of interaction between product categories.

Keywords: Marketing, market basket analysis, finite mixture model, variable selection, multivariate logistic regression, pseudo likelihood estimation, maximum likelihood approximation, multicategory purchase incidence models

1 Introduction

Thanks to advanced data acquisition tools (e.g., scanner checkouts or loyalty card programs) and data management systems, retailers have increasingly access to detailed transaction data. These transaction data are collected in the form of "market baskets", where one basket contains all products that a single customer purchases during one shopping trip at one store. Scientists and practitioners have likewise realized that these data hold an enormous amount of information on customers and prompted research on tools that are able to tap this hidden knowledge. Comprehensively, these tools are called market basket analysis models.

Market basket models are based on the assumption that purchases in different categories are not necessarily independent from each other but may have reciprocal interaction effects. The purchase of product category A, for example, might increase the purchase probability of category B above the baseline probability expected under stochastic independence (and vice versa) if the joint presence of categories A and B in one market basket increases a basket's utility by more than the sum of individual category utilities. This intercategory effect implies complementarity. Product category A might also decrease the purchase probability of category C below the baseline probability (substitution), and might not have any effect on the purchase probability of category D (independence). These cross-category effects have diverse causes, e.g. budget competition, consumption complementarity or substitution, marketing-mix complementarity or substitution, one-stop-shopping, coincidence, etc. (e.g., Manchanda et al., 1999; Boztuğ and Hildebrandt, 2008; Niraj et al., 2008).

During the last decade, the development of market basket analysis models has seen substantial growth (e.g., Russell et al., 1999; Seetharaman et al., 2005; for a summary article, see Boztuğ and Silberhorn, 2006). Research can be divided into an exploratory and an explanatory stream (Mild and Reutterer, 2003). Exploratory models determine the general nature of the relationship between a large set of product categories (e.g., at a retailer) by compressing transaction data to purchase patterns. On the other hand, explanatory models aim at explaining and quantifying intercategory marketing-mix effects for small subsets of product categories with disaggregated data (Boztuğ and Silberhorn, 2006).

Modeling the diversity of customers and segmenting consumers into homogeneous subgroups are focal problems in many marketing applications (Wedel and Kamakura, 1998). The development of appropriate tools is spurred by the inefficiency of treating a compound of diverse consumers as a homogeneous market and the consequent need of marketers to identify consumer segments for targeted marketing actions (Allenby and Rossi, 1999). For example, Abramson et al. (2000) find that the neglect of heterogeneity with regard to preferences and market response results in biased parameter estimates for a MNL model. Consumer heterogeneity can be modeled as stemming from observable (e.g., demographics or shopping behavior) or latent sources.

For market basket models, too, the need to implement heterogeneity has been recognized. Boztuğ and Reutterer (2008, p.295) stress that "the analytical focus for studying cross-category dependencies (...) needs to be shifted to a more disaggregate (i.e. individual or customer segment) level". Manchanda et al. (1999, p.98) argue that "[a]ccounting for heterogeneity may be crucial in understanding the true nature of the association across categories". Next to reproducing only a small number of product categories, treating customers as homogeneous is a pivotal source of bias in parameter estimation.

1.1 Literature Review on Multicategory Purchase Incidence Models

Although research in explanatory market basket analysis may include several steps of consumer decision making, i.e., purchase incidence, brand choice, and purchase quantity, our work focuses on multicategory purchase incidence, on "whether to buy" models (Seetharaman et al., 2005), as in the seminal paper of Russell and Petersen (2000; RP). The benefit of their multivariate logit (MVL) model is the translation of a joint basket purchase decision into a consistent set of conditional category purchase decisions (conditional choice specification). As such, demand interdependencies across product categories that are purchased together in one basket can be modeled and thoroughly analyzed. Even though derived differently, Hruschka (1991) as well as Hruschka et al. (1999) propose similar MVL models that determine category interactions based on aggregated joint purchase frequencies for category pairs and individual consumer baskets respectively.

These models disregard heterogeneity in purchase behavior and in reactions to marketing measures. Hruschka (1991) and Hruschka et al. (1999) as well as Dippold and Hruschka (2010) neither integrate heterogeneity in the model structure nor allow for customer-specific covariates. Likewise, Russell and Petersen (2000) do not account for latent heterogeneity. However, they include observable sources of heterogeneity as model covariates by letting customers differ in time since last category purchase, category loyalty, and mean basket size. The study by Boztuğ and Hildebrandt (2008) which is a slightly varied replication of the RP model uses the same covariates. But heterogeneity over market baskets of consumers, more specifically over their purchase habits, preferences and marketing-mix sensitivities, is also caused by unobservable factors and has to be inferred from the data (Wedel and Kamakura, 1998). In this sense, heterogeneity is not (comprehensively) implemented in the models explained above.

An approach commonly used in market basket analysis is inferring heterogeneity over the covariation of error terms and estimating the model over a hierarchical Bayes structure with MCMC simulations. The substantial example in the context of multicategory purchase incidence is the multivariate probit (MVP) model of Manchanda et al. (1999) where both observed and unobserved heterogeneity are accounted for. As observable causes of heterogeneity, a sociodemographic and a shopping trip variable contribute to household-specific base utilities. For the unobserved part, the remaining covariation over customers that can not be explained by observed sources is modeled and interpreted with a variance components approach. Chib et al. (2002) allow for unobserved heterogeneity in a MVP model by adding household-specific constants and household-category-specific effects to the latent utility of a basket. In order to determine cross-category price sensitivities with a MVP model, Duvvuri et al. (2007) implement individual-level category constants whose values correlate over consumers via a covariance matrix and are influenced by consumer demographics.[1]

Boztuğ and Reutterer (2008) adopt a completely different approach to segmentation by combining a data compression step with prototype-wise MVL models. Heterogeneous shopping behavior is modeled according to the assumption that a household's purchases can be described as a mixture over generic market baskets. In the first step, generic basket prototypes including their most important product categories are determined by a vector quantization algorithm, an online K-means clustering method. For the determination of prototypes, each individual basket is assigned to the prototype that explains the shopping behavior best. For the second stage of the estimation, a 'complete' household with all baskets is assigned to the best fitting generic basket type thus achieving a discrete segmentation. Hence, consumer segments are interpreted as household allocations to generic basket types. Then, a prototype-specific MVL model is estimated based on the assigned households. A basket's latent utility is complemented with observable consumer-category-specific covariates on loyalty and time since last purchase. Note that these two estimation steps do not interact.

With regard to most previous models and especially the RP model, the low number of included product categories is striking. Russell and Petersen (2000) remark in their conclusion that their model is not applicable to a larger number of product categories. However, studies show that ignoring purchased categories significantly biases the estimates of parameters included in the model. Chib et al. (2002) find that the neglect of heterogeneity induces overestimated cross-category effects. In contrast, Dippold and Hruschka (2010) demonstrate that augmenting the number of included product categories increases the percentage of pairwise independent categories.

Some publications with logit models incorporate a higher number of product categories. Hruschka (1991) and Hruschka et al. (1999) used datasets with 72 and 73 product categories respectively. In order to reduce the parameters to estimate, both approaches applied model selection methods operating with a forward-backward scheme on univariate logit models. That is, for every product category model, it deletes interaction effects that are determined as insignificant according to test statistics. In a further step, these category models are combined to consistent joint basket models. Boztuğ and Reutterer (2008) circumvent the estimation of a complex MVL model with all 65 product categories in their data set by estimating a single MVL model for every prototype that contains only the 4 or 5 most important or "distinguished" (p.305) categories thus limiting the computational burden exogenously. For

[1] Other examples for this approach that treat different multicategory problems are probit models by Ainslie and Rossi (1998) and Seetharaman et al. (1999). For logit models, Hansen et al. (2006) and Singh et al. (2005) impose a factor analytic structure on the covariance matrix that enables the interpretation of factors as household "traits" while distinguishing observable and unobservable effects.

a data set with 30 product categories, Dippold and Hruschka (2010) apply Bayesian variable selection techniques that endogenously determine which category pairs are independent. During one iteration, only the significant interaction parameters are estimated to determine substitutive or complementary relations.

Table 1: Comparison of Multicategory Purchase Incidence Models

Logit

Publication	Observable Heterogeneity	Latent Heterogeneity	Categories
Hruschka (1991)	Disregarded	Disregarded	72
Hruschka et al. (1999)	Disregarded	Disregarded	73
Russell & Petersen (2000)	Shopping Behavior	Disregarded	4
Boztuğ & Hildebrandt (2008)	Shopping Behavior	Disregarded	5
Boztuğ & Reutterer (2008)	Demographic/Shopping Behavior	Vector Quantization	65
Dippold & Hruschka (2010)	Disregarded	Disregarded	30

Probit

Publication	Observable Heterogeneity	Latent Heterogeneity	Categories
Manchanda et al. (1999)	Demographic/Shopping Behavior	Covariance Matrix	4
Chib et al. (2002)	Disregarded	Covariance Matrix	12
Duvvuri et al. (2007)	Demographic/Shopping Behavior	Covariance Matrix	6

1.2 Our Approach

Our approach of market segmentation is a model based on data that are readily available to any retailer, observed purchases and marketing information. To account for multicategory purchase decisions, we use a MVL model, the prevailing approach in market basket analysis (e.g., Boztuğ and Reutterer, 2008; Hruschka et al., 1999; Russell and Petersen, 2000). In this respect, we understand our work as an extension of the seminal model by Russell and Petersen (2000).

