The authors develop a class of mixtures of piece-wise exponential hazard models for the analysis of brand switching behavior. The models enable the effects of marketing variables to change nonproportionally over time and can, simultaneously, be used to identify segments among which switching and repeat buying behavior differ. Several forms of asymmetry in brand switching are accommodated. The authors provide an application to the analysis of scanner panel data on ketchup, which illustrates the implications for asymmetry, nonproportionality, and heterogeneity. The results show that the model predicts purchases and purchase timing in holdout data better than the models proposed by Kamakura and Russell (1989) and Vlachos and Jain (1991).

Implications for Asymmetry, Nonproportionality, and Heterogeneity in Brand Switching from Piece-wise Exponential Mixture Hazard Models

Assume the following hypothetical situation: A consumer visits a retail chain once a week. Last week, the consumer noticed that his or her regular brand of ketchup, Heinz, was reduced in price, but decided not to buy it. Prior to leaving home for this week's shopping trip, the consumer notices that the 32-ounce bottle of Heinz ketchup he or she purchased four weeks ago is almost empty. This week, the consumer purchases two large 32-ounce bottles of Heinz ketchup, which are still under promotion.

We use this simple hypothetical example to demonstrate a number of issues that are very pertinent to the modeling framework we present. We develop hazard models to delve into the question of when and why consumers switch brands. Hazard models have become widespread in their use in such fields as biology, medicine, economics, sociology, and psychology, but their application to marketing problems is of much more recent date (cf. Gönül and Srinivasan 1993; Vlachos and Jain 1991).

Our hazard models of brand switching deal with three issues. First, the previous example shows that because of the discrete nature of the timing of shopping trips, purchase decisions typically occur in discrete time intervals. Therefore, the framework we formulate to model brand switching behavior deals with such discrete time intervals through a so-called piece-wise exponential formulation of the hazard (e.g., Laird and Olivier 1981). As we discuss subsequently, this formulation enables us to include purchase quantities and multiple purchases into our model in a straightforward manner.

Second, from the example, it appears that consumers' sensitivity for marketing activities may change as time since the last-purchase elapses. Therefore, we specify nonproportional hazard functions, which do not assume the effects of the marketing instruments to be constant over time. The example also demonstrates that shopping trips in which a consumer does not purchase, though he or she is aware of the changes in, for example, price, convey information on marketing effectiveness. Therefore, we allow marketing instruments to vary between purchases (cf. Gupta 1991).

Third, segments of consumers may differ in their sensitivity to marketing instruments, as has been demonstrated in various applications of mixture regression models to marketing problems (cf. Wedel and DeSarbo 1994). In our mixture formulation, duration dependence varies across segments as well, which allows segments of consumers to switch and repeat-buy at different rates.
We detail the hazard model framework in the subsequent section. The main contribution of our approach is that we allow for nonproportionality and heterogeneity across segments in the investigation of asymmetries of brand switching, while using a piece-wise exponential hazard framework. We briefly outline the algorithm used for estimation. We describe an application to scanner panel data on ketchup, which illustrates the substantive insights into brand switching behavior that may be obtained with our method. We show that our approach predicts purchases in holdout data better than the (benchmarking) models of Vilcassim and Jain (1991) and Kamakura and Russell (1989), and outperforms the model of Vilcassim and Jain (1991) with respect to the prediction of interpurchase times.

THE HAZARD MODEL FRAMEWORK

Let:

\[ i = 1, \ldots, I \] households,
\[ h = 1, \ldots, H_i \] time intervals between purchases, or spells, for household \( i \),
\[ j = 1, \ldots, K \] brands bought on the last-purchase occasion,
\[ k = 1, \ldots, K \] brands bought on the present purchase occasion,
\[ r = 0, \ldots, R_{hi} \] weeks since last purchase,
\[ p = 1, \ldots, P \] covariates,
\[ s = 1, \ldots, S \) segments,
\[ t_{hi} \] time elapsed in weeks since the last purchase within spell \( h \), for household \( i \),
\[ X_{ikpr} \] the value of the \( p \)th marketing variable for brand \( k \) in week \( r \) of spell \( h \) of household \( i \), and
\[ Y_{ikhr} \] the number of equivalent units of brand \( k \) bought by household \( i \) in week \( r \) of spell \( h \).

