Erasmus University Rotterdam

Bachelor Thesis

Quantitative Marketing and Business Analytics

Accounting for Taste

A performance comparison of the Multinomial Logit, Mixed Logit and Latent Class Models

Author
C.P.F. de Nijs (412673)

Supervisor
A. Castelein

July 8, 2018

Abstract

Despite the abundance of advertisement and price promotion in marketing, consumption choices often rely just as heavily on personal taste. We compare the conventional Multinomial Logit discrete choice model, with a Mixed Logit specification, as to ascertain whether consumers behave hetero- or homogeneously. We consider both Multivariate Normal and discrete distributions for model parameters. Model performance is measured by in-sample information criteria, as well as out-of-sample forecasting performance. We find strong evidence of preference heterogeneity in both a product marketing setting, as well as a discrete choice survey experiment. By incorporating a Guadagni and Little (1983) loyalty variable, we extend this result to a dynamic choice setting, finding residual preference heterogeneity, even when brand loyalty is accounted for.

1 Introduction

Which university to attend, where to go on holiday, which health insurance to buy. Discrete choices permeate our economic life. The workhorse model to describe such choices was given by McFadden (1973) in the Multinomial Logit model (MNL). However, since the model is quite restrictive in its assumptions, more flexible models have since been proposed. The Mixed Logit Model (MIXL) allows for random coefficients, bypassing the coefficient homogeneity assumption inherent in the MNL. Whilst attribute coefficients are often assumed to be generated through a multivariate normal distribution, the Latent Class (LC) model assumes a discrete distribution for the MNL coefficients. This means that whilst individuals may have somewhat heterogeneous preferences and reactions to attributes, they all fit in some class of people that shares a set of coefficients.

In a partial replication of Fiebig et al. (2010), we aim to compare model performance of the MNL, LC, and MIXL in modeling household choices for crackers, as well as trail preference among mountain bikers. By means of information criteria, we may observe which model most efficiently matches consumer behavior in a product marketing setting. Given the different sets of assumptions underlying each model, this then teaches us the nature of human reactions to choice attributes; whether they are generally homogeneous or individual-specific.

We extend on Fiebig et al. (2010) by comparing MIXL and MNL performance when dynamics are incorporated through a Guadagni and Little (1983) loyalty variable. Fiebig et al. (2010) find significant improvements in model performance when moving from MNL to MIXL. This improvement may partly originate in the consideration for brand preference offered by MIXL. As alternative-specific constants are allowed to vary, choice agents are allowed individual preferences across alternatives. However, we may consider other methods of modeling brand preference, such as the inclusion of dynamics. If we can model brand preference through an additional covariate to indicate purchase history, MNL may make similar improvements in model fit, without needing to assume randomly distributed parameters. We estimate both MNL and MIXL with the addition of a Guadagni and Little (1983) loyalty variable to ascertain whether the performance superiority of MIXL over MNL holds up when both models account for brand preference through dynamics. This may provide valuable insights to product marketeers seeking to model client choices based on previous purchase behaviors.

We further extend the performance evaluations by Fiebig et al. (2010) by considering out-of-sample forecasting performance, as well as both revealed preference as stated preference data.

This allows us to observe whether the in-sample superior performance of MIXL over MNL in the stated preference data studied by Fiebig et al. (2010) holds up out-of-sample and for actual choice behaviors. We seek to thereby broaden the scope of existing knowledge on the added value of parameter randomization to forecasting performance, dynamic choice modeling and a product marketing setting.

We find MIXL to consistently outperform the MNL model on both in-sample information criteria, as well as out-of-sample forecast measures. The Latent Class model outperforms MNL, but underperforms relative to MIXL, suggesting that consumer preference heterogeneity is distributed continuously, rather than discretely. The improvement in fit from MIXL may be largely traced back to its consideration for brand preference. The inclusion of dynamics in MNL and MIXL brings their model performance closer together, but MIXL still outperforms MNL in this setting. In the context of cracker purchase decisions, we find purchase history to play an economically large role. Willingness to pay for a particular brand rises by \$0.53 on average if the brand was purchased in the period preceding the purchase decision.

We continue with a description of the data used in section 2. Section 3 gives a review of the literature, after which Section 4 describes the data used, with section 5 reporting results and section 6 concluding.

2 Literature Review

Given the abundance of discrete choices in economic behaviors, choice modeling has been treated extensively in the econometric literature. In proposing the MNL model, McFadden (1973) argues that economic choices are made according to some "common behavioral rule". Choice utilities are modeled to additively depend on choice characteristics, with the whole population sharing similar preferences for these characteristics.

Choice probabilities are then calculated according to Luce's (1959) choice axiom, thereby inferring the Independence of Irrelevant Alternatives (IIA) assumption. By calculating the probabilities as $\exp(utility_i)/\sum_j \exp(utility_j)$ for any choice i in choice set $\{1, \ldots, J\}$, the relative probabilities of two choices are independent of the presence or characteristics of any other choices in the set $\{1, \ldots, J\}$. A strict assumption, which has proven questionable by both empirical results and the proposition of alternative models (Currim, 1982). IIA implies that the addition of a new choice i_{new} to set $\{1, \ldots, J\}$ would have all the other choices surrender a part of their share of the market in proportion to previous market shares. This requires careful considerations in the construction of the choice set $\{1, \dots, J\}$ by the researcher. For example, if we were to model transport mode choices from the set {*bike, car, train*}, we could not add a secondary public transport option such as *bus*, as it is unlikely to receive proportionally as many customers from previous car owners as previous train users. Similarly, if we were to remove the car option, the IIA would expect the long distances previously traveled by car to now be undertaken as much by bike as by train, relative to previous market proportions. Alternative models yield more freedom in the construction of choice sets by circumventing the IIA assumption. However, the MNL remains a popular tool for analysis in choice sets with equally similar choices, as well as for preliminary analysis in more complicated datasets.

One such alternative model is the Mixed Logit Model (McFadden and Train, 2000). By modeling not just a population-wide "common behavioral rule" as per McFadden (1973), but the distribution over the individuals of behavioral rules, different actors are no longer assumed to have identical reactions to changes in choice attributes. Given that, in a marketing setting, individuals often exhibit a quite specific, seemingly random preference for a particular brand or product, the population-wide mean probabilities under MNL may provide an inaccurate estimation. The MIXL allows for a choice setting in which individuals have varying sensitivities to changes in price or other choice attributes, as well as variable base utilities for each choice. As such, if As individuals are allowed their own parameter vectors, the IIA assumption need no longer hold. Using the example of choice set {bike, train, car}, if a bus is added to the choice set, MIXl will estimate those individuals with parameters that favour public transport to have high utilities for the bus. Those who favour private transport will see lower utilities. As such, the bus option will disproportionally cannibalize on the market share of previous train users, in violation of the IIA. A specific case of the MIXL model is given by the Latent Class model. Kamakura and Russell (1989) argue that we may segment the individuals in a dataset, drawing their brand preferences and attribute sensitivities from a discrete distribution. This allows for choice probabilities to be based on both observable characteristics and unobservable ex-ante preference. In a study of retail scanner data, they find worthwhile gains in model likelihood to be made from estimating segment-specific parameters, with segments ordered by the estimated loyalty of the household. Whilst each segments comes at the cost of having to estimate another segment-specific parameter vector, the gains in model performance are such that the AIC is found to be optimal at 6 clusters. This shows significant heterogeneity among consumers in a supermarket shopping setting, as studied in this paper also. If this heterogeneity is clustered across different discrete segments, the Latent Class model may deliver greater performance than the MIXL model with normally distributed parameters. If there are distinct disjoint peaks in the distribution of parameters, the bell-curved normal distribution may provide a poor fit.

In a study of survey data from New Zealand car drivers, Greene and Hensher (2003) compare the performance of Latent Class and Mixed Logit model specifications, in estimating the effects of certain road attributes on road choice. They find model performance to be reasonably on par with each other, arguing that model choice should largely be based on whether the researchers wish to avoid distributional assumptions as in the LC model, or seeks flexibility in modeling the unobserved heterogeneity, provided in the choice of distribution in the MIXL model.

Given the availability of these tools for modeling consumer heterogeneity, we may wonder whether its use is warranted based on actual consumer behaviors. Specifically, whether consumers react to choice attributes homogeneously or heterogeneously. Lim et al. (2005) study household-level scanner data from a supermarket to determine whether reactions to product attributes differ among consumers. They find significant segmentation among consumers. Brand-loyal consumers act differently from less predictable consumers, and a distinction was found between those who consume great quantities of a product and those who don't. Market segmentation significantly improved forecast performance of brand choices for two out of four products studied. Whilst this result presents some evidence of the value of considering heterogeneous preferences, the authors consider a discrete distribution into two classes, without performing a comparison with continuously distributed parameter models. Better fitting models could perhaps be achieved by allowing more taste variation, as human behavior may not fall only into two categories.

Fiebig et al. (2010) provide another model comparison using stated preference data. The authors compare the MNL and MIXL model with the Scaled Multinomial Logit model (S-MNL), which assumes homogeneity in the vector of coefficients, but heterogeneity in the scale of this vector. They find great improvements in model performance, as measured by information criteria, through allowing coefficient heterogeneity. However, the authors argue that this result is more symptomatic of scale heterogeneity rather than heterogeneity in reactions to attributes. Whilst individuals do behave differently, a large part of this difference comes from the extremity of their reactions, not the nature of their reactions themselves. This lends some credibility once more to McFadden's assumption of a "common behavioral rule". If people behave as specified by S-MNL, they exhibit a common reaction to variations in attributes, but do so with different scales to their reaction.

We aim to elaborate on the current literature by extending the MNL model with a Guadagni and Little (1983) loyalty variable, giving this basic specification some consideration for dynamics. Modeling scanner supermarket data for coffee purchases, Guadagni and Little (1983) present an MNL specification with a parametric loyalty variable. They find significant loyalty effects for both brand and size of serving. Given the similarity of the data with the scanner data studied in this paper, we expect the addition of a loyalty variable to improve the performance of the basic MNL for this dataset also.

Fiebig et al. (2010) report a large improvement from estimating MIXL over MNL on all 10 datasets studied. However, we may wonder whether this is rooted in individual variation in attribute coefficients, or variation in the Alternative Specific Constants (ASCs). Out of 10 datasets studied by Fiebig et al., the 4 datasets with the smallest percentage improvement in log-likelihood from going from an MNL to a Generalized MNL (which nests MIXL), were those without any ASCs. By accounting for taste within MNL, by constructing a loyalty variable, we may already capture a large part of the performance improvement presented by MIXL.