As it was successfully applied to other logit models (e.g., Wedel et al., 1999; Andrews and Currim, 2002; Niraj et al., 2008; Song and Chintagunta, 2007), we propose a finite mixture model to describe purchase behavior that differs between consumer segments.[2] A large advantage of finite mixture models is the intuitive interpretation of the mixture components in the sense of market segments. For the purpose of segmentation, all available purchase information should be used. Therefore, our model explicitly allows for several basket purchases

[2]There has been a long discussion on the performance of finite vs. continuous mixture models (e.g., Wedel et al., 1999; for a short summary see Varki and Chintagunta, 2004). Whereas Allenby et al. (1998) and Allenby and Rossi (1999) argue that finite mixture models do not sufficiently represent consumer heterogeneity, especially when the number of segments is small and complete homogeneity within a segment is an unrealistic assumption, Andrews et al. (2002) do not find any performance superiority of the continuous over the finite mixture model. Even for a very limited number of segments (1 to 3), the continuous and the discrete model recover parameters and forecast holdout data equally well. Additionally, Wedel and Kamakura (1998) and Wedel et al. (1999) stress the consistency of the finite mixture model with the way management thinks about consumers in segments.

per individual. Hence, we do not segment single market baskets but households with differing numbers of baskets. We remark that our approach is in contrast to the idea of segmentation in Boztuğ and Reutterer (2008) where segment (= prototype) characteristics are determined on the basis of single baskets that are allocated to prototypes ignoring by which consumer they are purchased and which marketing measures had been present during purchase. The incorporation of marketing effects in our model (display and price reduction) is essential to separate segments with different reactions to marketing instruments.

We aim to include a number of product categories that is substantially larger than for the RP model who only investigated four categories. In order to handle the - with number of categories quadratically growing - amount of coefficients, we apply a Bayesian variable selection technique that limits the enormous number of intercategory relationships to the relevant interaction effects as in Dippold and Hruschka (2010).

In line with the RP model and Chib et al. (2002), we argue for the exclusion of sociodemographic covariates as observed sources of heterogeneity. First of all, the shopping baskets are often the only and always the most reliable and recent data that retailers have on their clients. Secondly and most importantly, the influence of sociodemographics on shopping behavior can be challenged because results of former research have been ambiguous. Generally, the possibility to derive marketing-mix sensitivities from demographic data to forecast purchase behavior is subject of a long discussion. Boatwright et al. (2004) identified an influence of demographic variables on promotional response, Hoch et al. (1995) a strong connection between demographics and price sensitivity. In contrast, other authors discovered only weak influence on marketing-mix sensitivities (e.g., Chintagunta and Gupta, 1994; Rossi et al., 1996). Similarly, Kim et al. (1999) find that on an individual level shopping pattern variables reveal more information on price elasticity than demographic data do.

To summarize, we find that there is no model we are aware of that determines significant interdependencies over a large number of product categories endogenously and additionally accounts for heterogeneity. Therefore, we propose a multivariate finite mixture logit model that can handle a large amount of parameters due to variable selection and fast calculation algorithms and identify segments at the same time, thus giving a precise picture of interaction effects between categories (section 2). The application to a data set is shown in section 3, before the article closes with a conclusion and remarks on future research (section 4).

2 Model

2.1 Finite Mixture Multivariate Logit Model (FM-MVL)

In the following, we introduce the finite mixture (FM) approach, the multivariate logit model and variable selection as substantial components of our model. As a widely-used approach in statistics, FM modeling is a natural way to represent a heterogeneous population that can be divided into K homogeneous components (Frühwirth-Schnatter, 2006). These subgroups are either known a priori or introduced to achieve a more flexible adjustment to the data while modeling a heterogeneous group. For the second case, the number of components K can vary from 1 to the number of observations I (McLachlan and Peel, 2000). Conceptually, it is assumed that every individual belongs to exactly one unknown component that has to be determined from the data. But as the data do not suffice to make unique assignments, only individual probabilities to belong to a component can be estimated (Wedel and Kamakura, 1998).

Each component $k = 1, ..., K$ has a non-negative weight, probability or mixing proportion π_k with $\sum_k \pi_k = 1$. The joint density of any random variable Y, such as purchase incidences of product categories over market baskets, and a component k is

$$P(Y, k) = P(Y|k)\pi_k \tag{1}$$

with Y following the distribution $P(Y|k) = P(Y|\beta_k)$ conditional on k with β_k as a vector of component-specific parameters. If the segment[3] indicator k is not observable, the marginal (unconditional) distribution of Y is derived as the mixture

$$P(Y) = \sum_k P(Y, k). \tag{2}$$

We introduce the following notation: a specific market basket i with $i = 1, ..., I$ contains binary purchase incidences $Y_i = [Y_{i1}, ..., Y_{iJ}]$ with $j = 1, ..., J$ indicating product categories. In other words, $Y_{ij} = 1$ denotes purchase of category j in basket i, otherwise $Y_{ij} = 0$.

In accordance with all earlier applications of logit models for consumer decision modeling (e.g., Guadagni and Little, 1983; McFadden, 1974), the MVL models assumes a utility-maximizing customer. As in McFadden's Random Utility Framework (1974), basket and category purchase probabilities are derived from the utility that a specific purchase decision delivers to the consumer. In detail, the deterministic part of the utility $V(Y_i)$ of a market basket i is formulated as

$$V(Y_i) = \sum_j (\alpha_j + \gamma_j DISP_{ij} + \xi_j PRED_{ij})Y_{ij} + \sum_{j<l} \theta_{jl} Y_{ij} Y_{il}.$$

Category constants α_j stand for the intrinsic utility of a product category j, whereas γ_j and ξ_j are category-specific marketing mix coefficients for display $DISP$ and price reduction $PRED$. The interaction effects θ_{jl} represent the relationship between categories j and l with $\theta_{jl} = 0$ denoting independence, $\theta_{jl} < 0$ substitution and $\theta_{jl} > 0$ complementarity. There is only an impact - additionally to the coefficients α_j, γ_j and ξ_j - on the utility of basket

[3]Segment is the marketing interpretation of a component in a finite mixture model. Therefore, the terms segment and component are used interchangeably in the text.

purchase if categories j and l are purchased together.[4] We define reflexive effects $\theta_{jj} = 0$ (e.g., Besag, 1977), as well as symmetric interdependencies, that is the influence of category j on category l is exactly the same as the influence of category l on category j, i.e. $\theta_{jl} = \theta_{lj}$ for $j \neq l$.

For a more detailed derivation of the MVL model in the context of market basket analysis, see Russell and Petersen (2000) or Boztuğ and Hildebrandt (2008). For the basic principles of MVL, refer to Besag (1974), Cox (1972), and Cressie (1993).

Combining FM segmentation and MVL modeling results in the intended FM-MVL model. As we assume heterogeneity over segments and homogeneity within segments, utility of a basket is determined with parameters specific to segment k

$$V(Y_i|k) = \sum_j (\alpha_{kj} + \gamma_{kj} DISP_{ij} + \xi_{kj} PRED_{ij}) Y_{ij} + \sum_{j<l} \theta_{kjl} Y_{ij} Y_{il}. \tag{3}$$

Accordingly, the basket purchase probability given segment k is

$$P(Y_i|k) = \frac{exp(V(Y_i|k))}{\sum_{Y^*} exp(V(Y^*|k))} \tag{4}$$

which is a multivariate logit model, with Y^* the set of all potential baskets.

This leads to the segment-specific conditional category purchase probabilities

$$P(Y_{ij}|Y_{il}, k) = \frac{exp(Y_{ij} V(Y_{ij} = 1|Y_{il}, k))}{1 + exp(V(Y_{ij} = 1|Y_{il}, k))} \tag{5}$$

with

$$V(Y_{ij} = 1|Y_{il}, k) = \alpha_{kj} + \gamma_{kj} DISP_{ij} + \xi_{kj} PRED_{ij} + \sum_{l \neq j} \theta_{kjl} Y_{il}. \tag{6}$$

The number of parameters to be initialized in this FM-MVL model ($npar = K[3 \cdot J + J(J-1)/2]$) is very large and considerably slows down the speed of calculation. However, only significant interactions influence utility and thus purchase probability. Previous studies prove that only a small percentage of interaction effects in market basket data indicate complementary or substitutive effects. Hruschka et al. (1999) show that purchases are independent for 95 percent of all category pairs in their data with over 70 categories. Dippold and Hruschka (2010) find 66 percent insignificant intercategory effects in a basket data set with 30 product categories.

