Juha Nuutinen

Segmentation of Consumers in Retailing of Probiotics - A Latent Class Approach

Mat-2.4108 Independent Research Project in Applied Mathematics

Helsinki, March 18, 2012

Supervisor: Professor Ahti Salo
# Contents

1 Background

2 Latent Class Models in the Marketing Literature
   2.1 Introduction to Latent Class Models ............................................. 2
   2.2 Applications of Latent Class Models in the Marketing Literature .......... 3

3 Model Specification and Evaluation Methods
   3.1 Within-Class Choice Model ............................................................ 4
   3.2 Class-Allocation Model .................................................................. 4
   3.3 Maximum Likelihood Estimation ...................................................... 6
   3.4 Segment Retention Criteria ............................................................. 6

4 Results and Analysis
   4.1 Determining the Number of Segments .............................................. 9
   4.2 Comparison of Models With and Without Segmentation .................... 12
      4.2.1 Parameter Estimates ................................................................. 12
      4.2.2 Performance of Models ............................................................. 14

5 Conclusions

References

Appendices
   Appendix 1: The Full Notational Convention ........................................ 19
   Appendix 2: Choice Model Specification .............................................. 21
1 Background

The latent class model incorporates heterogeneity across consumers by identifying latent consumer segments with heterogeneous preferences (Kamakura and Russell 1989, Chintagunta et al. 1991, Bucklin and Gupta 1992, Chintagunta 1993). The latent class model is frequently used in the marketing literature to consumer segmentation. It enables consumers across segments to have different preferences and different responses to marketing stimuli. Identification of these segments is shown to significantly increase the predictive performance of the choice model (e.g., Kamakura and Russell 1989, Fader and Hardie 1996).

In this study, the latent class model is calibrated on point-of-sale data from retailing of probiotics to healthcare personnel in 2008-2009. Two thirds of customers are included in the training set, which is used for calibration of the model. The remaining set of customers comprising the test set is reserved for testing and validation purposes. Loyalty variables are initialized with historical purchases in 2007. The forecasting performance is evaluated by predicting choice decisions in a future period, 2010-2011.
2 Latent Class Models in the Marketing Literature

2.1 Introduction to Latent Class Models

The latent class model assumes that consumers belong to one or several segments. Consumers in each segment have homogeneous preferences toward product attributes and they respond similarly to marketing stimuli. Consumers are assigned to these segments either deterministically or probabilistically based on their observed purchase history, attitudes, stated preferences, socio-demographic background or other covariates.

The latent class analysis typically consists of the following steps. First, the number of classes, or segments, is determined. Then, a class-allocation model is used to determine the probability of an individual being assigned to a specific segment as a function of her socio-demographic characteristics or observed choice history. Simultaneously, a within-class choice model determines the segment-specific choice probabilities for the different alternatives. The resulting choice probability for an alternative is given as a sum of within-class choice probabilities weighted by the class-allocation probabilities for that specific individual.

As an example, Figure 1 illustrates choice probabilities for a two-segment model with three alternatives in a tree structure. This example shows that calculation of the choice probability for an alternative is very straightforward once the class-allocation probabilities and the within-class choice probabilities are determined.
2.2 Applications of Latent Class Models in the Marketing Literature

Kamakura and Russell (1989) use observed purchase histories to probabilistic segmentation of households into homogeneous segments with different intrinsic brand utilities and sensitivities to price changes. Their research pioneered in incorporating consumer heterogeneity in choice models with scanner panel data in explaining brand choice behaviour. Bucklin and Gupta (1992) extend their approach to study both brand choice and category purchase incidence simultaneously while accounting for heterogeneity across consumers using the latent class approach.

Fader and Hardie (1996) propose the full use of product attributes to describe consumers’ preferences toward stock-keeping-units (SKUs). They show the advantages of using a more parsimonious attribute-based approach combined with the latent class analysis to achieve superior model fit with fewer parameters. Since then, the latent class analysis has been employed in most attribute-based choice models (e.g., Ho and Chong 2003, Decker and Scholz 2010).
3 Model Specification and Evaluation Methods

3.1 Within-Class Choice Model

We use the standard multinomial logit model in modelling the choice of an SKU. The total utility that the consumer \( i \) assigns to the alternative \( j \) at purchase occasion \( t \) is given as

\[
TU_{jt} = U_{jt} + \epsilon_{jt},
\]

where \( U_{jt} \) is the deterministic component of utility and \( \epsilon_{jt} \) is the random component of utility.

We employ the attribute-specific approach of Fader and Hardie (1996) in our within-class choice model. It incorporates attribute loyalty variables \( LOY_{ikt} \), lagged purchase indicators \( PPP_{jt} \) and \( PNPP_{jt} \), promotion variable \( PROM_{jt} \), regular price variable \( PRICE_{jt} \) and discount variable \( DISC_{jt} \). See Appendix 1 for the full notational convention. The deterministic component of utility is written as

\[
U_{jt} = \sum_{k \in k} \sum_{t \in t} I_{tk}(\alpha_{0tk} + \alpha_{1tk}LOY_{ikt}) + \beta_{1j}PPP_{jt} + \beta_{2j}PNPP_{jt} + \beta_{3j}PROM_{jt} + \beta_{4j}PRICE_{jt} + \beta_{5j}DISC_{jt},
\]

The multinomial logit model gives the choice probabilities in closed form without including the random components of utility. Thus, choice probabilities can be easily calculated without the use of simulation. The choice probability is given as

\[
p_{jt} = \frac{e^{U_{jt}}}{\sum_{z \in j} e^{U_{zt}}}.\]

For a complete specification of the choice model, see Appendix 2.