Piece-wise Exponential Formulation

We observe \( I \) households choosing among \( K \) brands in a product category. For household \( i \), we observe \( H_i \) time intervals between purchases. The time within spell \( h \) is partitioned into \( R_{hi} \) discrete intervals of equal length. Each spell can be split into one-hour, one-day, one-month, and so on, intervals. We use one-week intervals because of the underlying behavioral assumptions of consumers' weekly planning of shopping trips.

We assume that the sample of consumers arises from a population that is a mixture of \( S \) unobserved segments, in proportions \( \pi_1, \ldots, \pi_S \). Let \( T \) denote a random variable indicating the time before a switch or repeat occurs. Conditional on segment \( s \), we assume the hazard of switching from buying one unit of brand \( j \) to buying one unit of \( k \), in week \( r \) of spell \( h \), to be constant. The hazard is defined on the basis of the probability of a switch in a small time unit \( \Delta t \), where \( \Delta t \) approaches zero:

\[
\lim_{\Delta t \to 0} \frac{P(\text{switch} | t_{hi} \leq T < t_{hi} + \Delta t; T \geq t_{hi})}{\Delta t} = \lambda_{jk|h},
\]

which implies that the hazard is constant, the interpurchase time is exponentially distributed within each week, and the timing of purchases within a week is random. This assumption is based on the timing of purchases in a particular category being exogenously determined by the timing of the shopping trip. Our formulation has the advantage that exponential models of duration times are equivalent to Poisson models of the quantity of units purchased, and we, thus, allow for several units being bought within one week (e.g., Laird and Oliver 1981). We derive the survivor function for brand \( k \), the probability of no purchase of \( k \) being observed before \( t_{hi} = r \), if \( j \) was purchased last, as

\[
S_{jkhr}(t_{hi}) = \prod_{u = 1}^{r} e^{-\lambda_{jk|h}}.
\]

It may be observed that the overall survival rate, defined as the probability that no purchase of any brand is made before \( t_{hi} \),

\[
S_{jhr}(t_{hi}) = \prod_{k=1}^{K} S_{jkhr}(t_{hi}),
\]

involves the hazard rates of all competing brands. Discrete-time models have previously been used, for example, by Meyer (1990). However, the approach Meyer uses is not appropriate for the analysis of brand switching, because it does not deal with repeated spells and time-varying covariates. Contrary to Meyer, we use a mixture formulation of heterogeneity. The models we propose can be called mixture semi-Markov models of brand switching.

Specification of the Hazard

Conditional on the covariates \( X_{ikpr} = (X_{ikpr}, X_{ikph}) \) and segment \( s \), we formulate the log-hazard rate of a \( j-k \) switch as,

\[
\ln(\lambda_{jk|h}) = \Theta_{jk}(t_{hi}) + \Phi_{jk}(X_{ikhr}, X_{ikph}).
\]

Equation 1 explains the hazard rate from three factors. The first factor \( \Theta_{jk}(t_{hi}) \) is the baseline hazard function. We employ a step-function formulation of the baseline hazard:

\[
\Theta_{jk}(t_{hi}) = \alpha_{jks} + \sum_{l=1}^{L} \alpha_{jak} G_{l}(t_{hi}),
\]

where \( G_{l}(t_{hi}) = 1, \ldots, L \) represent some functions of \( t_{hi} \) that remain to be specified. We may, for example, use \( G_{1}(t_{hi}) = t_{hi} \), \( G_{2}(t_{hi}) = t_{hi}^{2} \), and \( G_{3}(t_{hi}) = \ln(t_{hi}) \). The second term in Equation 1 captures the direct effects of the covariates \( X_{ikhr} \).