[4] In contrast to probit models, there is no biasing effect of joint non-purchase that would be the most frequently occurring event. We also remark that we follow the cross-category effect definition by Hruschka et al. (1999). In contrast to Russell and Petersen (2000), cross-category effects do not depend on a household's typical basket size. This is justified because (1) the inclusion of basket size resulted only in a weak improvement of the LL value for holdout data in the RP model and because (2) our model already accounts for interaction effect variability by estimating different effects for different segments.

Here, we define the subset S_{kj} of other categories l that category j interacts with conditional on basket i belonging to segment k, i.e. $l \in S_{kj}$ if $\theta_{kjl} \neq 0$. To clarify, not only we allow for varying parameter values over segments but also for different subsets of included (i.e. significant) interaction effects. Given segment-specific sets of interacting parameters S_{kj}, we reformulate the deterministic segment-specific conditional category purchase utility as

$$V(Y_{ij} = 1 | Y_{il}, k) = \alpha_{kj} + \gamma_{kj} DISP_{ij} + \xi_{kj} PRED_{ij} + \sum_{l \in S_{kj}} \theta_{kjl} Y_{il}. \tag{7}$$

Our last assumption is that a household $h = 1, ..., H$ decides according to a utility function which is consistent and constant over all his purchases. Therefore, a household h with all his baskets is assigned to one segment k with probability

$$\varpi_{kh} = P(z_h = k | z_{\backslash h}, Y, \beta) \propto P(z_h = k | z_{\backslash h}) \prod_{i \in h} P(Y_i | \beta_k). \tag{8}$$

This probability depends on the assignment of other households $\backslash h$ to components, $z_{\backslash h}$, as well as on the parameters $\beta_k = (\alpha_k, \gamma_k, \xi_k, \theta_k)$ and on $P(Y_i|k) = P(Y_i|\beta_k)$, the likelihood of baskets $i \in h$ given segment k. Segment shares follow as $\pi_k = \sum_{h=1}^{H} \varpi_{kh}/H$.

2.2 Estimation

The standard approach to estimate parameters in a logit model is maximum likelihood (ML) with

$$ML(\beta) \equiv L(\beta) = \prod_i P(Y_i) \rightarrow max. \tag{9}$$

For market basket analysis, this would induce maximization over the joint basket purchase probabilities $P(Y_i)$ (ignoring segment indicators k) . This denominator, called normalizing constant or partition function,

$$Z(\beta) = \sum_{Y^*} exp(V(Y^*)) \tag{10}$$

is a sum over $|Y^*| = 2^J$ summands which causes enormous computational problems (Moon and Russell, 2004; Cressie, 1993) as it cannot be calculated efficiently (Carreira-Perpiñán and Hinton, 2005; Osindero et al., 2006) except for a comparatively small number of summands[5] (Hyvärinen 2006; Magnussen and Reeves, 2007; Wang et al., 2000). Consequently, the first derivative of the likelihood to be maximized - with respect to parameters - cannot be computed either. In the case of market basket analysis where the goal is to include as many

[5]With $J = 31$, Z has $2^{31} = 2,417,483,648$ elements, i.e. all possible market baskets. Huang and Ogata (2002) observe exponents between 9 and 15 to be the limit of computation.

product categories as offered in a supermarket, the solution to the ML problem is intractable.

It is also possible to work with the easier to calculate pseudo likelihood (PL; Besag, 1975) that is a heuristic approximation to the exact ML and avoids the calculation of the partition function (Cressie, 1993; Wang et al., 2000). The PL makes use of the fact that the basket purchase probabilities $P(Y_i|k)$ can be split into J conditional category purchase probabilities $P(Y_{ij}|Y_{il},k)$ and generally reads as

$$PL(\beta) = \prod_i \prod_j P(Y_{ij}|Y_{il};\beta) \tag{11}$$

which has to be maximized (Besag, 1974, 1975; Cressie, 1993). This decomposition is analogical to RP's "conditional choice specification" mentioned in section 1. To account for the K different segments, the pseudo likelihood components are not directly multiplied over market baskets. Instead, a weighted average per basket is calculated with the probabilities of a basket belonging to a segment as weights. The pseudo likelihood is

$$PL(\beta) = \prod_i (\sum_k \varpi_{ki} (\prod_j P(Y_{ij}|Y_{il};\beta_k))). \tag{12}$$

PL optimization has been used for various problems of spatial statistics (e.g., Moon and Russell, 2004) and provided sound estimates, also in the context of market basket analysis (e.g., Boztuğ and Hildebrandt, 2008). Sherman et al. (2006), for example, prove a good performance of PL estimation for autologistic functions that depends only weakly on the amount of correlation in the model, an important finding for market basket analysis where most product categories are independent but some are heavily interdependent.

Already in 1975, Besag (p.190) remarks that the downside of PL estimation is "that no sampling properties of the estimates are yet known". Although the PL parameter estimates are consistent with the ML estimates (Besag, 1975), they are usually not efficient (Besag, 1977; Cressie, 1993; Sherman et al., 2006; Wedel and Kamakura, 1998).[6] However, Hyvärinen (2006) showed in a simulation study that estimation errors of ML and PL were of similar size.

We propose to estimate the model parameters with PL optimization, where selected segment-specific parameters are sampled in a hybrid MC step (Neal, 1996; Shi et al., 2005). For details on the sampling of parameter values, please see the appendix.

2.3 Model Selection

So far, the question of how to find the optimal number of segments K has been ignored. In marketing literature, various methods to solve this problem - a crucial question because of

[6] The general disadvantage of wrong standard errors can be easily adjusted for as correct standard errors can be computed with bootstrapping (e.g., Efron and Tibshirani, 1998).

the direct impact on profit - have been proposed (e.g., Andrews and Currim, 2003; Nylund et al., 2007). Both over- and underestimation of segment number lead to inefficient marketing programs and misspecified product policy.

We decided against an endogenous determination of the number of segments in order to keep the estimation stable. Hence, the parameters are estimated for different numbers of segments in a first step. Afterwards, the best fitting model specification is chosen in accordance with a performance criterion.[7]

There is an ongoing debate about the appropriate criterion to use which has not been solved satisfactorily until today. The dominating approach is the application of information criteria (IC) that can have various specifications.[8] Generally, they consist of a model fit measure (traditionally, the log-likelihood LL) and a penalty value for model complexity (the number of estimated parameters) and have to be minimized. Well-established criteria are BIC (Schwarz, 1978), AIC (Akaike, 1974), or DIC (Spiegelhalter et al., 2002), the Bayesian version of AIC, which have been applied successfully in the context of finite mixture logit models. Unfortunately, they can not be calculated in our context because the log-likelihood value from maximum likelihood estimation can not be easily substituted with its analogon from pseudo likelihood estimation.

In order to render all estimations comparable, we use predictive model selection (Laud and Ibrahim, 1995; Carlin and Louis, 2000). The idea is "that good models, among those under consideration, should make predictions close to what has been observed " (Laud and Ibrahim, 1995, p.249). Hence, the model M_m is chosen which minimizes the expected value of a given discrepancy function $E[D(Y^s, Y|M_m)] = \frac{1}{S}\sum_s D(Y^s, Y|M_m)$ between observed data Y and sampled data sets Y^s $(s = 1, \dots, S)$ and penalizes in this way variability (and as such complexity) of a model. Laud and Ibrahim (1995) propose a loss criterion based on squared error. For our binary data, this corresponds to the absolute deviation.

$$E[D(Y^s, Y|M_m)] = \frac{1}{S}\sum_s\sum_i\sum_j(Y^s_{ij} - Y_{ij})^2 = \frac{1}{S}\sum_s\sum_i\sum_j|Y^s_{ij} - Y_{ij}| \tag{13}$$

The new data sets Y^s are sampled based on coefficient values saved during iterations while estimating model M_m. For every segment k, a saved iteration of parameter values β^s_k is randomly chosen. Given a basket's probability belonging to segment k conditional on β^s_k, a discrete assignment of Y_i and Y^s_i respectively to a segment k is sampled. β^s_k is used to update the original basket Y_i by 1000 Gibbs sampling steps to Y^s_i.

[7]This two-step approach is conventionally used in finite mixture models for multicategory choice (e.g., Song and Chintagunta, 2007).

[8]See, for example, Andrews and Currim (2002) for a complete tabulation including formulas for calculations.