The resulting MNL model is then compared with both the MIXL model, as in Fiebig et al. (2010), as well as the latent class model. This allows us to observe whether the superiority of the MIXL model's performance holds when the MNL includes dynamics, as well as when compared to discrete consumer heterogeneity through the Latent Class model.

3 Data

To compare the various choice models, we employ a panel dataset of household shopping behavior, as previously used by Jain et al. (1994). The data observes 136 households across 14 to 77 store visits per household. Each observation has the household purchasing crackers, choosing one of four possible brands. Table 1 reports descriptive statistics for brand choice and brand attributes across all visits. We see that Nabisco is by far the most popular brand with a market share of 54%, with the Private brand capturing 31% of the market, leaving 7% each for the Sunshine and Kleebler brands. On average, the private brand is the cheapest, with Kleebler reporting the highest mean price. Besides being frequently chosen, Nabisco is also most likely to be engaged in a feature promotion, as well as most likely to be in a special store display. Given the general homogeneity of the choice options, as well as the presence of some explanatory variables to motivate brand choice, this dataset lends itself well to the discrete choice models we aim to test.

Table 1: Descriptive Statistics

	Min	Max	Mean	St.Dev		Min	Max	Mean	St.Dev
Sunshine					Kleebler				
Chosen	0.00	1.00	0.07	0.26	Chosen	0.00	1.00	0.07	0.25
Display	0.00	1.00	0.13	0.34	Display	0.00	1.00	0.11	0.31
Feature	0.00	1.00	0.04	0.19	Feature	0.00	1.00	0.04	0.20
Price	49.00	129.00	95.70	13.29	Price	88.00	139.00	112.59	10.64
Nabisco					Private				
Chosen	0.00	1.00	0.54	0.50	Chosen	0.00	1.00	0.31	0.46
Display	0.00	1.00	0.34	0.47	Display	0.00	1.00	0.10	0.30
Feature	0.00	1.00	0.09	0.28	Feature	0.00	1.00	0.05	0.21
Price	0.00	169.00	107.92	14.48	Price	38.00	115.00	68.07	12.41

Table 1 reports the minimal and maximal values for both choice variables and brand attributes, as well as the mean and standard deviation. Prices are given in dollarcents.

Table 12 reports descriptive statistics for the second dataset used. As previously studied, and kindly supplied by Morey et al. (2002), the dataset reports the result of a discrete choice experiment among mountain bikers. Responses were recorded at a 1995 cycling trade show. Respondents were asked to participate in the survey, conditional on identifying as a mountainbiker. Given the choice between two mountain bike trails with five differing attributes, the experiment recorded choices between 5 randomized pairs for 289 mountain bikers. 269 mountain bikers reported a choice for each pair, the remaining 20 responded to 3 to 4 pairs, likely from suffering some interruption during the survey procedure. As all other information about these cyclist is correctly recorded, we see no reason to invalidate the remaining responses by these cyclists.

Other variables included are age and gender (reported separately in appendix A), as well as cycling experience metrics, bike characteristics and household characteristics. The average cyclist surveyed is 30 years of age and has 2 years of experience with mountain biking. Whilst survey data may not directly translate to actual behaviors, the experimental setting allows for extensive attribute variation to elicit consumer reactions.

Table 2: Descriptive Statistics Mountain Biking Data

Variable	Min	Max	Mean	St.Dev	N
Cyclist ID	1.0	289.0	144.9	83.5	1420
t	1.0	5.0	3.0	1.4	1420
Chose trail A	0.0	1.0	0.5	0.5	1420
Mileage trail A	7.0	21.0	13.5	5.7	1420
Vertical Climbing trail A	400.0	2200.0	1154.2	685.3	1420
Peaks trail A	1.0	4.0	1.7	1.0	1420
Single Track Miles in A	0.0	14.0	5.1	5.4	1420
Hikers/Equestrians on A	0.0	1.0	0.2	0.4	1420
Fee for trail A	1.0	8.0	3.0	2.2	1420
Mileage trail B	7.0	21.0	15.2	5.2	1420
Vertical Climbing trail B	400.0	2200.0	1363.7	673.5	1420
Peaks trail B	1.0	4.0	2.0	1.2	1420
Single Track Miles in B	0.0	21.0	9.6	6.7	1420
Hikers/Equestrians on B	0.0	1.0	0.8	0.4	1420
Fee for trail B	1.0	8.0	4.8	2.7	1420

4 Methodology

4.1 Multinomial Logit

The basic multinomial logit models the choice $y_{i,t}$ of a choice actor i at time t. The choice set $\{1, \ldots, J\}$ contains all available choice alternatives j. MNL assumes there to be a population-wide parameter vector β that determines the utility $U_{i,j,t}$ of a particular choice, conditional on choice characteristics X_j. For individual i choosing choice j at time t, $U_{i,j,t} = \alpha_j x'_{i,j,t}\beta + \epsilon_{i,j,t}$, where α_j gives the intercept. α_j may be interpreted as the base utility assigned to choice j before choice characteristics $x_{i,j,t}$ are accounted for. By Luce's (1959) choice axiom, this yields choice probabilities as in (1):

$$P(y_{i,t} = j|X_{i,t}) = \frac{exp(\alpha_j + x'_{i,j,t}\beta)}{\sum_{g=1}^{J} exp(\alpha_j + x'_{i,g,t}\beta)}; j = 1, ..., J; i = 1, ..., N; t = 1, ..., T. \quad (1)$$

We may deduce the inherent IIA assumption from the log-odds ratios of two competing alternatives j and l, $log(P(y_{i,t} = j|X_{i,t}))/log(P(y_{i,t} = l|X_{i,t})) = \alpha_j - \alpha_l + (x'_{i,j,t} - x'_{i,l,t})\beta$. The relative probability of choosing one alternative j to another option l, is dependent only on the attributes and intercepts of those two parameters. Therefore, if we add or remove other alternatives to the choice set $\{1, \ldots, J\}$, the relative probabilities remain constant, regardless of how similar the additional alternatives are to either j or l.

For a particular estimate of the parameter vector, $\hat{\beta}$, this leads to a likelihood function defined as

$$L(\hat{\beta}|y,x) = \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{J} \frac{exp(x'_{i,j,t}\hat{\beta})^{I_{y_{i,t}=j}}}{\sum_{g=1}^{J} exp(x'_{i,g,t}\hat{\beta})}, \tag{2}$$

where $I_{y_{i,t}=j}$ is an indicator function, taking the value of 1 on the subscripted condition, and 0 otherwise. Using Maximum Likelihood Estimation (MLE), we seek to find the parameter vector estimate which maximizes this likelihood. For computational stability, we do so by maximizing the log-likelihood function.

When alternative-specific constants (ASC) are included, one of these is fixed to 0 to identify the model. Any deviation in individual choices from the probabilities conditional on $X_{i,j,t}$ are part of the error term $\epsilon_{i,j,t}$. Heterogeneity in preferences or attitudes to choice attributes are also swallowed by this error term. This makes the MNL potentially less suited for panel data in which individuals may show certain choice preferences or individual-specific reactions to attribute changes. In the presence of unobserved heterogeneity, the MNL does not capture all the available information in the data, assuming all heterogeneity to be unobservable.

4.1.1 Loyalty

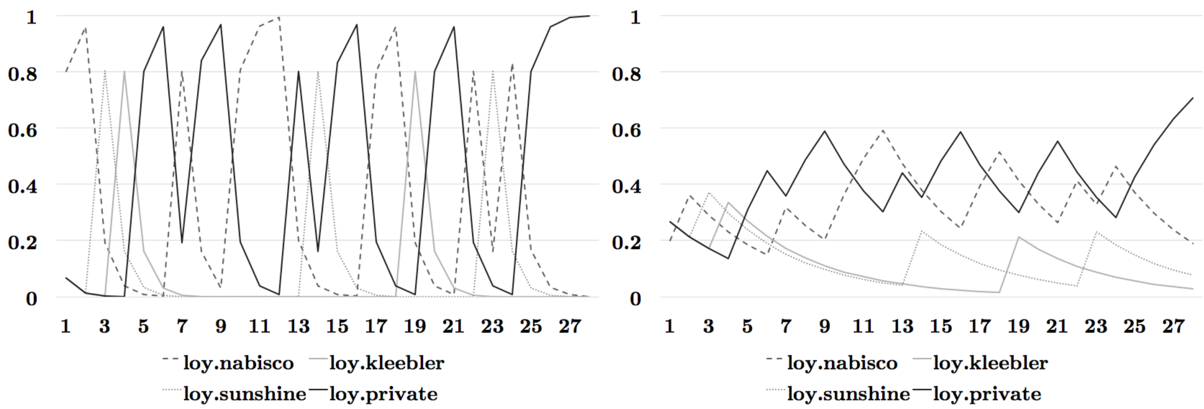
We may extend the MNL estimation slightly to have some consideration for dynamics, by including a loyalty variable as introduced by Guadagni and Little (1983). They propose a loyalty variable $b_{i,j,t}$ for choice j at time for individual i, such that

$$b_{i,j,t} = \phi b_{i,j,t-1} + (1-\phi) I_{y_{i,t-1}=j}, \tag{3}$$

where I an indicator function, $y_{i,t}$ gives the index of the choice made at time t by individual i, and $\phi \in [0,1]$. ϕ is estimated as a model parameter, along with a coefficient β_b to model utility dependency on the loyalty variable. Whilst the inclusion of $I_{y_{t-1}=j}$ as a regressor would also allow for dynamics in the MNL model, (3) gives a more forgiving approach to loyalty behaviors[1], that nests the option of simply including a lagged indicator function.

Figure 1 illustrates the behavior of the loyalty variable for two levels of ϕ, taking household number 78 of the supermarket scanner data as an example. When an individual deviates from his favourite brand, the relevant loyalty variable decays by the factor ϕ. For a high ϕ, this means

[1]Momentary deviations from an otherwise loyal customer yield a deterioration in the loyalty variable $b_{i,j,t}$ by a factor of ϕ, where $\phi \in [0,1]$. Using an indicator function $I_{y_{t-1}=j}$, this factor would always be 0. Ceteris paribus, this means that someone who has purchased brand j all their life but deviated to brand j^* would have an equal probability in the next period of purchasing brand j as someone who has never purchased brand j before. $b_{i,j,t}$ with $\phi > 0$ is more forgiving in these deviations, letting loyalty decay to 0 only after habitual neglect of brand j.



(a) Brand loyalty using $\phi = 0.2$

(b) Brand loyalty using $\phi = 0.8$

Figure 1: Household 78's brand loyalty variables over time

that individuals generally stick to one brand with only momentary deviations, as their loyalty hardly deteriorates from experimentalism. A low ϕ indicates rapid deterioration of loyalty, with strong switching behaviors. For $\phi = 0$, $b_{i,j,t}$ reduces to $I_{y_{t-1}=j}$.