3.2 Class-Allocation Model

Here we formulate the latent class model that is used to segment consumers into sets with homogeneous preferences and responses to marketing stimuli. The within-class
choice model presented in previous section gives the probability that the consumer $i$ chooses the alternative $j$ at shopping occasion $t$, conditioned on membership to segment $s$. The conditional choice probability is expressed as

$$p_{jt|s}^i = e^{U_{jt|s}^i} / \sum_{j' \in J} e^{U_{j't|s}^i}.$$  \hfill (4)

The general form of the class-allocation probability $q_{st}^i$ is given as

$$q_{st}^i = e^{\lambda_s + \lambda_{as}A_t^i + \lambda_{ps}P_t^i + \lambda_{ds}D_t^i} / \sum_{s' \in S} e^{\lambda_{s'} + \lambda_{as}A_t^i + \lambda_{ps}P_t^i + \lambda_{ds}D_t^i},$$  \hfill (5)

where $A_t^i$ describes consumer $i$’s attitudes at shopping occasion $t$, $P_t^i$ represents consumer $i$’s perceptions at shopping occasion $t$ and $D_t^i$ describes consumer $i$’s observed socio-demographic characteristics. However, most academic studies that analyse consumers’ purchase behaviour using scanner panel data or point-of-sale data do not incorporate consumers’ socio-demographic characteristics or observed choice behaviour in the class-allocation model. If these are omitted, the class-allocation probability reduces to

$$q_s = e^{\lambda_s} / \sum_{s' \in S} e^{\lambda_{s'}},$$  \hfill (6)

where parameters $\lambda_s$ are estimated from the data. In effect, this formulation assumes that all consumers have an equal probability of belonging to a certain segment. The class-allocation probability is also assumed time-invariant. The design of the formula ensures that the following conditions are met:

$$q_s \in [0,1] \text{ and } \sum_{s \in S} q_s = 1.$$  \hfill (7)

The unconditional choice probability is given as a sum of the within-class choice probabilities weighted by the class-allocation probabilities for that specific individual. It is expressed as

$$p_{jt}^i = \sum_{s \in S} q_s p_{jt|s}^i.$$  \hfill (8)
3.3 Maximum Likelihood Estimation

The models are estimated using the maximum likelihood procedure. The optimal parameters are obtained by maximizing the likelihood function. Generally it is more convenient to take the logarithm of the likelihood function and maximize the resulting log-likelihood function as it achieves its maximum value with the same parameter values as the likelihood function. The log-likelihood function is given as

$$LL = \ln \mathcal{L} = \sum_{i \in I} \sum_{s \in S} q_s \sum_{j \in J} \sum_{t \in T} PURCH_{jt}^i \ln(p_{jt}^i),$$

(9)

where $PURCH_{jt}^i$ is a binary purchase indicator, which has the value 1 if the consumer $i$ chooses product $j$ on shopping occasion $t$, 0 otherwise.

The parameters for different segments and the class-allocation coefficients can be estimated simultaneously or sequentially. Maximum likelihood estimation using the EM algorithm is the most commonly used estimation procedure. Note that latent class models do have local maximum solutions. This means that instead of converging to a global optimum, the solution may converge to a suboptimal local optimum. Large number of latent segments and dependencies across variables increase the likelihood of a local maximum solution. To avoid local optimum solutions, the analyst should keep the number of latent classes as few as necessary, use tight convergence criterion and run the estimation procedure multiple times with different initial values for parameters.

3.4 Segment Retention Criteria

The number of latent segments is determined by estimating the model for an increasing number of segments until there is no significant improvement in the model fit. Additional number of segments increases the log-likelihood of the model, but gains are smaller as the number of segments grows higher. Introducing a large number of segments results in overfitting the model, i.e., model explains mostly the noise in the data instead of capturing real aspects of choice behaviour.
Information criteria are used to choose a number of segments which minimises the chosen criterion. Generally, a smaller number of parameters are preferred to a larger number of parameters. Thus, information criteria incorporate penalties for the number of parameters. We introduce and compare several segment retention criteria commonly used in the marketing literature: AIC (Akaike 1974), BIC (Schwarz 1978), CAIC (Bozdogan 1987), AIC3 (Bozdogan 1994) and the test set log-likelihood (Andrews and Currim 2003a). See Andrews and Currim (2003b) for a performance review of segment retention criteria in regression-based marketing models.

Akaike information criterion (AIC) is given as

$$AIC = -2LL + 2k, \quad (10)$$

where $LL$ is the log-likelihood of the model specification and $k$ the number of parameters. AIC is the least conservative of the criteria that we present in this study, penalising only one unit of log-likelihood for each additional parameter.

Bayesian information criterion (BIC) penalises extra parameters more heavily than the AIC, thus favouring a more parsimonious model. It is given as

$$BIC = -2LL + k \ln N, \quad (11)$$

where $N$ is the number of observation in the calibration set.