2.4 Summary on Model Development

We aim at developing a heterogeneous MVL model. First of all, we estimate two simple benchmark models. Benchmark B1, the independence model, includes only main effects α_j which are computed as log odds. The second benchmark model B2 still assumes homogeneity ($K = 1$) and incorporates interaction effects reduced by variable selection. Next, we introduce our final model formulation with heterogeneity and marketing mix effects. The main model M is estimated optimizing PL for a varying K and includes a step of variable selection. Finally, we select the best model with predictive model selection.

We expect the following tendencies to show up during estimation:

E1: With a growing number of segments, PLL and D improve, especially if compared to the fit and selection criteria of a homogenous model.

E2: Less interaction effects are included in single segments of heterogeneous models than in the homogenous model.

E3: For the heterogeneous models, the segments differ w.r.t. included interaction effects.

E4: The heterogeneous estimation reveals more pronounced interaction effects than the homogeneous model.[9]

3 Empirical Study

3.1 Data Set

We use the IRI data set explained in detail by Bronnenberg et al. (2008) and Kruger and Pagni (2008) to test our model. The IRI data set contains information which product categories are bought by each household in several outlets. These household data are compiled to market baskets, i.e. a market basket is defined as the categories purchased by one household in one shop during one week.

We select a specific grocery store and use all available market baskets from 2001.[10] The resulting data set includes 1,794 households and 17,280 baskets, i.e., we have an average of 9.63 baskets bought per household. The average number of categories per basket is 3.52 (ranging between 1 and 17). The baskets include purchases in all 31 product categories provided in the original data. For absolute as well as relative purchase frequencies which vary

[9] We argue that the effects are not smoothed to medium strength over heterogeneous purchase behavior. Also the fact that less effects are estimated might contribute. A similar hypothesis was assumed by Boztuğ and Reutterer (2008) but motivated in a different way. Chib et al. (2002) proved the opposite effect.

[10] Our model specification does not include RP's category-specific HH variable with time since last category purchase ($TIME$) and loyalty ($LOYAL$). As this model is estimated over the purchases within one shop only neglecting purchases in other stores, we do not have complete information on a consumer's shopping history. Therefore, the values of $TIME$ and $LOYAL$ would not be meaningful. Besides, we already account for heterogeneity with the FM model and do not need auxiliary measures of consumer diversity.

between the two extremes RAZ (20 or 0.12 percent) and MIL (7,930 or 45.89 percent), see table 2. Histograms of the numbers of categories per basket as well as numbers of baskets per household are given in figure 1.

Table 2: Data Description

Name	Category Description	Freq.[a]	Name	Category Description	Freq.[a]
BEA	Beer & Ale	28	MIL	Milk	7930
BLA	Blades	279	MUK	Mustard & Ketchup	1321
CBV	Carbonated Beverages	5775	PAT	Paper Towels	2730
CIG	Cigarettes	176	PEB	Peanut Butter	1262
COF	Coffee	2189	PHO	Photography Supplies	78
COL	Cold Cereals	4830	RAZ	Razors	20
DEO	Deodorant	673	SAL	Salty Snacks	5930
DIA	Diapers	147	SHA	Shampoo	603
FAT	Facial Tissue	2156	SOU	Soup	3313
FRD	Frozen Dinners	1015	SPA	Spaghetti & Italian Sauce	2190
FRP	Frozen Pizza & Entrees	519	SUS	Sugar Substitutes	359
HHC	Household Cleaners	1277	TOT	Toilet Tissue	3466
HOD	Hot Dogs	1503	TOB	Toothbrush	324
LAD	Laundry Detergent	1766	TOP	Toothpaste	1196
MAB	Margarine & Butter	2699	YOG	Yoghurt	3825
MAY	Mayonnaise	1193			

[a] Absolute frequency of purchase incidence in all 17,280 market baskets.

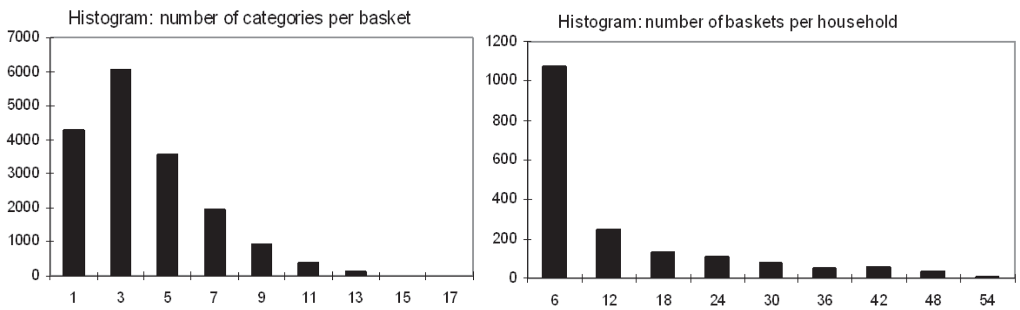

Figure 1: Description of IRI data

One advantage of the IRI data set is that it comprises information on the marketing actions run at the moment of purchase. For our analysis, we use price reduction and display. $PRED_{ij} = 1$ if a price reduction of at least 5 percent was given, otherwise $PRED_{ij} = 0$. $DISP_{ij} = 1$ if the purchased product was on display, otherwise $DISP_{ij} = 0$. For both mix measures, this information is available in 20 of 31 categories. Besides, the IRI data set includes demographic (e.g., age, family size) and socioeconomic (e.g., income, education, occupation) variables on the households in categorical form.

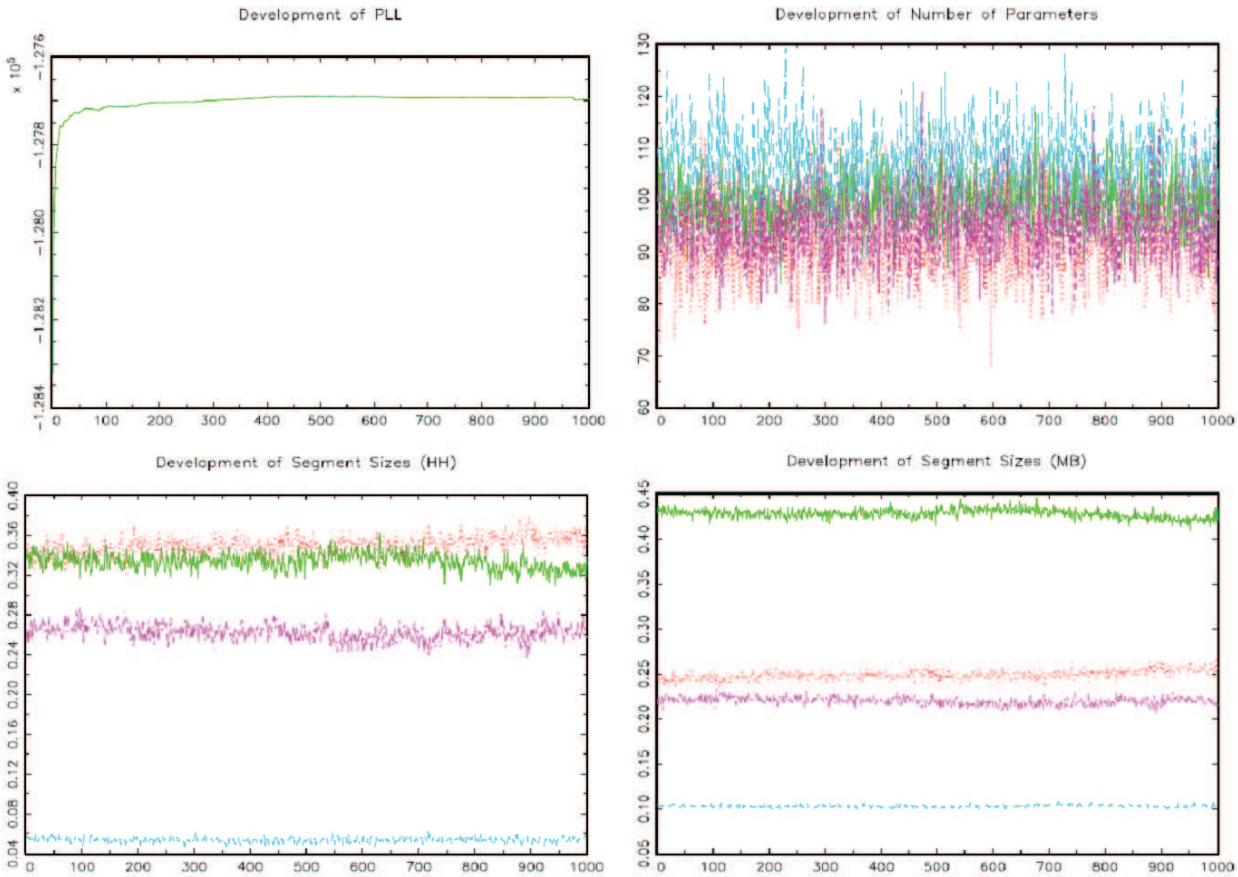

Figure 2: PLL value, number of parameters, and segment sizes for $M(K = 4)$

3.2 Results and Discussion

For the sake of practicability and interpretability, we limit the maximum number of segments to five. For every model, we use 2,000 burn-in runs followed by 1,000 iterations that are saved for parameter computation[11]. Parameters are calculated as means over the sequence of sampled values. Pseudo log-likelihood value PLL value and segment sizes stabilize quickly during estimation. The number of parameters also levels off in a range of approximately 20. Exemplarily, see figure 2 for the saved iterations of $M(K = 4)$.