$b_{i,j,1}$ is initialized at $t = 1$ at the value of ϕ for the choice j which was made at that time, and $\phi/(J-1)$, for all other brands. The inclusion of $b_{i,j,t}$ yields J total additional variables, requiring the estimation of 2 additional parameters, ϕ and $\beta_{loyalty}$. Choice probabilities are now computed as

$$P(y_{i,t} = j|X_{i,t}) = \frac{exp(\alpha_j + x'_{i,j,t}\beta + b_{i,j,t}\beta_{loyalty})}{\sum_{g=1}^{J} exp(\alpha_j + x'_{i,g,t}\beta + b_{i,g,t}\beta_{loyalty})}; j = 1, ..., J; i = 1, ..., N; t = 1, ..., T, \tag{4}$$

for two competing alternative j and l, this leads to the following log-odds ratios:

$$log(P(y_{i,t} = j|X_{i,t}))/log(P(y_{i,t} = l|X_{i,t})) = \alpha_j - \alpha_l + (x'_{i,j,t} - x'_{i,l,t})\beta + (b_{i,j,t} - b_{i,g,t})\beta_{loyalty}. \tag{5}$$

As $(b_{i,j,t} - b_{i,g,t})$ may be affected by the addition or removal of other alternatives to choice set $\{1, \ldots, J\}$, the IIA assumption need no longer hold. For example, if a close substitute to j, j^* is added to the choice set, $b_{i,j,t}$ is liable to change, as individuals may spread their choices equally over the two close substitutes. $b_{i,l,t}$ will be unaffected if j^* provides no substitute to l, causing a shift in the log-odds ratio.

Under MIXL, brand preference is also accounted for, through the estimation of randomly distributed intercepts α_j for each brand. This requires the estimation of J additional parameters σ_j for each brand, assuming a multivariate normal distribution of α_j with diagonal Σ (σ_j^2 denoting the elements of the diagonal). For non-diagonal Σ, allowing for correlated parameter values, this requires $\frac{1}{2}J(J+1)$ additional parameters. As such, using $b_{i,j,1}$ to model choice preferences is sure to require less model parameters.

4.2 Mixed Multinomial Logit

To take into account preference heterogeneity, we may want to allow for individual-specific parameters. In the MIXL model (McFadden and Train, 2000), utility is given by

$$U_{i,j,t} = (\beta + \gamma_i)x_{i,j,t} + \epsilon_{i,j,t}; j = 1, ..., J; i = 1, ..., I; t = 1, ..., T, \tag{6}$$

where β gives the population-wide mean of the parameter vector, and γ_i the individual-specific random deviation from that mean. γ_i may be specified to be from any distribution, with the multivariate normal being commonplace. In our estimation, we assume γ_i to be multivariate normal. We estimate the model both with and without an independence restriction on the individual elements of γ_i. Under independence, this implies a diagonal covariance matrix Σ, without independence Σ is unrestricted. As a shorthand, MIXL with diagonal Σ will be referred to as MIXL-I, with the unrestricted model referred to as MIXL-II.

The exact choice probabilities are given by their expected value over the distribution of the parameters, such that the probability of a specific individual i exhibiting a series of choices $y_{i,1}, , ..., y_{i,T}$ is given by

$$P(y_i|X_{i,t}, \beta, \Sigma) = \int (\prod_{t=1}^{T} \frac{\prod_{j=1}^{J} exp((\beta + \gamma_i)x_{i,j,t})^{I_{y_{i,t}=1}}}{\sum_{g=1}^{J} exp((\beta + \gamma_i)x_{i,g,t})}) f(\gamma_i|\Sigma)d\gamma_i, \tag{7}$$

where $\gamma_i \sim N(0, \Sigma)$, and β gives the population-wide mean parameter vector. Seeing as we integrate over γ_i, the ratio of two choice probabilities is a function of all the data. As such, the IIA assumption need not hold under MIXL Train (2009). Lacking a closed form of this integral, we approximate its value through Monte Carlo integration. Taking a set of D draws from the chosen distribution for γ_i, we may approximate the probabilities for each individual by taking their average probability over D draws of γ_i, such that

$$\hat{P}(y_i|X_{i,t}, \beta, \Sigma) = \frac{1}{D} \sum_{d=1}^{D} (\prod_{t=1}^{T} \frac{\prod_{j=1}^{J} exp((\beta + \gamma_i^{(d)})x_{i,j,t})^{I_{y_{i,t}=1}}}{\sum_{g=1}^{J} exp((\beta + \gamma_i^{(d)})x_{i,g,t})}), \tag{8}$$

where $\gamma_i^{(d)}$ is drawn from $N(0, \Sigma)$. These simulated choice probabilities are then used to compute the likelihood function, such that for an estimate $\hat{\beta}, \hat{\Sigma}$,

$$L(\hat{\beta}, \hat{\Sigma}|y, X) = \prod_{i=1}^{N} P(y_i|X_{i,t}, \hat{\beta}, \hat{\Sigma}). \tag{9}$$

Given the inherent variability in the finite amount of random draws, random draws are reused with the proper transformation to allow the BFGS algorithm to find consistent descent directions. This is done by drawing a k-length vector from the standard normal distribution and multiplying it by the Cholesky decomposition L of Σ, where $LL' = \Sigma$, L a lower-triangular matrix, and k the number of parameters. To force the decomposability of Σ, we directly estimate the elements of L, which are transformed post-estimation to the resulting covariance matrix, as well as a correlation matrix.

To obtain standard errors, we use a log-likelihood function which takes the elements of the covariance matrix Σ itself as inputs, rather than the elements of L. A finite-difference approximation of the hessian matrix is attempted using the covariance values at the estimated Likelihood maximum. The estimated standard errors are then given by the diagonal elements of the inverse of the estimated hessian. However, given the instability of the covariance matrix, even small deviations from the optimum may lead to the hessian becoming uninvertible. As a result of this, standard errors for the elements of the covariance matrix are obtained for the MIXL-II model without loyalty in the cracker dataset, but not for the model including loyalty, nor for the mountainbike dataset. Although this hampers conclusive interpretation of the parameter values, a performance comparison of the various models should still be equally possible.

Post-estimation, we may derive individual specific expected parameter values $E[\gamma_i|y_i, X_{i,t}, \beta, \Sigma]$. These may be used for interpretation, as well as for out-of-sample forecasting. Using Bayes rule, it can be shown that the conditional expectation

$$E[\gamma_i|y_i, X_i, \beta, \Sigma] = \frac{\int \gamma_i f(y_i|X_i, \gamma_i) d\gamma_i}{\int f(y_i|X_i, \gamma_i) d\gamma_i}. \tag{10}$$

We estimate the integrals through Monte Carlo integration, taking 500 draws $\gamma_i^{(d)}$ and averaging the terms $\gamma_i f(y_i|X_i, \gamma_i)$ and $f(y_i|X_i, \gamma_i)$ over these draws to obtain estimates of the integrals.

The Mixed Logit model, as compared to the standard MNL, allows for taste variation, relaxes the Independence of Irrelevant Alternatives assumption (IIA), and allows for temporal correlation among choices. In a setting where agents have different preferred choices and personal sensitivities to attributes, this formulation may prove more realistic. However, it comes at a cost of an increased amount of required parameters. Moreover, for a small number of draws in the simulated likelihood function, the model may converge to a fit of the random draws, rather than the actual consumer behaviors.

4.3 Latent Class Model

A specific case of this MIXL model is the Latent Class model (LCM) (see Kamakura and Russell (1989)), which assumes a discrete distribution for the parameter variation γ_i. Rather than assuming parameters to be completely individual-specific, the LCM assumes that preferences can be categorized into a finite number of groups. Specifically, that each individual belongs to some class c, with every class having a specific parameter vector $\beta + \gamma_c$, with β the population mean parameter vector. The probability of person i exhibiting a series of choice $y_{i,1}, ..., y_{i,T}$ is given by

$$P(y_i|X_{i,t}, \beta, \Sigma) = \sum_{c=1}^{C} d_{i,c} \frac{\prod_{j=1}^{J} exp((\beta + \gamma_c)x_{i,j,t})^{I_{y_{i,t}=1}}}{\sum_{g=1}^{J} exp((\beta + \gamma_c)x_{i,g,t})}, \tag{11}$$

where $d_{i,c} = 1$ if individual i belongs to class c, and 0 otherwise. The relevant log-likelihood function to these probabilities is given by $log(L) = \sum_i^N log(\sum_{c=1}^C d_{i,c}(\prod_{t=1}^T P(y_i|c, X_{i,t})))$ This is a missing data model, as $d_{i,c}$ is unknown for each i and each c. We therefore optimize the expected log-likelihood based on estimated probabilities $\hat{p}_c$ with which an individual may fall under a class c. We employ the expectation-maximization algorithm to estimate class probabilities and class-specific parameters γ_c. Starting with an initial guess for p_c of $1/C$, and randomly generated [2], the first step of the algorithm estimates posterior class probabilities, conditional on the current parameters γ_c, where

$$\hat{\pi}_{i,c} = \frac{p_c P(y_i|c, X_{i,t})}{\sum_{c=1}^C p_c P(y_i|c, X_{i,t})}. \tag{12}$$

Using these estimated personal probabilities $\pi_{i,c}$, the second step maximizes through MLE a weighted log-likelihood for each class, such that for each class we have $\hat{\gamma}_c$ that maximizes $\sum_{i=1}^N \sum_{c=1}^C \pi_{i,c} \sum_{t=1}^T log(P(y_{i,t}|X_{i,t}, \gamma_c))$. $\hat{p}_c$ is also set to its maximum likelihood estimate $\frac{1}{N} \sum_{i=1}^N \pi_{i,c}$. The algorithm iterates over expectation and maximization until the total log likelihood converges to a maximum. Conditional on being in a certain class, individuals are assumed

[2]Initial guesses for y_c are generated through a multivariate normal with mean and Σ as estimated through MIXL with diagonal Σ.

to behave in accordance with the IIA assumption. However, as the choice set does influence the estimated probabilities of belonging to a certain class, the IIA does not hold across classes or with respect to ex-ante probabilities. To take the choice set train, car, bike as an example. If a discrete class exists containing those individuals who need to travel great distances, the removal of the 'car' option from the choice set will make this class exclusively demand the train option, without violating the IIA within the cluster. Previous car users may then shift to the train option in disproportion to previous choice probabilities.