\[
\Phi_{jk}(X_{ikhr}) = \sum_{p=1}^{P} \beta_{jkp} X_{ikph}.
\]

The third term in Equation 1 specifies nonproportional effects of the covariates in a conceptually simple way through interactive effects of the covariates and duration:

\[
\Psi_{jk}(t_{hi}, X_{ikhr}) = \sum_{m=1}^{M} \sum_{p=1}^{P} \gamma_{jkm} G_{m}(t_{hi}) X_{ikph},
\]

where \( G_{m}(t_{hi}) \) denotes \( m = 1, \ldots, M \) functions of time—as in Equation 2. This term enables us to investigate the dynamics of the effects of the marketing instruments. For reasons of parsimony, we make \( M = 1 \) in Equation 4.

Tests for Asymmetry

Our approach enables a detailed investigation of the asymmetries in the rates of brand switching and repeat buying. To investigate the extent of asymmetry in competition, three basic types of restrictions are tested:
(A) If no restrictions are imposed, the effects of duration and covariates are estimated for all \( R^2 \) transitions from brand \( j \) to brand \( k \), within each of the \( S \) segments. Here, \( j = 1, \ldots, K; k = 1, \ldots, K \) indicates the full transition matrix and \( D^{(A)} = SK^2(P(I + M) + L + 1) + S - 1 \) parameters are estimated. Complete asymmetry in switching is allowed in this specification.

(B) If it can be assumed that for switches across all brands \( j \neq k \) to a specific brand \( k \), the parameters are equal, the number of distinct transitions for which parameters need to be estimated reduces to \( 2K \), and the number of model parameters to \( D^{(B)} = 2SK(P(1 + M) + L + 1) + S - 1 \). Here, \( j = 0, 1; k = 1, \ldots, K \), where \( j = 0 \) indicates a repeat-purchase and \( j = 1 \) indicates a switch. In this case, all switching rates to a certain destination brand are assumed to be equal.

(C) A further restriction arises when all switching and all repeat purchasing parameters are assumed to be identical. Here, there are only two possible transitions (switches or repeat-purchases), the subscript \( k \) drops out, \( j = 0, 1 \), and the number of estimated parameters reduces to \( D^{(C)} = 2S(P(I + M) + L + 1) + S - 1 \). In this model, switching behavior is symmetric.

These restrictions on the switching matrix can be imposed for all parameters or either of the sets of parameters in equations 2 to 4 and allow for partial asymmetries in the baseline hazard, marketing mix effects, or the dynamics of those effects. Note also that the three optional specifications are nested within one another.

**ESTIMATION**

The parameters of the model are estimated by maximizing the log-likelihood:

\[
\ln(L) = \sum_{i=1}^{I} \sum_{s=1}^{S} \ln \left( \pi_i \right) + \sum_{h=1}^{H} \left[ \prod_{j=1}^{J} \prod_{k=1}^{K} \left( \lambda_{j,k|i,h} \delta_{i|h} \delta_{|k} \right)^{y_{i|h}|_{j,k} \delta_{i|h} \delta_{|k}} \right]
\]

where \( \delta_{i|h} = 1 \) if spell \( h \) of household \( i \) ends in a purchase of brand \( k \), and \( \delta_{i|h} = 0 \) otherwise. We use the result that the parameters of piecewise-rectangular models can be estimated by solving systems of equations derived from the Poisson likelihood. This requires creating one record of data for each week for each respondent, and using the \( Y_{i,k|h} \) as a dependent variable. To obtain the estimates we use an Expectation-Maximization algorithm to maximize the log-likelihood function (cf. Wedel and DeSarbo 1994). Using this algorithm, we estimate the parameters of the model and simultaneously obtain a classification of consumers into segments. The Expectation-Maximization algorithm is a particularly attractive algorithm. It decreases the likelihood at each iteration cycle and converges to, at least, a local maximum. To minimize problems of local maxima, several (random) starts must be used.