Label switching, i.e. permutations of some component labels, is a known problem in mixture modeling (Celeux, 1998; Frühwirth-Schnatter, 2006; McLachlan and Peel, 2000). For our iterative process of parameter updating, we did not find any evidence of label switching, as can be seen in figure 2.

Results of the estimated models are summarized in various tables. Table 3 gives an overview of all measures calculated to compare the two types of estimation, especially the pseudo log-likelihood value PLL[12]. The PLL value for the independence model $B1$ is

[11] Test runs showed that the model and especially the household assignment stabilize quickly.
[12] For reasons of comparability, it is also calculated for models which approximate ML.

$-157,544.32$. $B2$, the homogeneous model with interactions but without marketing mix, achieves a PLL value of $-149,943.86$ with an average of 110.45 interactions.

Table 3: Model Calibration

	PLL	D
	Benchmarks	
$B1$	-157,544.32	94,113.89
$B2$	-149,943.86	93,735.75
	Main Model M	
$K=1$	-132,721.96	81,371.12
$K=2$	-129,717.60	79,233.55
$K=3$	-129,026.20	78,349.34
$K=4$	-127,696.47	77,367.14
$K=5$	-127,457.13	77,924.99

The introduction of marketing mix coefficients ($B2 \rightarrow M(K=1)$) results in a much higher increase of PLL than the introduction of different segments ($M(K=1) \rightarrow M(K>1)$). Still, fit can be improved by a model that explicitly allows for heterogeneity. A growing number of segments tends to enhance fit, as assumed in **E1**. Models with five segments already produce solutions with at least one very small segment. Therefore, a raise over five components does not seem reasonable. $M(K=4)$ is the best model according to D, the proposed model selection criterion.

Details on PL estimations are presented in table 4. For every model, we give the number of interaction effects in the different components.[13] The total number of parameters per segment is 31 constants + 40 marketing mix coefficients + individual number of interaction effects. We also provide relative component sizes w.r.t. households and market baskets. The upper line contains the percentage of households that were on average over all iterations allocated to one segment. The lower line contains the number of baskets that were on average allocated to one segment. As the number of baskets per household varies strongly over the households, these percentages apparently differ.

The larger part of interactions can be clearly deleted from the model with exclusion probabilities over 80 percent (see figure 3). Generally speaking, only a small percentage of interaction effects is determined as relevant. What is remarkable at first sight is that most models show a tendency to reveal two large segments which may be interpreted as 'normal' purchase patterns as well as at least one small component that contains households not fitting properly into other segments and so represents different types of 'abnormal' shopping. Consistently over different model dimensions, one of the large segments comprises many interaction effects, the other large segment contains only few interaction effects.

Returning to our tendency expectations, we look more detailed into which interaction

[13]We give only the number of interaction effects larger than .001.

Table 4: Result Details

	Main Model M				
	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
K=1					
# interactions	84.11				
K=2					
# interactions	5.63	70.62			
% of all households	53.12	46.88			
% of all baskets	40.34	59.66			
K=3					
# interactions	7.05	56.08	9.52		
% of all households	51.84	45.43	2.73		
% of all baskets	38.37	55.71	5.92		
K=4					
# interactions	41.19	15.79	3.81	18.34	
% of all households	33.05	5.30	35.90	25.75	
% of all baskets	42.11	10.32	24.77	21.80	
K=5					
# interactions	42.53	7.07	5.52	4.63	16.37
% of all households	26.87	46.15	14.21	1.73	11.04
% of all baskets	34.18	35.62	10.25	3.58	16.37

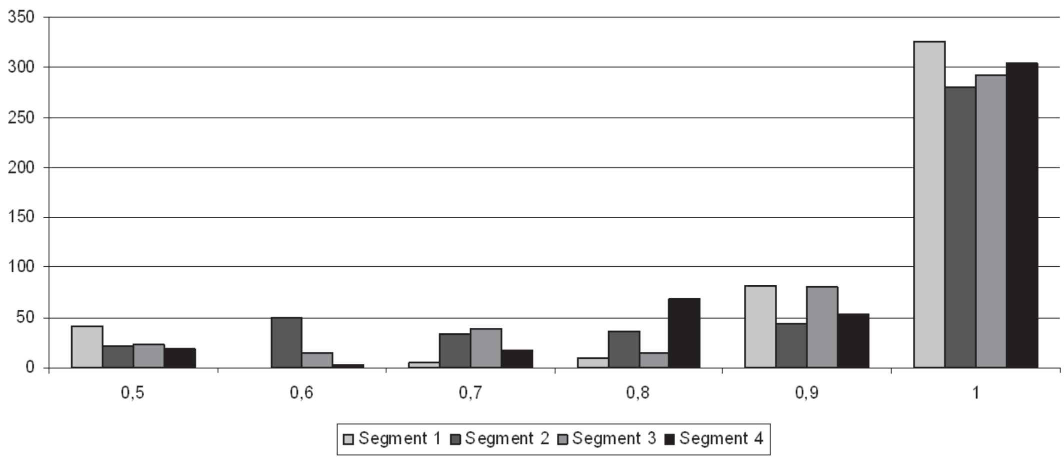

Figure 3: Histogram of Exclusion Probabilities of Interaction Effects in $M(K=4)$

effects are included. **E2** can be confirmed because the homogeneous segment of $M(K = 1)$ features more interaction effects than single segments of models $M(K > 1)$. This fact is depicted in figure 4, where black squares mirror exclusion probabilities of 0 percent, white squares reflect exclusion probabilities of 100 percent and the respective grey shades probabilities between 0 and 100 percent. The graphics show that the homogeneous version of M clearly overestimates the inclusion of interaction effects. For instance, the interactions BLA-DEO, CBV-SPA, and TOP-YOG appear in $M(K = 1)$ but in none of the four segments of $M(K = 4)$. In total, the homogeneous model estimates 34 interaction effects which are not relevant in any of the four segments of $M(K = 4)$. This underpins the importance of estimating a heterogeneous model because the homogeneous model would have led to completely different insights on how strongly product categories interact w.r.t. purchase probabilities.

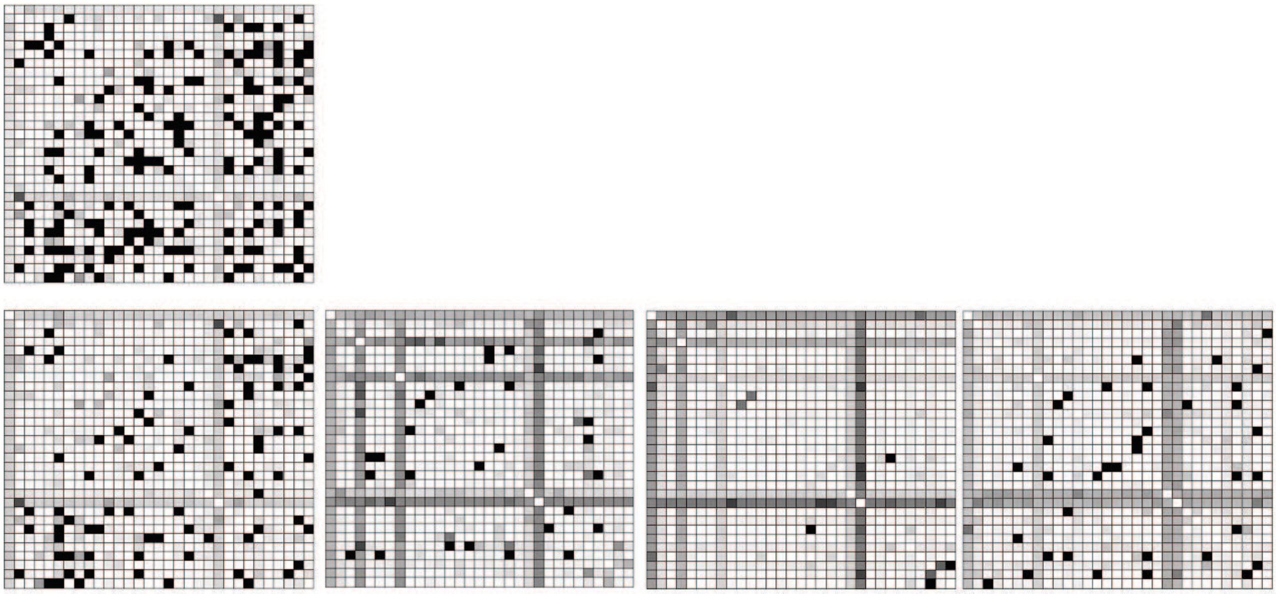

Figure 4: Exclusion Probabilities of Interaction Effects in $M(K = 1)$ vs. $M(K = 4)$

Considering **E3**, not only the quantity of included effects varies over components of one model but also the selection, although - with growing number of components - there may be interaction effects that are present in more than one segment. If the variable selection step had not been included into the mixture model but preceded it, the parameter estimates would have been biased. W.r.t. **E4**, no obvious trend is detected. The heterogeneous model neither reveals more nor less pronounced interactions.