Consistent Akaike information criterion (CAIC) is the one of the most conservative information criteria. It penalises models heavily for the number of parameters. The criterion is given as

$$CAIC = -2LL + k(\ln N + 1). \quad (12)$$

Modified Akaike information criterion (AIC3) has the best performance of giving the optimal number of segments with logit models using multinomial data (Andrews and Currim 2003b). In spite of its superior performance, it is rarely used in marketing studies. It is given as
Model performance can also be evaluated using cross-validation techniques. These techniques are used to analyse how well the results obtained from one data set generalize to a complementary data set. The hold-out method is one of the simplest cross-validation techniques (e.g., Halkidi and Vazirgiannis 2005, Hamel 2009). In the hold-out method, the data set is randomly partitioned into two complementary subsets, the training set and the test set. The model is calibrated on the observations in the training set, which generally comprise two thirds of the data. The remaining one third of observations is not looked at during the calibration of the model. If the introduction of additional latent segments increases the log-likelihood both in the training set and in the test set, then the improvement is most probably due to capturing real aspects of choice behaviour rather than fitting the model to the noise in the data.

In addition, the likelihood ratio index can be used to compare the performance of a model in different data sets or with different number of parameters. The statistic measures how well the model performs, with its parameters estimated, compared with a model in which all the parameters are set to zero. It is defined as

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

where $LL(\hat{\beta})$ is the value of the log-likelihood function at the estimated parameters and $LL(0)$ is the value of the log-likelihood function with all parameters set to zero. The log-likelihood ratio index has the value zero if the estimated model does no better than the null model and value one if the estimated model is perfectly able to predict each decision maker’s choices.
4 Results and Analysis

4.1 Determining the Number of Segments

In this section, we choose the number of segments to be used in comparison with the standard one-segment model. The use of fewer segments and parameters leads to a simple and elegant model with low computational requirements. Furthermore, the results can be communicated more easily. However, models with fewer segments are able to capture less detail and leave some aspects of choice behaviour unresolved. Increasing the number of segments allows the analyst to capture more detailed information on consumers’ choice behaviour. However, using too many segments increases the risk of overfitting, i.e., capturing random sampling fluctuations.

Table 1 presents log-likelihoods of our model with increasing number of segments and parameters. Log-likelihood is reported separately for the training set and the test set. The test set comprises independent set of customers whose purchases are not used in calibrating the model. We examine whether increasing the number of segments increases the model fit also in the independent test set. If the model is overfitting the data in the training set, then the increase in the test set log-likelihood is likely to be small. We use the test set log-likelihood as one of the segment retention criteria.

<table>
<thead>
<tr>
<th>Segments</th>
<th>Number of parameters</th>
<th>Increase in parameters</th>
<th>Training set log-likelihood</th>
<th>Test set log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18</td>
<td>13</td>
<td>-13,790</td>
<td>-7,402</td>
</tr>
<tr>
<td>2</td>
<td>31</td>
<td>16</td>
<td>-13,578</td>
<td>-7,337</td>
</tr>
<tr>
<td>3</td>
<td>47</td>
<td>16</td>
<td>-13,496</td>
<td>-7,310</td>
</tr>
<tr>
<td>4</td>
<td>63</td>
<td>16</td>
<td>-13,440</td>
<td>-7,277</td>
</tr>
<tr>
<td>5</td>
<td>79</td>
<td>16</td>
<td>-13,365</td>
<td>-7,244</td>
</tr>
<tr>
<td>6</td>
<td>95</td>
<td>16</td>
<td>-13,310</td>
<td>-7,251</td>
</tr>
<tr>
<td>7</td>
<td>111</td>
<td>16</td>
<td>-13,291</td>
<td>-7,236</td>
</tr>
<tr>
<td>8</td>
<td>127</td>
<td>16</td>
<td>-13,282</td>
<td>-7,246</td>
</tr>
<tr>
<td>9</td>
<td>143</td>
<td>16</td>
<td>-13,274</td>
<td>-7,229</td>
</tr>
</tbody>
</table>
We calculate the likelihood ratios in the training set and the test set with increasing number of segments. The likelihood ratio is calculated using Equation (14). The log-likelihood of the null model $LL(0)$ in the denominator is -18,402 in the training set and -9,836 in the test set. Otherwise, the likelihood ratios can be calculated using the information on the number of parameters and log-likelihoods in the training set and the test set presented in Table 1. The virtue of the likelihood ratio index is that it allows comparison of model performance across different data sets.

Figure 2 shows the likelihood ratios in the training set and the test set plotted with increasing number of segments. We observe a slight deterioration of performance as we move from the training set to the test set. The likelihood ratio for the one-segment model is 0.251 in the training set and 0.247 in the test set. Both models improve considerably as additional segments are introduced. This suggests that the latent segments are able to capture additional aspects of choice behaviour that are not explained by the one-segment model. The model improves in both data sets until five latent segments are added. The introduction of the sixth segment causes the likelihood ratio in the test set to decrease, which is a sign that the model is overfitting the training data. Based on test-set likelihood ratio, the five-segment model offers the best fit.