When the estimates are obtained, consumers can be assigned to segments on the basis of posterior probabilities; the separation of the segments can be examined using an entropy measure \( E_{i} \) (cf. Wedel and DeSarbo 1994). \( E_{i} \) is bounded between 0 and 1, a value close to 1 indicates that the segments are well-separated. In applications in which the number of segments \( S \) is unknown, we select the number of segments that yields the minimum value of the Consistent Akaike Information Criterion (CAIC). The CAIC is defined as \(-2 \max(\ln(L)) + D^{(R)}/\ln(N) + 1\), where \( D \) is the number of parameters and \( N \) the total number of observations.

The parameter estimates are asymptotically normal given \( S \); their variance-covariance matrix is obtained from the inverse of the Fisher information matrix. Possible nonrandomness of within week purchase patterns—caused by, for example, more regular timing of shopping trips within weeks—results in the mean and variance of \( Y_{i,k|h} \) not being equal, as the Poisson assumption implies. To deal with this, we assume the arising overdispersion to have the specific form: \( \text{Var}(Y_{i,k|h} | s) = \sigma^2 \mu_{i,k|h} \). \( \mu_{i,k|h} \) denotes the conditional expectation given \( s \) of \( Y_{i,k|h} \); even substantial violations of this assumed form have only small effects; cf. McCullagh and Nelder (1989, p. 199). Under this assumption, the estimates of the parameters obtained by maximizing the log-likelihood are unbiased, but the covariance matrix of the estimates for segment \( s \) is multiplied by an estimate of \( \sigma^2 \).

**APPLICATION TO KETCHUP SCANNER DATA**

**Data**

We applied the proposed model to the A. C. Nielsen Company scanner panel data on ketchup purchases collected in the Springfield, Mo. market. We distinguish three major national brands, Heinz, Hunts, and Delmonte, and an aggregated other-brands category. In addition, we code brand sizes as integer equivalent units of 14 ounces and use a random sample of 200 subjects. The analysis sample comprises the period from week 25 in 1986 to week 10 in 1987. Our study uses the data from week 10 in 1987 to week 25 in 1987 for validation. Information on price paid, display, and feature are available for each week in which a store visit was made by a household. If no store visit was made in a particular week, we inputted the information from the previous week. Some characteristics of the aggregate sample are shown in the first column of Table 4.

**Model Selection**

Exploratory analyses revealed that (1) the empirical hazard has a varying height across brands and is lower for switches than for repeat-purchases and (2) price, display, and feature, as well as \( t_{i|h} \), \( \ln(t_{i|h}) \), and \( t_{i|h}^2 \), display multicollinearity. Therefore, in Equation 2, we specify the duration distribution using \( G_1(t_{i|h}) = \gamma_{i|h} \). \( G_2(t_{i|h}) = \ln(t_{i|h}) \), \( G_3(t_{i|h}) = \text{LOG}(t_{i|h}) \), and \( G_4(t_{i|h}) = 0 \) for \( i > 1 \). In Equation 4, we use \( G_1(t_{i|h}) = t_{i|h} \) and \( G_2(t_{i|h}) = 0 \) for \( i > 1 \). To minimize the problems of collinearity, we restrict the present analyses to an investigation of direct (PRICE) and nonproportional (NPPPRICE) price-effects (the condition number of the X-matrix, 25.1, did not indicate serious collinearity).

First, we determine the appropriate value of \( S \), using the saturated Model A, which was estimated for \( S = 1 \) to \( S = 4 \). The CAIC statistic reaches a minimum for \( S = 2 \) segments (see Table 1). Second, given \( S \), we investigate whether asymmetry restrictions of models B and C can be imposed. In Model C we restrict all coefficients to be identical for all switches and repeats, but because of the exploratory results
the intercepts were allowed to vary across destination-brands, as in Model B (α_{j, k} \neq 0, 1; k = 1...4). Model C is the most parsimonious model, as evidenced by likelihood-ratio tests and the CAIC statistics (see Table 2). The S = 2 solution for Model C provides minimum CAIC, compared to the S = 1 and S = 3 models.) The entropy statistic, E_s = .835, indicates that the segments are reasonably well-separated. All models were estimated using ten random starts. The likelihoods of the ten S = 2 solutions for Model C were close, differing only by 2.0%.