The LCM assumes clustering in discrete choice behaviors. In a marketing context, this might mean that we have distinct groups that are price sensitive, some who are sensitive to marketing, and others who are unconditionally loyal to a particular brand, with each of these groups sharing a group-specific deviation from the mean parameter vector that describes these behaviors. We estimate the Latent Class model for 2 classes. We incorporate the Guadagni and Little (1983) loyalty variable in $X_{i,t}$. This allows us to observe not only whether consumers behave heterogeneously after accounting for brand loyalty, but also whether they do so in a continuously distributed manner (which would favor MIXL), or in discrete behavioral groups (which would favor LC).

4.4 Model Comparison

Model parameters are estimated through Simulated Maximum Likelihood Estimation (SMLE) for the MIXL model. We estimate MIXL both with and without the independence assumption on the distribution of the parameter vector β_i. MNL is estimated through Maximum Likelihood. For the household scanner data, we also estimate MNL, LC and MIXL with the Guadagni and Little (1983) loyalty variable described in 4.1.1, as to compare model performance in a setting which includes dynamics. As loyalty considerations are irrelevant to the survey context of the mountainbike dataset, these models are reserved for the first dataset. For all maximum likelihood estimations, the stats4 R package is used to optimize the likelihood functions by means of the BFGS algorithm.

For results of the full sample, the resulting models are then compared by means of the Akaike ($AIC = 2k - 2log(\hat{L})$, with k the number of parameters and $\hat{L}$ the estimated log-likelihood), Bayes ($BIC = log(n)k - 2log(\hat{L})$, with n the number of observations) and consistent Akaike ($CAIC = (log(n) + 1)k - 2log(\hat{L})$) criteria to ascertain which model best fits actual consumer behaviors. Higher values indicate a poorer balance between model fit, represented in the log-likelihood term, as well as parsimony, represented in the τk term, with τ some cost coefficient for k.

The three measures vary in their punishment of overfitting and therefore may yield different results. AIC is strictly less punishing than the others of high k, with CAIC the most strict. Models which are preferred by AIC are therefore likely to fit very well, whilst models which are preferred by CAIC are likely to fit the data very efficiently for the amount of parameters used. In a performance evaluation of the three measures at identifying the correct parametrization of simulated data of choice heterogeneity, Fiebig et al. (2010) find BIC and CAIC to yield the most accurate results.

We also consider out-of-sample performance. Splitting the cracker purchase dataset after the 12th choice occasion for each household, we estimate the models on the first 12 supermarket visits. The remaining household visits are then forecasted using each model as to ascertain out-of-sample performance. Choice probabilities are evaluated as described in the previous subsections, using the conditional expectations for parameter values. For MNL, this simply means that the estimated model parameters are used, whilst for MIXL we may use individual model parameters β_i, as calculated in (9). For the LC model, we use posterior class probabilities as in (11) to construct weighed probabilities $P(y_{i,t} = j) = \sum_{c=1}^{C} \pi_{i,c} P(y_{i,t} = j|c)$. For the mountainbike survey data, the first four responses of each cyclist are used to estimate the model, with the last response used to evaluate forecast performance.

Forecast performance is evaluated by the out-of-sample hit-rate (i.e. the fraction of correct predictions), as well as the F-1 statistic. Computed as $F_1 = \frac{\sum_{j=1}^{J} p_{jj} - p_{.j}^2}{1 - \sum_{j=1}^{J} p_{.j}^2}$ (with $p_{i,j}$ the fraction of observations where j was predicted and i realized), Veall and Zimmermann (1992) find it the most reliable statistic of evaluating prediction-realization tables, as determined through a simulation study comparing various performance measures. The F-1 statistic rewards correct predictions through the p_{jj} term, whilst quadratically punishing the number of predictions $p_{.j}$ of each alternative j. As such, for equal hit rates, models which spread their predictions across the choice set are favored over those which predict only one choice alternative. The statistic ranges from -1 to 1. Models with actual predictive power should expect a positive value Veall and Zimmermann (1992), with perfect prediction models attaining an F-1 value of 1.

Table 3: MNL and MIXL-I results mountainbike data

	MNL		MIXL-I		MIXL-II	
Number of Parameters	6		12		27	
AIC	1915		1809		1721	
BIC	1946		1746		1579	
Consistent AIC	1853		1734		1552	
Log-Likelihood	-951		-917		-888	
Variable	Coefficient	SE/CI	Coefficient	SE/CI	Coefficient	SE/CI
Trail Mileage	0.003	0.013	0.013	0.021	0.005	0.031
Feet of Climbing	0.050	0.077	0.058	0.135	0.202	0.193
Number of Peaks	0.145***	0.040	0.234***	0.069	0.351***	0.091
Miles Singletrack	0.020	0.013	0.034*	0.021	0.052**	0.023
Shared track	-0.563***	0.083	-0.896***	0.144	-1.159***	0.173
Trail fee	-0.039	0.031	-0.071	0.049	-0.113**	0.054
σ_1			0.123	[0.06, 0.23]	0.023	-
σ_2			0.987	[0.63, 1.54]	0.060	-
σ_3			0.408	[0.24, 0.71]	0.786	-
σ_4			0.080	[0.04, 0.16]	0.961	-
σ_5			0.519	[0.20, 1.37]	1.863	-
σ_6			0.260	[0.17, 0.39]	0.856	-

Table 3 reports parameter estimates for MNL and MIXL-I models. AIC, BIC and CAIC values are reported, as well as the log-likelihood for each estimation. σ parameters are estimated in a transformed manner to map their estimated values to their proper ranges of $[0, Inf)$. For MIXL-I we report a 95% confidence interval as inferred from the standard errors of the untransformed parameters. Significance is denoted by *** (1%), ** (5%) and * (10%), respectively.

5 Results

5.1 Full sample, Mountainbike survey

We estimate model parameters for basic MNL as well as MIXL. Table 3 reports the results. Parameter estimates are in accordance with expected trail preferences for cyclists. Trail length, elevation and number of peaks, as well as miles of singletrack trail all contribute to trail utility, increasing choice probability of the trail with the highest value for these features. Conversely, trail fees and the presence of hikers or equestrians negatively affect choice probabilities. Of these variables, only peak numbers and the presence of other trail users significantly affect the choice probability. As we move from MNL to MIXL-I, we see the singletrack mileage become significant at the 10% level. Under MIXL-II both singletrack mileage and the trail fee have a significant effect on choice probabilities.

Comparing Information Criteria across the two models, MIXL-I outperforms MNL on every

Table 4: Out-of-sample performance mountainbike data

	Hit-rate	F-1	f_A	f_B
MNL	55%	-0.283	77%	23%
MIXL	56%	0.097	58%	42%
MIXL-II	55%	0.077	58%	42%

Table 4 reports out-of-sample performance measures for MNL, MIXL-I and MIXL-II models in the mountainbike survey dataset. The hit-rate, as well as an F-1 statistic of out of sample performance are reported. The rightmost columns report for each model the frequency with which each option j is predicted, f_j

account. Reactions to choice attributes differ enough across cyclists to yield an improved model fit with random parameters. This is in line with results presented in Fiebig et al. (2010), who also find stated preference data to benefit in model fit from MIXL over MNL.

5.2 Out of Sample, Mountainbike Data

Table 4 summarizes the out-of-sample performance of MNL, MIXL-I and MIXL-II in predicting trail choice among mountainbikers. We see that the choices are difficult to predict, as each model has a hit-rate of around 55%. A random prediction (drawing predictions by predicting choice i with probability p_i) would yield a hit rate of 50.5%. Although Mixed Logit does not improve on the hit-rate generated through Multinomial Logit, an improvement may be seen in the F-1 statistic. The MNL has a negative F-1 value, realizing its hit rate the 'easy way', by predicting the most common choice 77% of the time. MIXL-I and MIXL-II have small, but positive F-1 statistic, as they realize a similar hit-rate, but do so through more daring predictions, spreading the forecasts more evenly over the two options.

5.3 Full sample, Household Scanner data

We estimate model parameters for MNL and MNL with loyalty, as well as MIXL-I (with independently distributed parameters) with and without loyalty. Table 5 reports the results. Reviewing the parameter estimates, the models largely behave as we would expect from the data. Fixing the ASC for Nabisco to 0, the ASCs are estimated as significantly negative for all other brands, representing their relative unpopularity. Price significantly detriments choice probability for all models estimated, with brand features and store displays improving choice probability.

Comparing MIXL-I without loyalty to basic MNL, we find a clear recurrence of the result presented in Fiebig et al. (2010). MIXL-I outperforms MNL on accounts of all information crite-

Table 5: Full Sample parameter estimates cracker choice

	MNL		MIXL-I		MNL with Loyalty		MIXL-I with Loyalty	
k	7		14		9		17	
AIC	6724.023		3421.130		3354.154		3238.881	
BIC	6760.618		3500.420		3402.948		3336.469	
CAIC	6775.718		3522.520		3420.047		3361.568	
$log(L)$	-3356.012		-1697.565		-1669.077		-1603.441	
Variable	Coefficient	SE/CI	Coefficient	SE/CI	Coefficient	SE/CI	Coefficient	SE/CI
$\alpha_{Kleebler}$	-1.941***	0.072	-3.488***	0.252	-0.433***	0.099	-0.869***	0.164
$\alpha_{Sunshine}$	-2.459***	0.080	-3.543***	0.221	-1.101***	0.108	-1.474***	0.137
$\alpha_{Private}$	-1.814***	0.101	-4.784***	0.384	-1.589***	0.143	-2.523***	0.264
β_{Price}	-0.032***	0.002	-0.051***	0.006	-0.039***	0.003	-0.043***	0.005
$\beta_{Feature}$	0.143**	0.061	0.426***	0.121	0.409***	0.091	0.315**	0.131
$\beta_{Display}$	0.321***	0.100	0.552***	0.178	0.634***	0.146	0.676***	0.188
ϕ					0.784	[0.76 , 0.8]	0.830	[0.78 , 0.86]
$\beta_{Loyalty}$					3.693***	0.087	3.367***	0.205
σ_1			2.269	[1.96 , 2.63]			0.000	-
σ_2			2.195	[1.78 , 2.71]			0.394	[0.16 , 0.96]
σ_3			0.912	[0.53 , 1.56]			0.000	-
σ_4			3.893	[3.38 , 4.48]			1.366	[0.99 , 1.89]
σ_5			0.049	[0.04 , 0.06]			0.030	[0.02 , 0.04]
σ_6			0.614	[0.34 , 1.11]			0.721	[0.47 , 1.11]
σ_7			0.381	[0.03 , 4.36]			0.763	[0.38 , 1.55]
σ_8							1.588	[1.12 , 2.25]

Table 5 reports parameter estimates for MNL and MIXL-I models, both with and without brand loyalty as a covariate. AIC, BIC and CAIC values are reported, as well as the log-likelihood for each estimation. σ and ϕ are estimated in a transformed manner to map their estimated values to their proper ranges of $[0, Inf)$ and $[0, 1]$ respectively. For those transformed variables, 95% confidence intervals are extrapolated from their untransformed estimates, rather than standard errors. Significance is denoted by *** (1%), ** (5%) and * (10%), respectively.

ria, by realising a two-fold improvement in log-likelihood at the cost of 7 additional parameters. Behaviorally, this indicates that individuals exhibit distinct reactions to marketing variables, as well as strong brand loyalty behaviors. The strongly variable ASCs indicate that there is significant unobserved heterogeneity in baseline product utility across customers. Intuitively this is quite sensible, as some customers may derive utility just from being used to a particular brand of cracker, or from the ease of not deliberating brand choice on each supermarket visit.