Considering obvious complements-in-use, for instance, for $M(K = 4)$, we often find them to have significant positive interaction effects. COL-MIL, TOB-TOP, DEO-TOP, and FAT-TOT are complementary pairs in two of four segments. MAY-MUK as well as FAT-PAT are complements in three of four segments. Other complements-in-use such as the pair HOD-MUK are independent purchases, probably due to different purchase cycles. Also products that serve as substitutes-in-use, e.g. FRD-HOD, are complements-in-purchase in two of four segments.

Regarding marketing mix coefficients, both display and price reduction exert the expected positive influence on the purchase probability of a category. Only for categories with very low purchase frequencies, e.g. beer and ale, the model sometimes fails to produce the correct sign in all segments. The marketing mix effects differ in strength between segments. We do not find any segment that has consistently either smaller or larger coefficients than the other segments in the model.

3.3 Descriptive Segment Comparison

Like Andrews and Currim (2002), we examine in a subsequent step if segments differ w.r.t. their sociodemographic variables as well as their shopping behavior. For the sake of comparison (see table 5), we assigned every household to the one segment that it was most often sampled to during iterations. This analysis is applied to $M(K = 4)$, which was determined as the best model. The table shows the means of basket size and number of baskets per household as well as the most important interaction effects per segment (i.e. with the largest absolute values) in descending order where italics indicate a substitutive relation. The last lines of the shopping behavior section contain the most often purchased categories per segment where bold characters indicate a purchase frequency over 50 percent and italics indicate a purchase frequency under 25 percent. We also give demographic and socioeconomic data: the mean of family size, the percentage of married heads of household and of households without children under 18 as well as the three most often appearing classes of household income[14]. Separated for female and male head of household, the table also shows the three most often appearing classes of age and education[15].

A oneway ANOVA shows that the segments differ significantly in mean basket size ($F = 59.038, p = .000$) and number of baskets per household ($F = 510.451, p = .000$). The most important interactions show some overlap, except for segment 3 which has the smallest mean basket size consequently resulting in few and only negative interactions of considerable size. Milk, salty snacks, carbonated beverages, and cold cereals, which are overall the four most often purchased categories, are also present in the six most often purchased categories per segment. This overlap is unexpectedly large.

With regard to sociodemographic variables, there is some variation between segments, for example in family size ($F = 13.621, p = .000$). Whereas the modes of categorical variables are surprisingly consistent over segments, χ^2-tests show that their distribution varies (for all $\chi^2 > 65, p = .000$). Segment 2 tends to have the youngest households with the highest income, the best education, the largest family size, the highest percentage of married heads of household and the lowest percentage of households without children under 18 years; its

[14]Household income in thousand US$.

[15]HS: high school; C: college; TS: technical school; PG: postgraduate work. SC: some college, what means that the person left college without a degree.

Table 5: Detailed comparison of segments

	$k = 1$	$k = 2$	$k = 3$	$k = 4$
Shopping Behavior				
Basket Size	3.82	6.28	1.77	2.52
Number Baskets	12.45	20.70	6.68	8.14
Most	TOB/TOP, DIA/SHA	PAT/TAT, HHC/SUS	*TOB/YOG*	DEO/TOP, TOB/TOP
important	BLA/TOP, MAY/MUK	FRP/FRD, MAB/SUS	*MIL/SOU*	FAT/TAT, FRP/FRD
interactions	FRP/LAD, PEB/TAT	MAY/MUK, COF/MIL		PAT/TAT, COL/SPA
Most often	**MIL**, SAL	**CBV, SAL**	*MIL, COL*	CBV, SAL
purchased	COL, CBV	**MIL**, TAT	*SAL, MAB*	YOG, MIL
categories	*YOG, TAT*	COL, PAT	*SOU, CBV*	*COL, SOU*
Sociodemographics				
Family Size	2.62	2.85	2.23	2.52
Income	45-55/25-35/75-100	100+/75-100/45-55	25-35/35-45/45-55	35-45/25-35/45-55
% Married	70.01	80.95	56.00	65.87
% No Children	72.70	71.43	84.00	76.24
Male Age	65+/55-64/45-54	45-54/55-64/65+	65+/55-64/45-54	65+/45-54/55-64
Female Age	65+/55-64/45-54	45-54/55-64/35-44	65+/55-64/45-54	65+/45-54/55-64
Male Education	HS/C/SC	C/HS/TS	HS/C/TS	HS/SC/C
Female Education	HS/SC/C	C/HS/PG	HS/SC/TS	HS/SC/C

opposite is segment 3.

4 Conclusions and Future Research

We introduce a finite mixture multivariate logit model comprising a variable selection step to analyze binary market basket data. We find that a finite mixture model significantly improves fit over a homogeneous model and reveals segments with different reactions to marketing instruments, shopping behavior and sociodemographic characteristics.

Knowledge on categories which interact is of benefit to a manager who wants to increase the number of purchases in other categories. To this end she can make an appropriate category more attractive to customers by, e.g., sales promotion, advertising or assortment decisions. If she bases such decisions on the homogeneous model, she overestimates the effect of such measures because the homogeneous model implies 34 interactions which are not considered relevant according to the four segment model.

We give a few examples for the grocery store whose data we analyzed based on the four segment model. For this purpose we draw mainly on the results given in table 5. If purchases across all considered categories should be increased, categories with several strong positive interactions should be selected, e.g., categories TOP, TAT, and PAT. On the other hand, marketing measures in categories which only affect purchases of few other categories are clearly inappropriate to achieve this goal. For our data this applies to the categories COF, HHC, and MAB each of which only has one strong interaction in segment 2 which is the smallest segment both with respect to the number of households and the number of baskets.

Management could also take advantage from information on differences between market segments. For example, if management considers high income households as target group, the two strong positive interactions of SUS in segment 2 suggest sales promotion measures in this category.

Several model variations are worthwhile considering. For instance, the use of a continuous mixture model might augment model accuracy but may be less helpful from a managerial point of view. The optimal number of segments might also be estimated endogenously by determining a prior and drawing the number in an additional sampling step (McLachlan and Peel, 2000).

From a management point of view, it would be interesting to model marketing mix effects not only on category constants but also on intercategory effects. This is a precondition to separate consumption complementarity from marketing-mix complementarity effects, i.e. the direct impact of a marketing activity for one category on the purchase probability of a second category (Niraj et al., 2008), and thus to control explicitly for demand and profit impacts of marketing measures. The obvious problem of this extension is an explosion of parameters. For our data set we would have to cope with $3 \cdot (31 \cdot 30) = 2790$ additional parameters for each segment.

Another idea for a further development is modeling the purchase quantity decision, which has not been achieved in full extent. Niraj et al. (2008) propose a two-stage bivariate logit model for purchase/non-purchase in the category (stage 1) and for one unit/more than one unit purchased (stage 2). The exact quantity is not referred to. Only Wang et al. (2007) explicitly model the exact number of purchased products; they propose a multivariate poisson model for the analysis of cross-category store brand purchasing in five product categories.

Appendix: Pseudo Likelihood Estimation

With this method, the FM-MVL model is estimated in $t = 1, ..., T$ iterations over three sampling steps. First, the probability of each household h belonging to segment k is estimated. Given an assignment of households to components based on the probabilities in *step 1*, the components are independent (Shi et al., 2005). As mentioned before, we assume that not all interaction effects contribute to the explanation of purchase behavior and include a variable selection step that determines which effects are not significantly different from zero, i.e. which categories are pairwise independent (*step 2*). This subalgorithm is located between segmentation and parameter estimation. Afterwards, component-specific parameters are drawn by a hybrid Monte Carlo step (Duane et al., 1987; Horowitz, 1991; Neal, 1996) that simulates a dynamical (Hamiltonian) system (*step 3*). The advantage of a hybrid MC algorithm is the possibility to suppress random walks for a sequence of steps L thus enabling a guided search for new parameter values (Neal, 1996). Within one iteration, *step 2* and *step 3* are performed for every component k. If possible, we suppress the iteration index t in the following for ease of exposition.