Results

The parameter estimates of the S = 2 Model C are presented in Table 3. Figure 1 presents the estimated baseline hazard functions for repeat buying and switching (calculated at the average price-levels) for both derived segments. Table 4 presents descriptive statistics for the two segments.

Segment 1 constitutes 45.6% of the sample. In this segment, the baseline hazards of repeat buying for Heinz and Delmonte ketchup increase substantially (see Figure 1), as does the baseline switching hazard for Heinz. The results imply strong inventory depletion effects: The corresponding purchase probabilities increase sharply among consumers who have not yet made a purchase and whose inventories are depleting. Both repeat-purchases and switches are significantly affected by price, but the effect on switches is approximately 50% higher. Apparently, price-discounts are less effective in attaining last-brand loyal behavior than in inducing switching behavior. The repeat-buying hazard shows a positive NPPRICE effect. This indicates that the relative sensitivity of consumers to repeat-purchase in response to a price discount decreases. (The same effect for switches is not significant.) In other words, consumers in this segment become less likely to repeat-purchase in response to a price-promotion if they have not repeat-purchased. A possible explanation is that as time elapses and stocks deplete, the need for ketchup rises and the drive of these loyal buyers to buy their favorite brand increases whether or not the brand is promoted. Compared to the other brands, Heinz is more effective in both keeping its own buyers and drawing buyers from its competitors. Table 4 shows that in Segment 1, Heinz has a market share of 86% and almost 80% of purchases are repeat-purchases. The average interpurchase time is approximately five weeks. This segment appears to be a segment of predominantly last-purchase loyal Heinz buyers.

In Segment 2 (54.4% of the sample), the baseline hazards of repeat buying are inverse-u-shaped and much lower than those in Segment 1 (see Figure 1). The peaks occur at approximately three to four weeks. The baseline hazards of switching are much higher than those of repeat buying, and they increase monotonically, showing an inventory depletion effect. The hazards of switching and repeat buying increase about equally as a brand offers a price discount (PRICE). The repeat-purchase hazard displays a significant negative NPPRICE effect, which indicates that the impact of a price discount on repeat-purchases increases as the time since the last purchase elapses. One explanation is that these consumers wait until their favorite brand is being promoted. The positive NPPRICE effect on switching indicates that the effects of promotions on switches decrease over time. In other words, the longer these consumers have resisted the enticements to switch, the harder it becomes to persuade them to do so with a promotional discount. Again, Heinz is most effective in retaining its buyers and drawing buyers from other brands. The average interpurchase time in this segment is seven weeks. The market share of Heinz in Segment 2 is half of that in Segment 1, whereas the shares of Delmonte, Hunts, and the other brands are three to four times as high (see Table 4). Almost 60% of all purchases in Segment 2 consist of switches. Segment 2 accounts for nearly 20% less purchase volume than Segment 1 (see Table 4). Thus, Segment 2 can be designated as a segment of brand-switchers.

Comparisons

We compare the predictive validity of our approach with that of Vilcassim and Jain’s (1991) hazard model and Kamakura and Russell’s (1989) latent class model, subsequently denoted as the V&J and K&R models, respectively. These models are considered as benchmarks for hazard models for brand-switching (V&J) and models for heterogeneity in brand choice (K&R). Note that our model has several conceptual advantages over (1) the V&J model of allowing for non-proportional price effects, time-varying

Table 3

<table>
<thead>
<tr>
<th>Segment 1</th>
<th>Segment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Repeat</strong></td>
<td><strong>Switch</strong></td>
</tr>
<tr>
<td>WEEK</td>
<td>-0.008</td>
</tr>
<tr>
<td>LOG(WEEK)</td>
<td>1.303*</td>
</tr>
<tr>
<td>PRICE</td>
<td>-1.756*</td>
</tr>
<tr>
<td>NPPRICE</td>
<td>0.044*</td>
</tr>
<tr>
<td>HEINZ</td>
<td>-1.043*</td>
</tr>
<tr>
<td>DELMONTE</td>
<td>-1.452*</td>
</tr>
<tr>
<td>HUNTS</td>
<td>-4.582*</td>
</tr>
<tr>
<td>OTHERS</td>
<td>-6.630*</td>
</tr>
</tbody>
</table>

*Denotes p < .01.
covariates, purchase quantities, and segments and (2) the K&R model of allowing for brand switching, purchase timing, time-varying covariates, and purchase quantities.