![Figure 2: Likelihood ratio in the training set and the test set with increasing number of segments](image)

We also use the segment retention criteria introduced in section 3.4 in determining the optimal number of segments. The criteria values with increasing number of segments are presented in Table 2. The training set comprises 6,853 independent purchase observations (N) for 1,194 individual customers in the calibration period 2008-2009.
Each additional segment adds 16 new parameters. Note, however, that we use the same values for loyalty variables across all categories as the computational cost of calculating loyalty variables separately for each segment would be unmanageable. This does not reduce the generality of the results, but reduces considerably the complexity of estimation.

Table 2: Segment retention criteria for increasing number of segments

<table>
<thead>
<tr>
<th>Segments</th>
<th>CAIC</th>
<th>BIC</th>
<th>AIC3</th>
<th>AIC</th>
<th>Test Set LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27,757</td>
<td>27,739</td>
<td>27,634</td>
<td>27,616</td>
<td>-7,402</td>
</tr>
<tr>
<td>2</td>
<td>27,460</td>
<td>27,429</td>
<td>27,248</td>
<td>27,217</td>
<td>-7,337</td>
</tr>
<tr>
<td>3</td>
<td><strong>27,454</strong></td>
<td><strong>27,407</strong></td>
<td>27,133</td>
<td>27,086</td>
<td>-7,310</td>
</tr>
<tr>
<td>4</td>
<td>27,499</td>
<td>27,436</td>
<td>27,068</td>
<td>27,005</td>
<td>-7,277</td>
</tr>
<tr>
<td>5</td>
<td>27,506</td>
<td>27,427</td>
<td>26,967</td>
<td>26,888</td>
<td>-7,244</td>
</tr>
<tr>
<td>6</td>
<td>27,555</td>
<td>27,460</td>
<td><strong>26,906</strong></td>
<td>26,811</td>
<td>-7,251</td>
</tr>
<tr>
<td>7</td>
<td>27,674</td>
<td>27,563</td>
<td>26,916</td>
<td><strong>26,805</strong></td>
<td>-7,236</td>
</tr>
<tr>
<td>8</td>
<td>27,813</td>
<td>27,686</td>
<td>26,946</td>
<td>26,819</td>
<td>-7,246</td>
</tr>
<tr>
<td>9</td>
<td>27,953</td>
<td>27,810</td>
<td>26,976</td>
<td>26,833</td>
<td><strong>-7,229</strong></td>
</tr>
</tbody>
</table>

Figure 3 illustrates the values of information criteria with increasing number of segments.
Different criteria achieve their minimums with different numbers of segments. Generally, GAIC and BIC are very conservative metrics. They often suggest using a very low number of segments which leads to underfitting the data. Therefore, we prefer AIC3 and AIC in determining the number of segments.

The six-segment model and the seven-segment model offer the best fit to the data in the training set according to AIC3 and AIC criteria, respectively. Based on earlier research, AIC criterion has a slight tendency to overfit the data (Andrews and Cummings 2003b), which suggests choosing the six-segment model proposed by the AIC3 criterion. However, the introduction of the additional sixth segment causes the log-likelihood in the test set to reduce. This suggests that the six-segment model is overfitting the data in the training set. Therefore, we choose the five-segment model for further analysis and comparison with the standard one-segment model.

### 4.2 Comparison of Models With and Without Segmentation

In this section, we analyse the differences in parameters between the five-segment model and the one-segment (unsegmented) model. We also compare performance of the models in the calibration period and the forecasting period, and for the training set and the test set.

#### 4.2.1 Parameter Estimates

The parameter estimates of the one-segment model are easily understandable as the model gives single values to represent the consumers’ preference toward a product attribute. Multi-segment models have as many preference parameters for each product attribute as there are segments, which complicates the interpretation of parameters. In this section, we show how the parameter estimates of the multi-segment model can be analysed.

Table 3 shows the parameter estimates for the five-segment model and the one-segment model. As can be seen, parameters of the five-segment model are very hard to interpret in a meaningful way. For example, it seems that the consumers in the segment 4 are
very likely to buy brand 2 and brand 4, are extremely brand loyal and prefer capsules and chewable tables. These effects are balanced as opposite preferences in other segments. Issues involved in interpretation of parameter estimates is probably one of the main reasons why segment-specific parameter estimates are rarely reported in the marketing literature. Besides increasing the cost of estimation and complicating parameter interpretability, high number of segments also results in a higher risk of multicollinearity issues.