Comparing the MNL and MIXL-I models with the Guadagni and Little (1983) loyalty variable included, we see similar results. In the MNL model, ϕ is estimated at 0.784. Revisiting the examples presented in Figure 1, we may recall that this corresponds to a loyalty variable that takes time to build up to a high value over the purchase history, but decreases little from momentary deviations in brand choice. This corresponds to strong loyalty behaviors and a lower switching propensity.

An intermediate purchase at time t of Sunshine after a long history of purchasing Nabisco only decreases the Nabisco Loyalty variable by 22%. For the choice made at time $t + 1$, this means that greater utility from purchase history is assigned to the brand chosen abundantly in the period $[1, t - 1]$, than to the previously chosen brand in period t. This is also seen when we estimate an MNL model using lagged indicator variables as regressors (the specific case of $\phi = 0$), which is estimated with a log-likelihood of 2254 and an AIC of 4492. Large gains in model fit may be made from including long-term purchase history over recent purchase history. $\beta_{loyalty}$ is estimated as 3.693 and is strongly significant with a standard error of 0.087.

Table 6: Mixed Logit parameter estimates with non-diagonal Σ

		$log(L)$:	-1658	BIC:	3439	AIC:	3386	CAIC:	3474
Parameter	Coefficient	SE	ρ_{i1}	ρ_{i2}	ρ_{i3}	ρ_{i4}	ρ_{i5}	ρ_{i6}	ρ_{i7}
$\alpha_{nabisco}$	-	-	1.00	0.39	0.84	0.43	-0.22	0.41	0.66
$\alpha_{kleebler}$	-3.456***	0.297		1.00	0.39	0.13	-0.29	0.28	-0.10
$\alpha_{sunshine}$	-3.470***	0.235			1.00	0.54	0.28	-0.11	0.51
$\alpha_{private}$	-4.562***	0.358				1.00	0.49	-0.30	0.47
β_{price}	-0.045***	0.006					1.00	-0.97	-0.02
$\beta_{feature}$	0.337**	0.135						1.00	0.20
$\beta_{display}$	0.621***	0.183							1.00

Table 6 reports estimation results for Mixed Logit with unrestricted Σ. Information criteria and log-likelihood are reported, as well as parameter estimates and standard errors. Concomitantly, a correlation matrix is supplied, where ρ_{ij} reports the correlation between parameters i and j, in the order of appearance in the left-most column. Significance is denoted by *** (1%), ** (5%) and * (10%), respectively.

Comparing MNL and MIXL-I with the loyalty variables included, we find MIXL-I outperforming MNL on all accounts. AIC, BIC and CAIC all favor MIXL-I, even in a setting which already explicitly models dynamics. This suggests that the improved fit is not simply a results of modeling unobserved brand preferences through heterogeneous ASCS, but that there is also a distinct difference among consumers in reactions to price, promotions and store displays.

We may note, however, that the improvement when moving from fixed to random parameters is less dramatic than in the setting without the loyalty variables. Whilst we saw a change of 49.5% previously, we now see a decrease of 4.0% in log likelihood. As the benefit of mixed logit is partially swallowed by an explicit consideration for loyalty, the differences between the two models become smaller.

We see this also in $\sigma_1, \sigma_2, \sigma_3, \sigma_4$, which denote the standard deviation of the ASC for Nabisco, Kleebler, Sunshine and Private, respectively. Initially estimated to be significant and quite large relative to their parameters, they are estimated at 0 for two of the brands in MIXL-I with loyalty,

Table 7: Mixed Logit parameter estimates with non-diagonal Σ and loyalty

	LogL:	-1553	BIC:	3264	AIC:	3195	CAIC:	3309	ϕ:	0.86
Parameter	Coefficient	SE	ρ_{i1}	ρ_{i2}	ρ_{i3}	ρ_{i4}	ρ_{i5}	ρ_{i6}	ρ_{i7}	ρ_{i8}
$\alpha_{nabisco}$	-	-	1.00	0.46	0.98	0.47	-0.12	0.73	0.30	-0.02
$\alpha_{kleebler}$	-1.318***	0.224		1.00	0.35	0.64	-0.45	0.90	0.37	-0.21
$\alpha_{sunshine}$	-2.047***	0.250			1.00	0.49	-0.09	0.68	0.32	-0.06
$\alpha_{private}$	-2.741***	0.369				1.00	0.14	0.63	0.49	-0.47
β_{price}	-0.049***	0.008					1.00	-0.56	-0.06	-0.13
$\beta_{feature}$	0.090	0.171						1.00	0.40	-0.16
$\beta_{display}$	0.464*	0.267							1.00	-0.21
$\beta_{loyalty}$	3.611***	0.203								1.00

Table 7 reports estimation results for Mixed Logit with unrestricted Σ and the inclusion of a loyalty variable. Information criteria and log-likelihood are reported, as well as parameter estimates and standard errors. Concomitantly, a correlation matrix is supplied, where ρ_{ij} reports the correlation between parameters i and j, in the order of appearance in the left-most column. Significance is denoted by *** (1%), ** (5%) and * (10%), respectively.

and become much less pronounced for the other two. As we account for brand preference through purchase history, ASC variability seems to lose its added value.

σ_8, which denotes the standard deviation of the $\beta_{loyalty}$ parameter, shows a 95% confidence interval of [1.12, 2.25]. This indicates that loyalty behaviors are also heterogeneous among households, meaning some consumers derive large amounts of utility from purchasing the same product they did previously, whilst others may be more fickle and shop based on marketing variables, rather than purchase history.

Table 6 reports estimation results for MIXL-II. The appendix reports the full covariance matrix. We see another performance improvement, relative to MIXL-I. Estimating parameter covariances on top of parameter heterogeneity yields a lower AIC, BIC and CAIC. Despite requiring the estimation of 35 parameters, the improvement in fit outweighs parsimony issues, according to the information criteria.

Observing the correlation matrix in Table 6, we see that surprisingly, $\alpha_{nabisco}$ negatively correlates with β_{price}, despite being the more expensive brand. This may be explained by some part of the population finding nabisco the superior brand, but only purchasing it on the event of a discount. Conversely, the cheapest, private brand positively correlates with β_{price}, making consumers who prefer the cheapest brand less price-sensitive. This could indicate a subset of the population who generally prefer the cheapest brand, despite fluctuations in the prices of other brands.

Sensitivity to one marketing variable (price, feature or display), generally seem to indicate greater sensitivity to all marketing variables. Especially for price and feature sensitivity, which are estimated to be almost perfectly correlated.

Table 7 reports estimation results for MIXL-II with the inclusion of brand-specific loyalty variables. MIXL-II with loyalty outperforms MIXL-I and MNL with loyalty on all information criteria. Despite being the model with the greatest number of parameters required at 45, it fits the observed choices closely enough to yield lower BIC, AIC and CAIC values.

ϕ is estimated at 0.86, higher than in both the MNL and MIXL-I models. This indicates very strong loyalty variables, with low switching behaviors. Loyalty under MIXL-II takes long to build up, but once a customer has revealed a brand preference, temporary deviations do little to detriment his loyalty to his favorite brand. Households are likely to return to their most-bought brand, even if they have not done so recently.

The appendix reports the full covariance matrix. The correlations observed in Table 6 generally persist. $\beta_{loyalty}$ negatively correlates with all other parameters. This indicates that brand features and displays have less influence on brand-loyal customers, but also that loyal customers are more price sensitive than others. This would suggest that households may deviate from their preferred brand for large discounts in other brands, but return when no such offers are being held.

5.3.1 Latent Class Model

Table 8 reports parameters from the estimation of a Latent Class model, using two classes. A Guadagni and Little (1983) loyalty variable is included for each alternative. The LC model outperforms MNL with loyalty as per the AIC, but fails to outperform MNL on the BIC and CAIC measures. This indicates that even though we observed strong improvements in model performance through the inclusion of parameter heterogeneity, a discrete modeling of consumer preferences poorly represents the household shopping data. Brand preferences, loyalty behaviors and reactions to marketing attributes seem to be more aptly modeled under an assumption of normality, than under the assumption of a finite number of preference clusters.

The clusters identified by the algorithm vary greatly in their alternative-specific constants. The first class is more sensitive to the display and feature marketing variables, whilst showing less pronounced brand constants. The second class is weighed heavily towards Nabisco. Loyalty variables are insignificantly different across the two classes.

However, we should note that the Latent Class likelihood surface is prone to local maxima

Table 8: Latent Class MLE results

	Class 1		Class 2	
Parameter	Estimate	SE/CI	Estimate	SE/CI
$\beta_{kleebler}$	0.279	0.325	-1.300***	0.455
$\beta_{sunshine}$	-0.593*	0.321	-1.901***	0.366
$\beta_{private}$	-0.691	0.705	-2.824***	0.541
β_{price}	-0.033***	0.010	-0.048***	0.011
$\beta_{feature}$	1.874***	0.670	-1.379**	0.674
$\beta_{display}$	1.141**	0.551	0.075	0.439
$\beta_{loyalty}$	3.342***	0.401	3.486***	0.270
ϕ	0.843	[0.78, 0.87]	0.756	[0.72, 0.78]
p_c	0.620		0.380	
Log-Likelihood	-1641		AIC	3319
BIC	3436		CAIC	3435

Table 8 reports parameter estimates and standard errors for a latent class model with C = 2. For ϕ, a transformation is used to map an estimated parameter to the proper range [0, 1] of ϕ. As such, direct standard errors were not generated. As an alternative, a 95% confidence interval is presented for the estimates of ϕ. Significance is denoted by *** (1%), ** (5%) and * (10%), respectively.

that may cause the EM-algorithm to converge outside the global maximum. Although care was taken to perform the algorithm 20 times with different random initializations of $\pi_{i,c}$, this does not guarantee a global maximum. Since the estimation of the class-specific parameter ϕ_c makes the evaluation of the likelihood more computationally expensive (as compared to a choice models without loyalty), a greater number of random draws proved to be beyond the resources of this study. Future research may focus on investigating discrete heterogeneity in a dynamic context with more thoroughness in this area.