Sampling Details

Step 1: The probability of one household h belonging to component k given the component assignments z of all other households $\backslash h$ is evaluated by a Gibbs sampling step according to

$$\varpi_{kh} = P(z_h = k | z_{\backslash h}, Y, \beta) \propto P(z_h = k | z_{\backslash h}) \prod_{i \in h} P(Y_i | \beta_k) \tag{14}$$

with $\beta_k = (\alpha_k, \gamma_k, \xi_k, \theta_k)$ and $P(Y_{i \in h} | \beta_k) = P(Y_{i \in h} | k)$ (Shi et al., 2005). With these probabilities, the complete household assignment to segments z is sampled for one iteration resulting in K subsets of households with a total of I_k market baskets Y_k that are assigned to a segment k. The intractable joint basket probability $P(Y_i | \beta_k)$ is replaced with the respective PL value $\prod_j P(Y_{ij} | Y_{il}, \beta_k)$. $P(z_h = k | z_{\backslash h})$ is calculated as in Shi et al. (2005).

Step 2: We make use of an algorithm explained in Geweke (2005) for linear regression and applied in Dippold and Hruschka (2010) for a homogeneous logit model. Given its contribution to the explanation of purchase behavior, the probability of being zero $\rho_{kjl} = p(\theta_{kjl} = 0 | \theta_{kjn}((n \neq l) \cap (n \in S_{kj})), Y_k, h_{kj})$ is determined for every single interaction[16] parameter

[16] For model stability, variable selection is not applied to category constants or marketing mix coefficients.

conditioned on the other parameters determined as significant within the component as

$$\begin{aligned}
(15) \qquad & \rho_{kjl} \propto \underline{p}\ exp(-h_k \sum_i d^2_{ikjl}/2) \\
with \qquad & \underline{p} = 0.5, \\
& d_{ikjl} = \tilde{Y}_{ikj} - \sum_{(n \neq l) \cap (n \in S_{kj})} \theta_{kjn} Y_{il} \\
and \qquad & \tilde{Y}_{ikj} = -ln[-\frac{ln \upsilon_1}{1 + exp(V(Y_{ij}|Y_{il}, k))} - \frac{ln \upsilon_2}{(1 + exp(V(Y_{ij}|Y_{il}, k)))(1 - Y_{ij})}]
\end{aligned}$$

being the stochastic utility of a purchase in category j for a market basket i given segment k with random uniform numbers $\upsilon_1, \upsilon_2 = U(0,1)$, as proposed in Tüchler (2008). The precision h_k is distributed as

$$(16) \qquad h_k \sim \chi^2(I_k + \underline{\nu})/(sse_k + \underline{s}^2)$$

with the sum of squared errors sse_k and the number of baskets assigned to category k, I_k, as well as prior values for variance $\underline{s}^2$ and free parameters $\underline{\nu}$.

Only if this probability ρ_{kjl} is smaller than a random uniform number $U(0,1)$, the respective interaction parameter θ_{kjl} is sampled in the next step.

Step 3: Assuming an a priori independence of parameters β_k across segments, we can sample segment-specific significant parameters independently for each segment with

$$(17) \qquad P(\beta_k|z, Y) = P(\beta_k|Y_k) \propto P(\beta_k) \prod_{i \in I_k} P(Y_i|\beta_k)$$

where $P(\beta_k)$ serves as prior. Again, $P(Y_i|\beta_k)$ is replaced with $\prod_j P(Y_{ij}|Y_{il}, \beta_k)$. Characteristic of the hybrid MC, the conditional parameter sampling density is treated as proportionate to a fictitious potential energy ε , i.e.

$$(18) \qquad P(\beta_k|Y_k) \propto exp(-\varepsilon)$$

as if the parameters β_k were the position parameters in a dynamical system. For every parameter within β_k, a respective momentum vector ϕ_k with $npar$ elements is defined. Together with the associated kinetic energy $\kappa = 0.5\phi'_k\phi_k/\lambda$ (mass $\lambda = 1$ as suggested by Rasmussen (1996)), we get the total energy of the system $H = \varepsilon + \kappa$. This total energy H facilitates draws from the joint distribution of the parameters in the dynamical system because

$$(19) \qquad P(\beta_k, \phi_k|Y_k) \propto exp(-H) = exp(-\varepsilon - \kappa).$$

This relation is used for updating position and momentum parameters in a stepwise process. Holding the total energy H constant within one iteration allows for a sequence of L directed improvements of β values. These substeps L are called leap-frog steps. With one leapfrog step and stepsize η, the parameter update for one parameter of β_k, e.g. θ_{kjl} with respective momentum variable ϕ_{kjl}, reads as (Shi et al., 2005)

$$\begin{aligned}
\phi_{kjl}(\frac{\eta}{2}) &= \phi_{kjl} - \frac{\eta}{2}\frac{\partial\varepsilon(\beta_k)}{\partial\theta_{kjl}} \\
\theta_{kjl}(\eta) &= \theta_{kjl} + \eta\phi_{kjl}(\frac{\eta}{2})/\lambda \\
\phi_{kjl}(\eta) &= \phi_{kjl}(\frac{\eta}{2}) - \frac{\eta}{2}\frac{\partial\varepsilon(\beta_k(\eta))}{\partial\theta_{kjl}}
\end{aligned} \tag{20}$$

We use $L = 2$ and $\eta = 0.003$ for burn-in, afterwards $L = 5$ and $\eta = 0.0001$. After one iteration, the value for the kinetic and consequently also the total system energy are slightly changed by perturbating ϕ_k thus exploring the whole phase space (Neal, 1996). In a Metropolis-Hastings step, the new estimates are proposed as new parameters with probability $min(1, exp[H(\beta_k^{t-1}, \phi_k^{t-1}) - H(\beta_k^t, \phi_k^t)])$. Further details on these sampling steps can be found in Neal (1996) and Shi et al. (2005).

References

Abramson, C., Andrews, R.L., Currim, I.S., Jones, M., 2000. Parameter Bias from Unobserved Effects in the Multinomial Logit Model of Consumer Choice. *Journal of Marketing Research*, 37 (4), 410-426.

Ainslie, A., Rossi, P.E., 1998. Similarities in Choice Behavior Across Product Categories. *Marketing Science*, 17 (2), 91-106.

Akaike, H., 1974. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19 (6), 716-723.

Allenby, G.M., Arora, N., Ginter, J., 1998. On the Heterogeneity of Demand. *Journal of Marketing Research*, 35 (3), 384-389.

Allenby, G.M., Rossi, P.E., 1999. Marketing Models of Consumer Heterogeneity. *Journal of Econometrics*, 89 (1-2), 57-78.

Andrews, R.L., Ainslie, A., Currim, I.S., 2002. An Empirical Comparison of Logit Choice Models with Discrete versus Continuous Representations of Heterogeneity. *Journal of Marketing Research*, 39 (4), 479-487.

Andrews, R.L., Currim, I.S., 2002. Identifying Segments with Identical Choice Behaviors Across Product Categories: An Intercategory Logit Mixture Model. *International Journal of Research in Marketing*, 19 (1), 65-79.

Andrews, R.L., Currim, I.S., 2003. A Comparison of Segment Retention Criteria for Finite Mixture Logit Models. *Journal of Marketing Research*, 40 (2), 235-243.

Besag, J., 1974. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B*, 36 (2), 192-236.

Besag, J., 1975. Statistical Analysis of Non-Lattice Data. *The Statistician*, 24 (3), 179-195.

Besag, J., 1977. Efficiency of Pseudolikelihood Estimation for Simple Gaussian Fields. *Biometrika*, 64 (3), 616-618.

Boatwright, P., Dhar, S., Rossi, P.E., 2004. The Role of Retail Competition, Demographics and Account Retail Strategy as Drivers of Promotional Sensitivity. *Quantitative Marketing and Economics*, 2 (2), 169-190.

Boztuğ, Y., Hildebrandt, L., 2008. Modeling Joint Purchases with a Multivariate MNL Approach. *Schmalenbach Business Review*, 60 (October), 400-422.

Boztuğ, Y., Reutterer, T., 2008. A Combined Approach for Segment-Specific Market Basket Analysis. *European Journal of Operational Research* 187 (1), 294-312.

Boztuğ, Y., Silberhorn, N., 2006. Modellierungsansätze in der Warenkorbanalyse im Überblick. *Journal für Betriebswirtschaft*, 56 (2), 105-128.

Bronnenberg, B.J., Kruger, M.W., Mela, C.F., 2008. The IRI Marketing Data Set. *Marketing Science*, 27 (4), 745-748.

Carlin, B.P., Louis, T.A., 2000. *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton: Chapman & Hall/CRC, Carreira-Perpiñán, M.A., Hinton, G.E., 2005. On Contrastive Divergence Learning. In R.G. Cowell, & Z. Ghahramani (Eds.), *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, Jan 6-8, 2005, Savannah Hotel, Barbados.