Table 3: Parameter estimates for the five-segment and the one-segment models

<table>
<thead>
<tr>
<th></th>
<th>5-Seg. Model</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>1-Seg.</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand 1</td>
<td>0.00</td>
<td>-2.27</td>
<td>2.38</td>
<td>-6.90</td>
<td>2.12</td>
<td>1.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand 2</td>
<td>-0.08</td>
<td>-5.88</td>
<td>3.69</td>
<td>68.72</td>
<td>-1.14</td>
<td>0.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand 3</td>
<td>0.90</td>
<td>-2.23</td>
<td>-0.96</td>
<td>61.03</td>
<td>-1.40</td>
<td>0.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand 4</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>1.65</td>
<td>-1.60</td>
<td>1.37</td>
<td>1.09</td>
<td>4.04</td>
<td>1.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>-0.35</td>
<td>0.76</td>
<td>1.59</td>
<td>2.27</td>
<td>-3.93</td>
<td>1.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capsule</td>
<td>1.44</td>
<td>-0.39</td>
<td>-0.36</td>
<td>12.02</td>
<td>-0.04</td>
<td>1.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chewable</td>
<td>-0.97</td>
<td>1.08</td>
<td>-0.02</td>
<td>10.18</td>
<td>1.26</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drops</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand loyalty</td>
<td>2.94</td>
<td>0.81</td>
<td>4.69</td>
<td>72.58</td>
<td>1.96</td>
<td>1.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size loyalty</td>
<td>0.79</td>
<td>1.87</td>
<td>0.64</td>
<td>1.35</td>
<td>3.70</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form loyalty</td>
<td>0.86</td>
<td>1.63</td>
<td>1.02</td>
<td>0.51</td>
<td>4.78</td>
<td>0.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous prom. purchase</td>
<td>-0.35</td>
<td>1.38</td>
<td>0.44</td>
<td>2.31</td>
<td>0.73</td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous non-prom. purchase</td>
<td>1.00</td>
<td>1.61</td>
<td>0.33</td>
<td>0.37</td>
<td>1.13</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Promotion</td>
<td>1.63</td>
<td>2.08</td>
<td>0.00</td>
<td>-6.53</td>
<td>2.01</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular price</td>
<td>2.12</td>
<td>-10.41</td>
<td>-4.84</td>
<td>-0.42</td>
<td>-8.23</td>
<td>-1.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discount</td>
<td>0.55</td>
<td>3.19</td>
<td>-4.40</td>
<td>-0.60</td>
<td>-0.28</td>
<td>-1.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segment size</td>
<td>0.23</td>
<td>0.08</td>
<td>0.38</td>
<td>0.16</td>
<td>0.15</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A general approach in presenting the preferences of individual segments is to present a table of each preference segment’s favourite attributes (e.g., Decker and Scholz 2010). Table 4 presents favourite attribute levels for each preference segment. This form of presentation is easily understandable and communicable.

Table 4: Segment-specific favourite attribute levels

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand</td>
<td>Brand 3</td>
<td>Brand 4</td>
<td>Brand 2</td>
<td>Brand 2</td>
<td>Brand 1</td>
</tr>
<tr>
<td>Size</td>
<td>Large</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Large</td>
</tr>
<tr>
<td>Formula</td>
<td>Capsule</td>
<td>Chewable</td>
<td>Drops</td>
<td>Capsule</td>
<td>Chewable</td>
</tr>
<tr>
<td>Segment size</td>
<td>0.23</td>
<td>0.08</td>
<td>0.38</td>
<td>0.16</td>
<td>0.15</td>
</tr>
</tbody>
</table>

4.2.2 Performance of Models

We compare the performance of the five-segment model and the one-segment model in predicting the consumers’ choice behaviour. Performance comparison is done for different periods of time and different sets of customers. We compare the log-likelihoods of the two models in the calibration period and the forecasting period, and in the training set and the test set.

Table 5 presents log-likelihoods for the two models in different periods and for different sets of customers. Additionally, we are able to compare the performance of the two models in predicting the purchases of new products that are introduced after the calibration period. These forecasts are based on the consumers’ preferences toward the attribute levels of the new products, given that these attribute levels are already present in the market.
Table 5: Log-likelihood for different data periods and sets of customers

**Existing products**

<table>
<thead>
<tr>
<th></th>
<th>Five-segment model</th>
<th>One-segment model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training set</td>
<td>Test set</td>
</tr>
<tr>
<td>Calibration period</td>
<td>-13,365</td>
<td>-7,244</td>
</tr>
<tr>
<td>Forecasting period</td>
<td>-6,860</td>
<td>-3,790</td>
</tr>
</tbody>
</table>

**New products**

<table>
<thead>
<tr>
<th></th>
<th>Five-segment model</th>
<th>One-segment model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training set</td>
<td>Test set</td>
</tr>
<tr>
<td>Forecasting period</td>
<td>-1,557</td>
<td>-780</td>
</tr>
</tbody>
</table>

The results show that the increase in log-likelihood in the calibration period 2008-2009 is significant both in the training set and the test set as we move from the one-segment model to the five-segment model. However, the five-segment model performs worse in the forecasting period 2010-2011 across all samples (existing and new products, training set and test set). The result is surprising as we would have expected the five-segment model to perform better in all samples. The result suggests that consumers’ tastes have significantly changed over time. This is explained by the increasing popularity of probiotics in recent years, which has induced the manufacturers to develop new and enhance existing attribute levels to better meet the changing tastes of consumers.
5 Conclusions

This aim of this study is to show how the latent class analysis can be used in consumer segmentation. It offers a detailed presentation on the necessary steps of incorporating the heterogeneity across consumers using the latent class approach. The results suggest that a choice model can be significantly improved by applying the latent class analysis to identify latent preference segments.

In this study, we implemented an attribute-specific choice model in the multinomial logit framework. We used consumer-level point-of-sale data from retailing of probiotics to healthcare personnel in 2007-2011. We included two thirds of consumers in the training set and calibrated the model with their purchases of probiotics in the calibration period 2008-2009. The model was then calibrated with different specifications using increasing number of latent segments. We used segment retention criteria and test set log-likelihood in choosing the number of segments for further analysis. Based on our analysis, the five-segment model offered the best fit to the data. We then compared the performance of the five-segment model with the performance of the one-segment model. The results show that the five-segment model offers superior fit to the data in the calibration period. However, the five-segment model performs worse in the forecasting period, which is explained by the changes in consumers’ preferences over time.