5.3.2 Parameter Distributions under MIXL

To ascertain the distribution of household-specific attribute coefficients and brand preferences, we may calculate the conditional expectations of γ_i for each household. Figures 2 to 7 report histograms of $\beta_{i,price}$ under MIXL-I and MIXL-II, as well as of $\beta_{i,loyalty}$ and $\alpha_{i,nabisco}$. For the price parameter under MIXL-I and MIXL-II, we see that the distribution of the parameter is slightly skewed to the left. Close to the mean of the parameter of -0.039 we see a large peak, otherwise the parameter seems to be distributed quite normally. Under MIXL-II, we see the parameter take more extreme values. As the parameter becomes subject not only to its own normal distribution, but also the covariances with other parameters, more extreme values become feasible. Across both histograms, discrete clusters (in the form of separated peaks) are

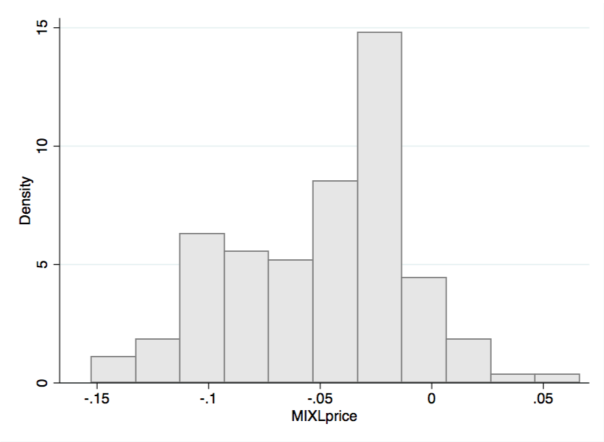


Figure 2: β_{price} distribution MIXL-I

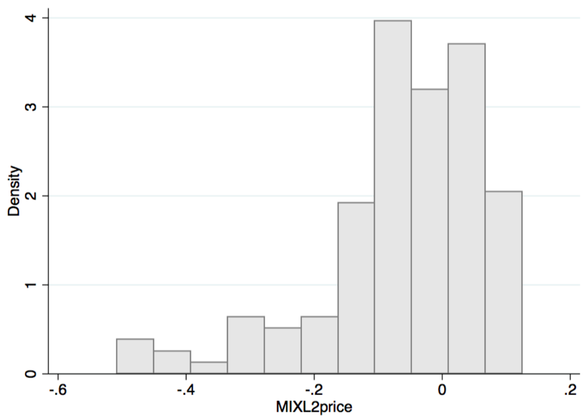


Figure 3: β_{price} distribution MIXL-II

not readily apparent.

Figure 4 and 5 show the variety of brand preference in the dataset. Taken from the estimation without loyalty variables, the two histograms show the variety of alternative-specific constants across households. As every household exhibits some preference, the tails of the distribution are fat, with the mean parameter estimate of 0 appearing no more frequently than values even on the edges of the distribution. As the distribution is not single-peaked, clustering is feasible. This is consistent with the α_j parameters being the most variable in the Latent Class analysis presented in table 8.

$\beta_{i,loyalty}$, shown in figures 6 and 7 appears to be distributed nearly normal, with a negative skew. Nearly all households exhibit some consideration for their purchase history, with a few outliers which approach 0. These do not recur under MIXL-II, where all households have $\beta_{i,loyalty} > 2$. Here too, parameter clustering does not readily appear, which may lend some explanation to the poor performance of the Latent Class model.

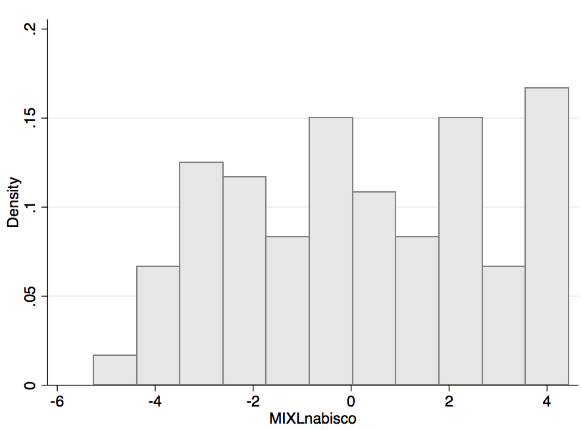


Figure 4: $\alpha_{Nabisco}$ distribution MIXL-I

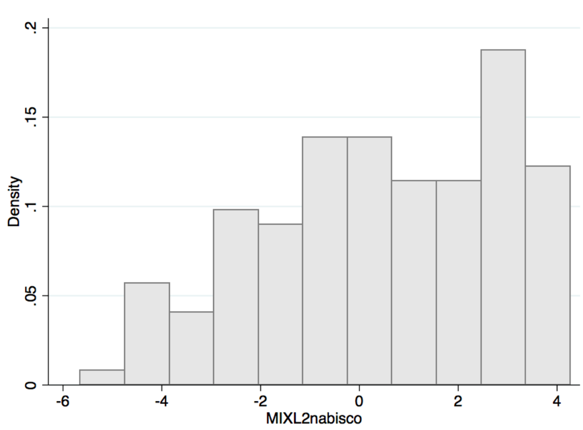


Figure 5: $\alpha_{Nabisco}$ distribution MIXL-II

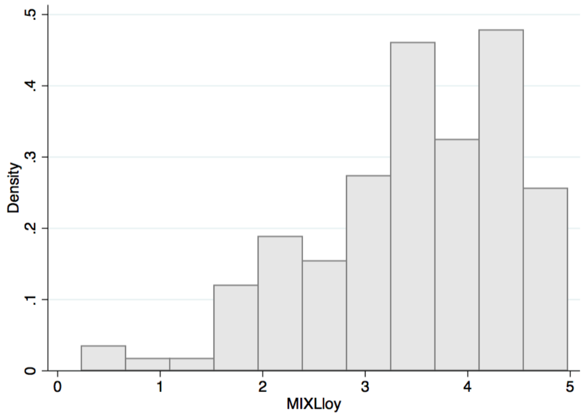


Figure 6: $\beta_{loyalty}$ distribution MIXL-I

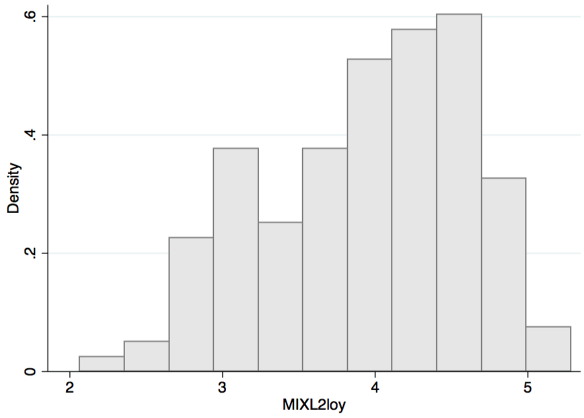


Figure 7: $\beta_{loyalty}$ distribution MIXL-II

5.4 Out-of-sample performance, household scanner data

Table 9: Forecasting performance cracker data

Model	$\sum_{j=1}^{J} p_{jj}$	F-1	$p_{.kleebler}$	$p_{.nabisco}$	$p_{.private}$	$p_{.sunshine}$
MNL	57.0%	-0.32	0.0%	79.6%	20.4%	0.0%
MIXL-I	81.8%	0.67	4.5%	53.7%	39.2%	2.6%
MIXL-II	80.3%	0.65	4.5%	54.9%	37.5%	3.1%
MNL (loyalty)	85.5%	0.73	3.3%	56.1%	36.9%	3.7%
MIXL-I (loyalty)	85.7%	0.74	3.6%	55.6%	37.7%	3.1%
MIXL-II (loyalty)	85.8%	0.75	3.9%	54.2%	37.8%	4.2%
LC (loyalty)	69.8%	0.49	2.0%	76.4%	20.1%	1.3%

Table 9 reports out-of-sample performance measures for MNL, MIXL-I, MIXL-II and the Latent Class model. Hit-rates are of correct predictions are reported, as well as an F-1 statistic and prediction frequencies for each model.

Table 9 reports out-of-sample performance for MNL and MIXL-I, both with and without loyalty variables included. For basic MNL, we find that the model is able to predict the purchase on 57% of choice occasions. It does so by predicting Nabisco as a choice 79.6% of the time, and Private otherwise. Forecasting performance in the marketing setting is much better than in the mountain bike dataset, as a random prediction, based on in-sample choice frequencies would realize a hit-rate of 40.3%. Using the conditional expectation of γ_i for each individual for MIXL-I and MIXL-II forecasts, much higher hit rates of over 80% are realized. Observing brand preferences and price reactions by individuals in the first 12 observations yield large improvements in our ability of predicting cracker choice in future store visits. F-1 statistics for MIXL-I and MIXL-II are quite similar, suggesting that for forecasting, independently distributed parameters suit the cracker dataset just as well as covarying parameters.

Adding the loyalty covariates increases performance for both MNL and MIXL. MNL now makes more diverse predictions, predicting Kleebler and Sunshine on a respective 3.3% and 3.7% of choice occasions. A higher hit-rate of 85.5 % is realised, as well as an improved F-1 statistic. As in the full-sample performance measures, MNL and MIXL performance is much closer when dynamics are accounted for. MIXL, for both diagonal or dense Σ, realizes a slightly higher hit-rate than MNL, with slightly higher F-1 measures also. Although the increased performance of MIXL over MNL extends to their forecasting abilities, the differences are very slight once dynamics are accounted for.

The Latent Class model (with two classes), with loyalty variables included, performs very poorly relative to the other models. The classes identified in the first 12 observations lead the model astray in forecasting the subsequent store visits. The forecast profile is quite similar to basic MNL, predicting mostly Nabisco and Private.