To enable an empirical comparison of our model with the V&J and K&R models, we disregard quantities purchased from the validation data. The brand switching specification used in the V&J model is identical to that of our model, and we specify two segments in the K&R model.

On the basis of the estimates of the three models, we predicted both choices and durations in the holdout data set. Thus, we calculated the transition-specific survivor functions, $S_{jkh}(t)$. Purchases, conditional on duration, are predicted using a Monte Carlo simulation (500 replications) based on the values of the survivor function. Duration, given a purchase, is predicted as that point in time when the survival function drops below $0.5$. Because the K&R model is not a hazard model, only holdout choices were predicted with this model.

The percentage of choice hits, averaged over the 500 Monte Carlo runs, was 37.99% for V&J, 40.79% for K&R, and 49.70% for our model. The average correlation of actual and predicted duration was .10 for V&J and .23 for our model, whereas the average Root Mean Squared prediction errors were 6.55 and 5.39, respectively, which amounts to 19.3% and 15.9% of the range of duration in the holdout sample. Choice prediction of our model appears to be more than 20% better than that of the V&J and K&R models. Our model also predicts the timing of purchases better than the V&J model by over 20%.

Table 4
AGGREGATE LEVEL AND SEGMENT CHARACTERISTICS

<table>
<thead>
<tr>
<th>Shares (%)</th>
<th>Aggregate</th>
<th>Segment 1</th>
<th>Segment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEINZ</td>
<td>65.38</td>
<td>85.67</td>
<td>42.13</td>
</tr>
<tr>
<td>DELMONTE</td>
<td>15.65</td>
<td>8.14</td>
<td>25.54</td>
</tr>
<tr>
<td>HUNTS</td>
<td>10.56</td>
<td>4.65</td>
<td>17.44</td>
</tr>
<tr>
<td>OTHERS</td>
<td>8.41</td>
<td>1.55</td>
<td>14.89</td>
</tr>
<tr>
<td>Number of purchases</td>
<td>962</td>
<td>421</td>
<td>541</td>
</tr>
<tr>
<td>Repeat-purchases (%)</td>
<td>39.29</td>
<td>76.67</td>
<td>52.51</td>
</tr>
<tr>
<td>Interpurchase time (week)</td>
<td>5.70</td>
<td>4.77</td>
<td>7.18</td>
</tr>
<tr>
<td>Volume (%)</td>
<td>100.00</td>
<td>58.67</td>
<td>41.33</td>
</tr>
<tr>
<td>Size (%)</td>
<td>100.00</td>
<td>45.70</td>
<td>54.30</td>
</tr>
</tbody>
</table>
CONCLUSIONS

We have presented a piece-wise exponential mixture hazard approach for the analysis of brand switching behavior. The model is capable of predicting the quantity purchased, considering the changes in the store environment between purchases of the product category, allowing for heterogeneity both in the effects of marketing variables and the duration dependence across segments of consumers, and considering nonproportional effects of marketing variables. We have empirically demonstrated that, in our application, these features lead to an improvement of the prediction of both purchase timing and brand choice over previously published models. Various managerial implications of our model, with respect to asymmetry in brand switching, nonproportional effects of price, and heterogeneity, were discussed.

Possible avenues for further research are the inclusion of competitive effects in our model, Monte Carlo studies on the relative performance of continuous and discrete time hazard models, and comparisons between the latent class framework and other approaches to heterogeneity, such as varying parameter models.

REFERENCES