Our approach can be extended to studying consumer segments in other product categories to find common behavioural patterns. Additionally, more specific background information on consumers’ socio-demographic characteristics could provide us with a basis for more detailed segmentation. This would enable more actionable information on consumers’ characteristics that can be used, for example, to targeted marketing campaigns.
References


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


Appendices

Appendix 1: The Full Notational Convention

All indices are written in lowercase. We write consumer index $i$ in superscript to highlight its importance in segregating different consumers from each other. All other indices are written in subscript. The indices are presented in Table 6.

Table 6: Indices

<table>
<thead>
<tr>
<th>Index</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>shopping occasion</td>
</tr>
<tr>
<td>$i$</td>
<td>consumer</td>
</tr>
<tr>
<td>$j$</td>
<td>product</td>
</tr>
<tr>
<td>$k$</td>
<td>attribute</td>
</tr>
<tr>
<td>$l$</td>
<td>level of attribute</td>
</tr>
</tbody>
</table>

Different types of notation are used for different categories of variables. Capital letters are used for components of utility, independent variables and other variables. The Greek alphabet is used for stochastic variables and parameters that are estimated from the data. Independent variables are presented in Table 8, other variables in Table 9 and parameters in Table 10.

Table 7: Components of utility

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$TU_{it}$</td>
<td>Total utility</td>
</tr>
<tr>
<td>$U_{jt}$</td>
<td>Deterministic component of utility</td>
</tr>
<tr>
<td>$\varepsilon_{jt}$</td>
<td>Stochastic component of utility</td>
</tr>
<tr>
<td>$V_{jk}$</td>
<td>Attribute-specific component of deterministic utility</td>
</tr>
<tr>
<td>$W_{jk}$</td>
<td>SKU-specific component of deterministic utility</td>
</tr>
</tbody>
</table>
**Table 8: Independent variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$LOY_{ikt}$</td>
<td>Attribute loyalty – exponentially weighted average of past purchases treated as binary variables indicating whether or not the household purchased this brand, size, form, etc.</td>
</tr>
<tr>
<td>$PPP_{jt}$</td>
<td>Previous promotional purchase – 1 if the consumer’s previous purchase was a promotional purchase of attribute level $l$, 0 otherwise</td>
</tr>
<tr>
<td>$PNPP_{jt}$</td>
<td>Previous non-promotional purchase – 1 if the consumer’s previous purchase was a non-promotional purchase of attribute level $l$, 0 otherwise</td>
</tr>
<tr>
<td>$PROM_{jt}$</td>
<td>Promotion – 1 if the product $j$ is promoted at shopping occasion $t$, 0 otherwise</td>
</tr>
<tr>
<td>$PRICE_{jt}$</td>
<td>Regular price – undiscounted gross price of the product $j$ at shopping occasion $t$ denoted as gross price / wholesale price $1 + \text{mark-up}%$</td>
</tr>
<tr>
<td>$DISC_{jt}$</td>
<td>Discount – Promotional price decrease of the product $j$ at shopping occasion $t$ as a percentage ($\geq 0$) of regular price</td>
</tr>
</tbody>
</table>

**Table 9: Other variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_{tkj}$</td>
<td>Attribute indicator describing which attribute levels are characterized by the product $j$</td>
</tr>
<tr>
<td>$S_{ikt}$</td>
<td>Share of purchases -variable describing certain attribute level’s share of purchases on a shopping occasion</td>
</tr>
<tr>
<td>$PREF_{ikt}$</td>
<td>Attribute preference toward an attribute level</td>
</tr>
<tr>
<td>$A_{jt}$</td>
<td>Available indicator – 1 if the product $j$ is available at shopping occasion $t$, 0 otherwise</td>
</tr>
<tr>
<td>$PURCH_{jt}$</td>
<td>Purchase indicator – 1 if the product $j$ is purchased at shopping occasion $t$, 0 otherwise</td>
</tr>
</tbody>
</table>

**Table 10: Parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{0tk}$</td>
<td>Attribute level-specific intercept term (FH model only)</td>
</tr>
<tr>
<td>$\alpha_{1k}$</td>
<td>Attribute-specific coefficient of $LOY_{ikt}$ variable</td>
</tr>
<tr>
<td>$\gamma_k$</td>
<td>Carry-over constant of attribute $k$</td>
</tr>
<tr>
<td>$\beta_{0j}$</td>
<td>SKU-specific intercept term (GL model only)</td>
</tr>
<tr>
<td>$\beta_{1j}$</td>
<td>SKU-specific coefficient of $PPP_{jt}$ variable</td>
</tr>
<tr>
<td>$\beta_{2j}$</td>
<td>SKU-specific coefficient of $PNPP_{jt}$ variable</td>
</tr>
<tr>
<td>$\beta_{3j}$</td>
<td>SKU-specific coefficient of $PROM_{jt}$ variable</td>
</tr>
<tr>
<td>$\beta_{4j}$</td>
<td>SKU-specific coefficient of $PRICE_{jt}$ variable</td>
</tr>
<tr>
<td>$\beta_{5j}$</td>
<td>SKU-specific coefficient of $DISC_{jt}$ variable</td>
</tr>
</tbody>
</table>
Appendix 2: Choice Model Specification