Celeux, G., 1998. Bayesian Inference for Mixtures: The Label Switching Problem. In R. Payne, & P. Green (Eds.), *COMPSTAT 98* (pp. 227-232). Heidelberg: Physica.

Chib, S., Seetharaman, P.B., Strijnev, A., 2002. Analysis of Multi-Category Purchase Incidence Using IRI Market Basket Data. In P.H. Franses, & A.L. Montgomery (Eds.), *Econometric Models in Marketing 16* (pp. 57-92). Amsterdam: JAI.

Cox, D.R., 1972. The Analysis of Multivariate Binary Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 21 (2), 113-120.

Cressie, N.C., 1993. *Statistics for Spatial Data. Revised Edition*. New York: John Wiley & Sons, Inc.

Dippold, K., Hruschka, H., 2010. Variable Selection for Market Basket Analysis. Working Paper, University of Regensburg.

Duane, S., Kennedy, A.D., Pendleton, B.J., Roweth, D., 1987. Hybrid Monte Carlo. *Physics Letters B*, 195 (2), 216-222.

Duvvuri, S.D., Ansari, A., Gupta, S., 2007. Consumers' Price Sensitivities Across Complementary Categories. *Management Science* 53 (12), 1933-1945.

Efron, B., Tibshirani, R.J., 1998. *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability, Vol. 57.* Boca Raton: Chapman & Hall/CRC.

Frühwirth–Schnatter, S., 2006. *Finite Mixture and Markov Switching Models*. New York: Springer.

Geweke, J., 2005. *Contemporary Bayesian Econometrics and Statistics*. Hoboken, NJ: John Wiley & Sons, Inc.

Geyer, C.J., Thompson, E.A., 1992. Constrained Monte Carlo Maximum Likelihood for Dependent Data. *Journal of the Royal Statistical Society. Series B*, 54 (3), 657-699.

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

Gupta, S., Chintagunta, P.K., 1994. On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models. *Journal of Marketing Research*, 31 (1), 28-36.

Hansen, K.T., Singh, V., Chintagunta, P.K., 2006. Understanding Store-Brand Purchase Behavior Across Categories. *Marketing Science*, 25 (1), 75-90.

Hoch, S.J., Kim, B.-D., Montgomery, A.L., Rossi, P.E., 1995. Determinants of Store-Level Price Elasticity. *Journal of Marketing Research*, 32 (1), 17-29.

Horowitz, A.M., 1991. A Generalized Guided Monte Carlo Algorithm. *Physics Letters B*, 268 (2), 247-252.

Hruschka, H., 1991. Bestimmung der Kaufverbundenheit mit Hilfe eines probabilistischen Meßmodells. *zfbf Zeitschrift für betriebswirtschaftliche Forschung*, 43, 418-434.

Hruschka, H., Lukanowicz, M., Buchta, C., 1999. Cross-Category Sales Promotion Effects. *Journal of Retailing and Consumer Services*, 6 (2), 99-105.

Huang, F., Ogata, Y., 2002. Generalized Pseudo-Likelihood Estimates for Markov Random Fields on Lattice. *Annals of the Institute of Statistical Mathematics* 54 (1), 1-18.

Hyvärinen, A., 2006. Consistency of Pseudolikelihood Estimation of Fully Visible Boltzmann Machines. *Neural Computation*, 18 (10), 2283-2292.

Kim, B.-D., Srinivasan, K., Wilcox, R.T., 1999. Identifying Price Sensitive Consumers: the Relative Merits of Demographic vs. Purchase Pattern Information. *Journal of Retailing*, 75 (2), 173-193.

Kruger, M.W., Pagni, D., 2008. IRI Academic Data Set Description, version 1.31. Chicago: Information Resources Incorporated.

Laud, P.W., Ibrahim, J.G., 1995. Predictive Model Selection. *Journal of the Royal Statistical Society. Series B*, 57 (1), 247-262.

Magnussen, S., Reeves, R., 2007. Sample-based Maximum Likelihood Estimation of the Autologistic Model. *Journal of Applied Statistics*, 34 (5), 547-561.

Manchanda, P., Ansari, A., Gupta, S., 1999. The "Shopping Basket": A Model for Multi-Category Purchase Incidence Decisions. *Marketing Science*, 18 (2), 95-114.

McFadden, D., 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105-142). New York: Academic Press, Inc.

McLachlan, G.J., Peel, D., 2000. Finite Mixture Models. New York: John Wiley & Sons, Inc.

Moon, S., Russell, G.J., 2004. A Spatial Choice Model for Product Recommendations. Working Paper, Tippie School of Business, University of Iowa.

Neal, R.M., 1993. Bayesian Learning for Neural Networks. Lecture Notes in Statistics 118. New York: Springer.

Niraj, R., Padmanabhan, V., Seetharaman, P.B., 2008. Research Note: A Cross-Category Model of Households' Incidence and Quantity Decisions. *Marketing Science*, 27 (2), 225-235.

Nylund, K.L., Asparouhov, T., MuthÂŽen, B.O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling*, 14 (4), 535-569.

Osindero, S., Welling, M., Hinton, G.E., 2006. Topographic Product Models Applied to Natural Scene Statistics. *Neural Computation*, 18 (2), 381-414.

Rasmussen, C.E., 1996. Evaluation of Gaussian Processes and Other Methods for Non-linear Regression. PhD Thesis, University of Toronto.

Rossi, P.E., McCulloch, R.E., Allenby, G.E., 1996. The Value of Purchase History Data in Target Marketing. *Marketing Science*, 15 (4), 321-340.

Russell, G.J., Ratneshwar, S., Shocker, A.D., Bell, D., Bodapati, A., Degeratu, A., Hildebrandt, L., Kim, N., Ramaswami, S., Shankar, V.H., 1999. Multi-Category Decision-Making: Review and Synthesis. *Marketing Letters*, 10 (3), 319-332.

Russell, G.J., Petersen, A., 2000. Analysis of Cross Category Dependence in Market Basket Selection. *Journal of Retailing*, 76 (3), 319-332.

Schwarz, G., 1978. Estimating the dimension of a model. *The Annals of Statistics*, 6 (2), 461-464.

Seetharaman, P.B., Ainslie, A., Chintagunta, P.K., 1999. Investigating Household State Dependence Effects Across Categories. *Journal of Marketing Research*, 36 (4), 488-500.

Seetharaman, P.B., Siddhartha, C., Ainslie, A., Boatwright, P., Chan, T., Gupta, S., Mehta, N., Rao, V., Strijnev, A., 2005. Models of Multi-Category Choice Behavior. *Marketing Letters*, 16 (3-4), 239-254.

Sherman, M., Apanasovich, T.V., Carroll, R.J., 2006. On Estimation in Binary Autologistic Spatial Models. *Journal of Statistical Computation and Simulation*, 76 (2), 167-179.

Shi, J.Q., Murray-Smith, R., Titterington, D.M., 2005. Hierarchichal Gaussian Process Mixtures for Regression. *Statistics and Computing*, 15 (1), 31-41.

Singh, V.P., Hansen, K.T., Gupta, S., 2005. Modeling Preferences for Common Attributes in Multi-Category Brand Choice. *Journal of Marketing Research*, 42 (2), 195-209.

Song, I., Chintagunta, P.K., 2007. A Discrete-Continuous Model for Multicategory Purchase Behavior of Households. *Journal of Marketing Research*, 44 (4), 595-612.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society. Series B*, 64 (4), 583-639.

Tüchler, R., 2008. Bayesian Variable Selection for Logistic Models Using Auxiliary Mixture Sampling. *Journal of Computational and Graphical Statistics*, 17 (1), 76-94.

Varki, S., Chintagunta, P.K., 2004. The Augmented Latent Class Model: Incorporating Additional Heterogeneity in the Latent Class Model for Panel Data. *Journal of Marketing Research*, 41 (2), 226-233.

Wang, H. (D.), Kalwani, M.U., Akcura, T., 2007. A Bayesian multivariate Poisson regression model of cross-category store brand purchasing behavior. *Journal of Retailing and Consumer Service*, 14 (6), 369-382.

Wang, L., Liu, J., Li, S.Z., 2000. MRF parameter estimation by MCMC method. *Pattern Recognition*, 33 (11), 1919-1925.

Wedel, M., Kamakura, W.A., 1998. *Market Segmentation. Conceptual and Methodological Foundations. International Series in Quantitative Marketing.* Boston: Kluwer Academic Publishers.

Wedel, M., Kamakura, W.A., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., Johnson, R., Lenk, P., Neslin, S., Poulsen, C.S., 1999. Discrete and Continuous Representations of Unobserved Heterogeneity in Choice Modeling. *Marketing Letters*, 10 (3), 219-232.