5.5 Tracing MIXL performance

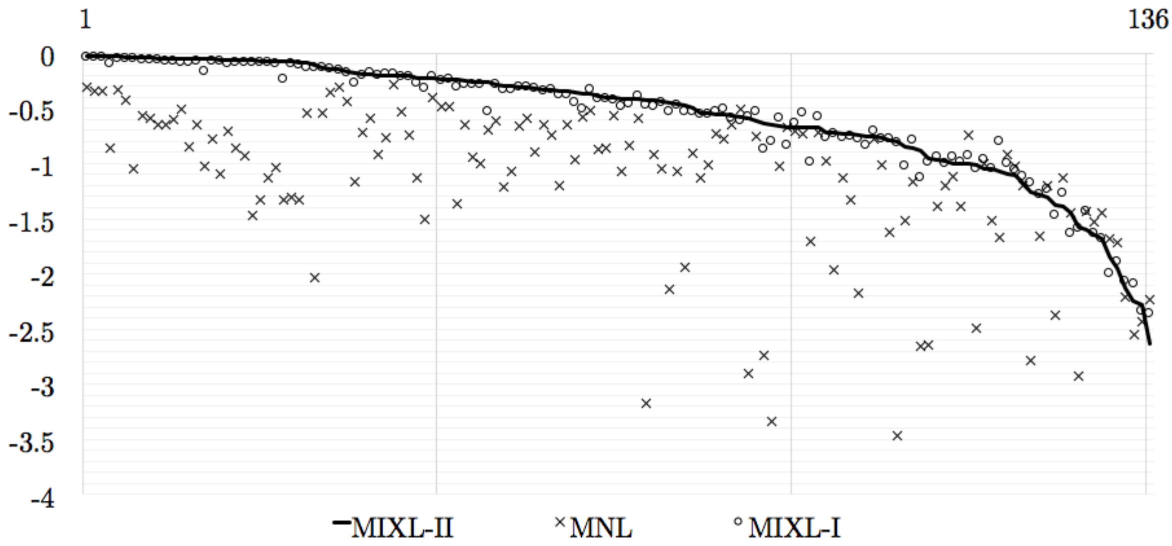


Figure 8: Observation-average log-likelihood contributions per household

Figure 8 reports the observation-average log-likelihood contributed to the total model log-likelihood per household, for the full sample of Cracker purchase data. Households are sorted left to right by their log-likelihood contribution to the MIXL-II model, represented by the black line. MNL (circles) and MIXL-I(crosses) log-likelihood contributions are graphed concomitantly.

To ascertain why Mixed Logit outperforms MNL on all accounts, we may consider the specific households in which the largest differences in model fit are realised. Figure 8 plots household log-likelihood under MNL, MIXL-I and MIXL-II (all without loyalty variables included), sorted by their observation-average log-likelihood values for MIXL-II. From left to right, this shows us the households who have near perfect fits under MIXL-II, to those which MIXL-II is unable to fit at all on the far right.

We see that MIXL-I and MIXL-II log-likelihood values are very similar. MNL log-likelihood values are much more variable. Although there exists some correlation between MNL household log-likelihood values and MIXL values, MNL fit for the middle third of the graph is not much better than the latter third. In other words, MNL fit is consistently poor after the first 45 households, and not much better for those first 45, as sorted by MIXL-II log-likelihood values.

To identify behavioral differences among the households with near-perfect MIXL-II fit (the left-most 45 households in figure 8) and the worst-fitting households (the right-most 46 households in figure 8), we estimate an MNL model with loyalty variables on each subsample separately.

Table 10 reports the results.

The most prominent difference can be seen in $\beta_{loyalty}$. Whilst the two subsamples are estimated to have similar values for ϕ, $\beta_{T1,loyalty}$ is estimated at 5.14, whilst $\beta_{T3,loyalty}$ is estimated at 0.25. Clearly the first subsample is much more brand-loyal than the latter. As MIXL-II is able to fit household-specific values to $\alpha_{i,j}$ for each brand, such loyalty behaviors may be accounted for. Under MNL, these individuals are subject to the same α_j as the whole population. Seeing as ϕ is estimated to be very close between the two subsamples, it seems that the latter group is not more fickle with their brand preferences, they simply do not attach value to their own purchase history.

Considering the marketing variables, T3 has positive estimates for $\beta_{feature}$ and $\beta_{display}$, whilst T1 does not. This suggests that marketing variables affect T1 more than they do T3, which is in line with the observed loyalty behaviors.

Table 10: Type I and Type III MNL estimates with loyalty

Parameter	T1	SE	T3	SE
$\beta_{kleebler}$	-1.574	0.418	-1.192	0.120
$\beta_{sunshine}$	-2.429	0.527	-1.105	0.116
$\beta_{private}$	-3.927	0.637	-0.634	0.140
β_{price}	-0.079	0.014	-0.021	0.003
$\beta_{feature}$	-0.287	0.429	0.404	0.091
$\beta_{display}$	-0.244	0.758	0.601	0.157
$\beta_{loyalty}$	5.143	0.331	0.249	0.079
ϕ	0.774		0.802	

Table 10 reports MNL parameter estimates for the disjoint subsamples T1 and T3. A Guadagni and Little (1983) loyalty variable is included for each choice alternative. T1 includes the 45 households with the highest MIXL-II log-likelihood values, T3 the 46 the households with the lowest MIXL-II log-likelihood values.

5.6 Willingness to Pay analysis

To shed more light on the relative importance of household purchase histories, we re-estimate the MIXL-I model with and without loyalty, in Willingness to Pay (WTP) space. This means that we fix $\beta_{price} = -0.01$ and leave all other parameters unrestricted. The subsequent estimated values for β_k, with k some choice characteristic, may then be interpreted as a household's willingness to pay for that characteristic, as the price of alternative j would have to go down by 100 cents to yield an equivalent benefit to the utility of alternative j as β_k.

For the display and feature parameters, this is straightforward. If we see a $\beta_{feature}$ of 1.5 in

WTP space, this mean that the household would pay \$1.50 more to enjoy the same utility from a featured product as they do from an unfeatured, cheaper one. For the α_j parameters, we need to consider differences between alternatives. If α_j is strongly negative in WTP space, but still higher than the $\alpha_{j'}$ parameters of the other alternatives, we estimate a positive WTP for brand j over the others.

For the loyalty variables, the process of obtaining WTP values is twofold. Firstly, we observe the estimated value for ϕ. From this, we know that purchasing brand j at time $t-1$ increases the loyalty variable $b_{i,j,t}$ by $(1-\phi)$, yielding an increase in Utility of $\beta_{loyalty}(1-\phi)$ from purchasing that particular brand at time t-1. By observing the distribution of $\beta_{loyalty}$, we may calculate the distribution of WTP increases from brand j being the last chosen brand.

Table 11 reports the result, showing the mean and tenth to ninetieth percentiles of the distribution of each parameter under MIXL-I, as well as under MIXL-I with loyalty. As per MIXL-I, Nabisco is valued on average as being worth \$3.94 more to the households in our sample. This is a striking difference with the MNL estimate in WTP space, which only reports a \$0.87 difference. The majority of households has a positive WTP for featured and displayed products, of \$0.70 and \$0.99 on average. We may note a large difference in the MNL WTP estimates of \$0.26 and \$0.53. As brand preference is unaccounted for under MNL, the marketing variables are estimated to be less valuable than they really are. After controlling for brand loyalty, as also seen in the MNL estimate with loyalty included, we find they do much more to increase household's WTP for a product.

In MIXL-I with loyalty, we see that the alternative-specific constants are much more equal across the population, seeing as loyalty behaviors are specifically accounted for. We find that the purchase of a brand at time $t-1$ leads to an increase in WTP for that brand at time 1 of \$0.53 on average. As $\beta_{loyalty}$ is also heterogeneous, this may range from \$0.24 to up to \$0.81 for the 10% most loyal households. Seeing as the mean price of a box of Nabisco crackers is \$1.07, this means that 90% of households are willing to pay at least 22% more for Nabisco if they purchased it on their last store visit.

Table 11: Willingness to Pay distributions for choice attributes

Model	WTP for	Mean	P_{10}	P_{25}	P_{50}	P_{75}	P_{90}	MNL estimate
MIXL	Nabisco	-0.56	-3.99	-2.36	-0.56	1.25	2.88	0.48
	Kleebler	-3.29	-5.08	-4.24	-3.29	-2.35	-1.50	-1.48
	Sunshine	-3.47	-4.40	-3.96	-3.47	-2.98	-2.53	-1.61
	Private	-4.50	-8.84	-6.78	-4.50	-2.21	-0.15	-0.39
	Feature	0.70	-0.06	0.30	0.70	1.11	1.47	0.26
	Display	0.99	-0.12	0.40	0.99	1.57	2.09	0.53
MIXL with loyalty	Nabisco	0.16	0.16	0.16	0.16	0.16	0.16	-0.56
	Kleebler	-0.92	-1.85	-1.41	-0.92	-0.43	0.01	-1.04
	Sunshine	-1.17	-1.17	-1.17	-1.17	-1.17	-1.17	-1.18
	Private	-1.82	-3.70	-2.81	-1.82	-0.83	0.06	-0.97
	Feature	0.47	-0.40	0.01	0.47	0.92	1.33	0.57
	Display	0.86	-0.01	0.40	0.86	1.32	1.74	0.94
	Chosing j at t-1	0.53	0.24	0.38	0.53	0.68	0.81	0.78

Table 11 reports MIXL-I estimates of the distribution of the Willingness to Pay for choice characteristics. With β_{price} fixed to -0.01, the mean parameter values are given for each choice characteristic, as well as percentile values $P_{.}$ of the distribution of the parameters. In the right-most column, we report the estimated parameter value under MNL in WTP space.

6 Discussion and Conclusion

Under the choice framework presented by McFadden (1973), the Multinomial Logit model, discrete choices are made by individuals, according to some "common behavioral rule". Through other models that cast aside this assumption, we show that this axiom fails to hold up in a marketing setting, as well as in a discrete choice experiment among mountainbikers. As in Fiebig et al. (2010), we find sufficient preference heterogeneity for the MIXL model to outperform MNL on in-sample information criteria, and out-of-sample forecasting performance. We extend this result to dynamic choice models, finding enough residual preference heterogeneity for MIXL to still outperform MNL, even after brand preferences are accounted for.

A Latent Class model with two classes reports unconvincing results in and out of sample. Observation of posterior parameter estimates under MIXL also yields little reason to believe that consumers exhibit discrete choice heterogeneity. For a marketing purpose, we recommend a MIXL model with a continuous parameter distribution.

We find that for conventional MNL and MIXL-I, the largest improvement is seen in households who are particularly brand loyal. The smallest difference in performance for the two models is found for households who show less consideration for their purchase history, and a higher sensitivity to marketing variables.

In estimating the model in WTP space, we find that households are willing to pay $0.53 more on average for a brand which they also purchased on their last store visit. This puts the benefit of product sampling on par with store displays and brand features. We also find that WTP estimates for marketing variables are higher once choice dynamics are accounted for. The use of the basic MNL model in ascertaining return on investment for marketing efforts may lead marketeers astray if preference heterogeneity is ignored.