Total Utility

Consider a consumer $i$ confronted with a choice from a set of feasible alternatives $j \in J$ at shopping occasion $t$. Here the alternatives are products in a certain category. We assume that the consumer has made a decision to purchase an item from the category. The consumer chooses the product from among all feasible alternatives which maximizes his total utility. The total utility $TU_{jt}^i$ is divided into two parts and can be expressed as

$$TU_{jt}^i = U_{jt}^i + \varepsilon_{jt}^i,$$  \hspace{1cm} (15)

where

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

$\varepsilon_{jt}^i = \text{stochastic component of utility.}$

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

$$P(\varepsilon_j \leq \varepsilon) = e^{-e^{-\varepsilon}}.$$  \hspace{1cm} (16)

The distribution has a mean which corresponds to Euler’s constant (0.57722). A more general form of Equation (16) would include a location parameter to set the distribution mean to zero and a scale parameter that implies the degree of heterogeneity among the consumers. However, we can omit these parameters without any loss of generality. The location parameter is common to all alternatives and therefore does not have an effect to choice probabilities. Scaling of parameters to obtain optimal model fit is built into the estimation step making a scaling parameter unnecessary.
Deterministic Component of Utility

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

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

where

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

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

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

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

\[ U_{jt}^i = V_{jt}^i + W_{jt}^i, \]  

(18)

where

\[ V_{jt}^i \] is the attribute-specific component of utility, and
\[ W_{jt}^i \] is the SKU-specific component of utility.

**Attribute-Specific Component of Utility**

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

\[ V_{jt}^i = \sum_{k \in K} \sum_{l \in L} I_{tkj} PREFERENCES_{ltk}^i, \]  

(19)

where

\[ I_{tkj} = \text{binary variable, which has the value 1 if the product } j \text{ has the } l^{\prime}\text{th level of attribute } k, 0 \text{ otherwise, and} \]
\[ PREFERENCES_{ltk}^i = \text{preference of } l^{\prime}\text{th level of attribute } k \text{ at shopping occasion } t \text{ for consumer } i. \]

\( PREFERENCES_{ltk}^i \) consists of attribute-specific intercept terms, loyalty variables, previous promotional purchase variables and previous non-promotional purchase variables. The expression can be written as

\[ PREFERENCES_{ltk}^i = [\alpha_{0lk} + \alpha_{1k} LOY_{ltk}^i]. \]  

(20)

where

\[ \alpha_{0lk} = \text{attribute level-specific intercept term (parameter, FH model only),} \]
\[ \alpha_{1k} = \text{attribute level-specific coefficient of } LOY_{ltk}^i \text{ variable (parameter),} \]
\( L_{\text{tkt}} \) = attribute level-specific loyalty variable, which is defined as exponentially weighted average of past purchases treated as binary variables indicating whether or not the household purchased a product with attribute level \( l \) of attribute \( k \),

The parameters \( \alpha_{0lk} \) and \( \alpha_{lk} \) are estimated from the data.

**Attribute Level-Specific Intercept Term**

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

**Attribute Loyalty**

\( L_{\text{tkt}} \) variable describes the consumers’ loyalty toward product attributes, e.g., brand, size, form, formula and price. The loyalty is defined in a similar fashion to Guadagni and Little (1983) and Fader and Hardie (1996). In addition to their definition of loyalty, we take into account that multiple products in a category can be bought simultaneously on each shopping occasion. The loyalty variable is defined as the exponentially weighted average of past purchases of the product attribute, treated as binary variables indicating whether or not the household purchased a product with the attribute on earlier shopping occasions. To emphasize the shopping occasion index \( t \), we write \( L_{\text{tkt}}(t) = L_{\text{tkt}} \) and \( S_{\text{tkt}}(t) = S_{\text{tkt}} \). The loyalty variable is expressed with

\[
L_{\text{tkt}}(t + 1) = \gamma_k L_{\text{tkt}}(t) + (1 - \gamma_k) S_{\text{tkt}}(t)
\]

where

\( \gamma_k = \) carry-over constant for product attribute \( k \), \( \gamma_k \in [0,1] \), and

\( S_{\text{tkt}} = \) share of purchases on shopping occasion \( t \) that have attribute level \( l \) of product attribute \( k \).
The carry-over constant $\gamma_k$ determines the share of $LOY_{ikt}^i$ variable carried over to the next shopping occasion $t + 1$. It is a parameter which is to be estimated for each product attribute $k$ separately so that it offers the model the best fit to the data. The sum of loyalties across product attributes must sum to 1.

We allow for multiple products being bought at one shopping occasion. Therefore we must introduce variable $S_{ikt}^i$ that describes the share of purchases that have attribute level $l$ of attribute $k$ at shopping occasion $t$. $S_{ikt}^i$ is simply defined as the number of products bought that have attribute level $l$ of attribute $k$ divided by the number of total category purchases on the shopping occasion $t$. We define

$$S_{ikt}^i = \frac{\sum_{j \in J} I_{lk_j}^i PURCH_{jk}^i}{\sum_{j \in J} PURCH_{jk}^i}$$

(22)

where

$I_{lk_j}^i = $ binary variable (see Equation (19)), which has the value 1 if the product $j$ has the $l$’th level of attribute $k$, 0 otherwise, and

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

The purchase indicator $PURCH_{jk}^i$ is defined

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

(23)

To start up the loyalty variable, the initial values for $LOY_{ikt}^i$ variable must be defined. Understandably, the loyalty variable cannot be defined without the purchase history. Therefore the loyalty variable is not defined at the first shopping occasion $t = 1$. We initialize the loyalty variable at the second shopping occasion $t = 2$ as follows

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

(24)

where
\( N_k \) = total number of attribute levels \( l \) associated with product attribute \( k \), and

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

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

**SKU-Specific Component of Utility**

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

\[
W_{jt}^i = \beta_{0j} + \beta_{1j}PPP_{jt}^i + \beta_{2j}PNPP_{jt}^i + \beta_{3j}PROM_{jt}^i + \beta_{4j}PRICE_{jt}^i + \beta_{5j}DISC_{jt},
\]

where

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

\( \beta_{1j} \) = SKU-specific coefficient of \( PPP_{jt}^i \) variable (parameter),

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

\( \beta_{2j} \) = SKU-specific coefficient of \( PNPP_{jt}^i \) variable (parameter),

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

\( \beta_{3j} \) = SKU-specific coefficient of \( PROM_{jt}^i \) (parameter),

\( PROM_{jt}^i \) = SKU-specific promotion variable, which has the value 1 if the product \( j \) of category \( c \) is promoted on the shopping occasion \( t \), 0 otherwise,

\( \beta_{4j} \) = SKU-specific coefficient of \( PRICE_{jt}^i \) (parameter),

\( PRICE_{jt}^i \) = SKU-specific price variable.
\[ \text{PRICE}_j^t = \text{SKU-specific regular price variable, (undiscounted) gross price of the product } j \text{ at shopping occasion } t \text{ denoted as gross price / wholesale price } = 1 + \text{mark-up } \% , \]

\[ \beta_{sj} = \text{SKU-specific coefficient of } \text{DISC}_{jt}^i \text{ (parameter), and} \]

\[ \text{DISC}_{jt}^i = \text{a non-negative SKU-specific discount variable, which has the discount of product } j \text{ at the shopping occasion } t \text{ as a percentage of normal price, 0 otherwise.} \]

All parameters \( \beta \) are estimated from the data.

**Previous Promotional and Previous Non-Promotional Purchase**

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

The previous promotional purchase variable \( \text{PPP}_{jt}^i \) describes whether or not the product \( j \) was on promotion and was bought by consumer \( i \) at the previous shopping occasion. We define

\[ \text{PPP}_{jt}^i(t + 1) = \begin{cases} 1 & \text{if one of customer } i\text{’s purchases at the shopping occasion } t \text{ is a promotional purchase of product } j, \\ 0 & \text{otherwise.} \end{cases} \]  

(26)

We use a similar approach to modelling the previous non-promotional purchases. We define previous non-promotional variable \( \text{PNPP}_{jt}^i \) with

\[ \text{PNPP}_{jt}^i(t + 1) = \begin{cases} 1 & \text{if one of consumer } i\text{’s purchases at the shopping occasion } t \text{ is a non-promotional purchase of product } j, \\ 0 & \text{otherwise.} \end{cases} \]  

(27)

**Promotion**

The effect of promotion is determined with \( \text{PROM}_{jt} \) variable. We define
\[ PROM/jt = \begin{cases} 
1 & \text{if product } j \text{ was on promotion at the consumer } i's \text{ tth shopping occasion}, \\
0 & \text{otherwise.} 
\end{cases} \quad (28) \]

**Regular Price**

Variable \( PRICE/ijt \) captures the effect of regular price to consumer choice. We define

\[
PRICE/ijt = \frac{\text{Gross Price}_j}{\text{Wholesale Price}_j} = 1 + \text{Markup\%-}i_jt \quad (29)
\]

**Discount**

The effect of discount is captured in \( DISC/ijt \) variable. We define

\[
DISC/ijt = \text{Discount (as \% of Regular Price)} \quad (30)
\]

**The FH Model**

The final presentation for the deterministic component of utility in the FH model becomes

\[
U/jt = V/jt + W/jt = \sum_{k \in K} \sum_{l \in L} I/ltj (\alpha_{0lk} + \alpha_{1lk} L OY/ltkj) + \beta_{1j} PP/ijt + \beta_{2j} PNPP/ijt + \beta_{3j} PROM/ijt + \beta_{4j} PRICE/ijt + \beta_{5j} DISC/ijt \quad (31)
\]

The model uses the attribute level-specific intercept terms \( \alpha_{0lk} \) to define the consumer’s preference toward attribute levels. The FH model represents a parsimonious approach to choice modelling at the SKU-level with insightful parameter estimates describing consumers’ preferences toward product attributes.

**Purchase Probability**

The consumer chooses the product with the highest utility among the set of choice alternatives. The probability that the consumer \( i \) chooses alternative \( j \) from the set of feasible alternatives \( J \) at shopping occasion \( t \) is
\[ p_{jt}^{i} = \frac{e^{u_{jt}^{i}}}{\sum_{z \in j} e^{u_{zt}^{i}}}. \] (32)

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