Future research may give more elaborate consideration to the Latent Class model. A large number of classes, as well as more thorough efforts to ensure the EM-algorithm reaches a global maximum proved beyond the resources of this research. Secondly, the performance of MIXL in a dynamic choice setting could be further explored. We may consider whether MIXL models with loyalty could be made more parsimonious by restricting the variances of the alternative specific constants to 0, as brand preference heterogeneity is already accounted for through loyalty behaviors. Moreover, we are curious to see whether the importance of scale heterogeneity over preference heterogeneity found by Fiebig et al. (2010) extends to the dynamic setting.

References

Currim, I. S. (1982). Predictive testing of consumer choice models not subject to independence of irrelevant alternatives. *Journal of Marketing Research*, 19(2):208–222.

Fiebig, D., Keane, M., Louviere, J., and Wasi, N. (2010). The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. 29:393–421.

Greene, W. H. and Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37(8):681–698.

Guadagni, P. M. and Little, J. D. C. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science*, 27(1):29–48.

Jain, D. C., Vilcassim, N. J., and Chintagunta, P. K. (1994). A random-coefficients logit brand-choice model applied to panel data. *Journal of Business and Economic Statistics*, 12(3):317–328.

Kamakura, W. and Russell, G. (1989). A probabilistic choice model for market segmentation. 26.

Lim, J., Currim, I. S., and Andrews, R. L. (2005). Consumer heterogeneity in the longer-term effects of price promotions. *International Journal of Research in Marketing*, 22(4):441 – 457.

Luce, D. (1959). Individual choice behavior: A theoretical analysis. *Journal of the American Statistical Association*, 115:239.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press New York, New York, NY, USA.

McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470.

Morey, E. R., Buchanan, T., and Waldman, D. M. (2002). Estimating the benefits and costs to mountain bikers of changes in trail characteristics, access fees, and site closures: choice experiments and benefits transfer. *Journal of Environmental Management*, 64(4):411 – 422.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Veall, M. R. and Zimmermann, K. F. (1992). Performance measures from prediction- realization tables. *Economics Letters*, 39(2):129 – 134.

A. Mountainbike data

Table 12: Descriptive Statistics Mountain Biking Data

Variable	Min	Max	Mean	St.Dev	N
Cyclist ID	1.0	289.0	144.9	83.5	1420
Age	10.0	69.0	29.8	10.0	1400
Gender	0.0	2.0	0.8	0.4	1367
Mountain Biking Experience	1.0	3.0	2.0	0.7	1415
Is a Mountainbiker	0.0	1.0	0.6	0.5	1415
Respondent has raced	0.0	1.0	0.4	0.5	1420
NORBA racing category	0.0	3.0	0.4	0.9	1365
USCF racing category	0.0	7.0	0.3	1.0	1395
Days spent Mountain Biking per season	0.0	7.0	4.1	1.7	1299
Considers Mountain Biking training	0.0	1.0	0.5	0.5	1357
Suspension bike	0.0	1.0	0.4	0.5	1362
Clipless pedals	0.0	1.0	0.3	0.5	1357
Cost of bike	0.0	4000.0	832.7	659.8	1332
Age of bike	0.0	11.0	2.7	2.1	1346
Married	0.0	2.0	0.6	0.5	1386
Children	0.0	10.0	0.6	1.1	1371
Household Spending	13.0	2200.0	908.0	682.6	1297
Hourly Wage	0.2	150.0	16.0	12.9	1192

B. MIXL-II covariance matrices

Table 13: MIXL-II covariance matrix cracker data

	$\alpha_{nabisco}$	$\alpha_{kleebler}$	$\alpha_{sunshine}$	$\alpha_{private}$	β_{price}	$\beta_{feature}$	$\beta_{display}$
$\alpha_{nabisco}$	0.04***(6.8E-5)						
$\alpha_{kleebler}$	0.115***(3.9E-3)	1.552***(2.0E-2)					
$\alpha_{sunshine}$	0.059(3.0E-4)	0.133***(1.1E-2)	0.091(7.9E-4)				
$\alpha_{private}$	0.783(2.7E-3)	6.738***(2.7E-3)	1.25***(2.8E-4)	70.602***(2.7E-2)			
β_{price}	-0.053(6.2E-3)	-1.209(1.7E-3)	-0.061(5.6E-4)	2.47***(7.5E-3)	4.56(1.5E-5)		
$\beta_{feature}$	0.166***(3.0E-5)	1.264***(5.8E-3)	0.231***(3.2E-4)	5.975(2.5E-3)	-1.363(2.3E-4)	1.284(2.5E-3)	
$\beta_{display}$	0.085(7.4E-2)	0.65***(2.0E-1)	0.136(8.3E-2)	5.924(1.2E+1)	-0.182(8.9E-2)	0.648(4.4E+0)	2.03***(2.8E-1)

Table 13 reports parameter variances and covariances for the MIXL-II model estimated on the full sample of household shopping data. Standard errors, given in parentheses, are obtained through a finite-difference approximation of the hessian, using step-size 10^{-7}. The approximated hessian is inverted, such that the diagonal elements of the inverse give the approximate standard errors. Significance is denoted by *** (1%), **(5%) and *(10%).

Table 14: MIXL-II covariance matrix cracker data with loyalty

	$\alpha_{nabisco}$	$\alpha_{kleebler}$	$\alpha_{sunshine}$	$\alpha_{private}$	β_{price}	$\beta_{feature}$	$\beta_{display}$	$\beta_{loyalty}$
$\alpha_{nabisco}$	0.04							
$\alpha_{kleebler}$	0.12	1.55						
$\alpha_{sunshine}$	0.06	0.13	0.09					
$\alpha_{private}$	0.78	6.74	1.25	70.60				
β_{price}	-0.05	-1.21	-0.06	2.47	4.56			
$\beta_{feature}$	0.17	1.26	0.23	5.98	-1.36	1.28		
$\beta_{display}$	0.08	0.65	0.14	5.92	-0.18	0.65	2.03	
$\beta_{loyalty}$	-0.01	-0.47	-0.03	-6.95	-0.47	-0.32	-0.53	3.05

Table 14 reports estimated variances and covariances of the MIXL-II model with loyalty variables included, estimated on the full sample of household shopping data.

Table 15: MIXL-II covariance matrix mountainbike data

	$\beta_{mileage}$	$\beta_{ft.climbing}$	$\beta_{no.peaks}$	$\beta_{singletrack}$	$\beta_{sharedtrack}$	β_{fee}
$\beta_{mileage}$	0.001					
$\beta_{ft.climbing}$	-0.001	0.004				
$\beta_{no.peaks}$	0.005	-0.016	0.617			
$\beta_{singletrack}$	-0.001	0.000	0.569	0.923		
$\beta_{sharedtrack}$	0.006	-0.022	1.398	1.241	3.472	
β_{fee}	0.002	-0.007	0.382	0.129	1.181	0.732

Table 15 reports the estimated covariance matrix for MIXL-II estimated on the mountainbike dataset.

C. Prediction-Realisation tables

Table 16: Prediction-Realization tables cracker data

				Realized		
Predicted	MNL	kleebler	nabisco	private	sunshine	F_1:0.27
	kleebler	0	0	0	0	0.0%
	nabisco	76	768	406	72	79.6%
	private	23	113	179	23	20.4%
	sunshine	0	0	0	0	0.0%
		6.0%	53.1%	35.2%	5.7%	57.0%
				Realized		
Predicted	MIXL-I	kleebler	nabisco	private	sunshine	F_1:0.69
	kleebler	44	16	11	4	4.5%
	nabisco	34	771	56	30	53.7%
	private	20	87	513	31	39.2%
	sunshine	1	7	5	30	2.6%
		6.0%	53.1%	35.2%	5.7%	81.8%
	MIXL-II			Realized		
Predicted	MNL	kleebler	nabisco	private	sunshine	F_1:0.66
	kleebler	46	14	10	4	4.5%
	nabisco	34	768	65	44	54.9%
	private	19	86	495	23	37.5%
	sunshine	0	13	15	24	3.1%
		6.0%	53.1%	35.2%	5.7%	80.3%
				Realized		
Predicted	MNL-loy	kleebler	nabisco	private	sunshine	F_1:0.75
	kleebler	47	5	1	1	3.3%
	nabisco	37	808	58	29	56.1%
	private	13	58	520	21	36.9%
	sunshine	2	10	6	44	3.7%

		6.0%	53.1%	35.2%	5.7%	85.5%

				Realized		
Predicted	MIXL-I loy	kleebler	nabisco	private	sunshine	F_1:0.76
	kleebler	51	5	1	2	3.6%
	nabisco	32	804	51	36	55.6%
	private	15	64	529	18	37.7%
	sunshine	1	8	4	39	3.1%
		6.0%	53.1%	35.2%	5.7%	85.7%

				Realized		
Predicted	MIXL-II loy	kleebler	nabisco	private	sunshine	F_1:0.76
	kleebler	52	6	4	2	3.9%
	nabisco	29	797	47	26	54.2%
	private	17	66	527	18	37.8%
	sunshine	1	12	7	49	4.2%
		6.0%	53.1%	35.2%	5.7%	85.8%

				Realized		
Predicted	LC loy	kleebler	nabisco	private	sunshine	F_1:0.49
	kleebler	24	8	1	0	2.0%
	nabisco	68	835	295	71	76.4%
	private	7	33	286	8	20.1%
	sunshine	0	5	3	14	1.3%
		6.0%	53.1%	35.2%	5.6%	69.8%

Table 16 presents prediction/realisation table for MNL and MIXL-I both with and without loyalty variables included. Concomitantly, hit-rates are presented, as well as an F_1 measure of out-of-sample fit, computed as $F_1 = \frac{\sum_{j=1}^{J} p_{jj} - p_{.j}^2}{1 - \sum_{j=1}^{J} p_{.j}^2}$, where p_{ij} is the fraction of all out-of-sample observations where i was observed and j predicted. Column indexes denote realized values, whilst row indexes denote predicted values.

Table 17: Prediction-Realization tables Mountainbike data

MNL		Realized		
		B	A	
Predicted	B	30	31	22.7%
	A	90	118	77.3%
		44.6%	55.4%	55.0%
MIXL-I		Realized		
		B	A	
Predicted	B	57	55	41.6%
	A	63	94	58.4%
		44.6%	55.4%	56.1%
MIXL-II		Realized		
		B	A	
Predicted	B	56	57	42.0%
	A	64	92	58.0%
		44.6%	55.4%	55.0%

Table 17 reports predictions-realisation tables for 269 out-of-sample observations in the mountainbike dataset. Hit rates are given in the bottom left for each table.