On the Stability of Store Format Choice

Giuseppe Galata

Randolph E. Bucklin

Dominique M. Hanssens[1]

November 1999

[1]Giuseppe Galata is Marketing Manager for Italy, Enron Europe, London, United Kingdom. Randolph E. Bucklin is Associate Professor and Dominique M. Hanssens is Bud Knapp Professor of Management, Anderson School at UCLA, 110 Westwood Plaza, Los Angeles, CA 90095. We thank Mr. Doug Honold and Ms. Tara Merrill of Information Resources, Inc. for providing the data used in this study and David Bell for assistance with the use of the data. We also thank the editor and three anonymous reviewers for their comments and suggestions.

On the Stability of Store Format Choice

This paper investigates the nature of segmentation in store choice behavior where competing supermarkets offer Every Day Low Price (EDLP) and High-Low Promotional Price (HILO) formats. Previous research, both theoretical and empirical, has held that supermarkets offering EDLP and HILO price formats compete for the business of *the same shoppers* (e.g., Lal and Rao 1997) – though they may do this in different ways for different segments. The authors propose an alternate segmentation structure where *inter-format* competition (i.e., HILO versus EDLP) occurs for the business of some shoppers while *intra-format* competition (i.e., HILO versus HILO and EDLP versus EDLP) takes place for the business of other shoppers. Relatively stable demographic traits (e.g., income and family size) are hypothesized to characterize the shoppers in the various segments, implying that the choice of price format also will be relatively stable for many supermarket shoppers.

The authors provide empirical evidence for their framework from the scanner data records of households shopping among five nearby supermarkets (two EDLP, three HILO) in a large city in the midwest U.S. First, a store switching analysis reveals only modest levels of inter-format switching, but a large extent of intra-format switching. Second, a multinomial logit model of store choice estimated with latent classes reveals three distinct shopper segments: one that is almost exclusively EDLP, one that is almost exclusively HILO, and one that switches among both EDLP and HILO price formats. The estimated size of the two intra-format segments together is substantially larger than the inter-format segment. This suggests that, depending upon market conditions, competition within price formats may be a more significant factor in supermarket retailing than competition between price formats.

Keywords: Store Choice, Shopping Behavior, Retail Price Format, Choice Models, Segmentation

On the Stability of Store Format Choice

Understanding the nature of competition among supermarkets is an important and active area of inquiry for both researchers and practitioners in retailing. In particular, supermarkets offer the opportunity to study retail competition based primarily on price format and service levels, given that most supermarkets stock a comparable, though not identical, assortment of goods. Attesting to the growing interest in the topic, a number of papers have recently appeared on the subject of price format competition among supermarkets (e.g., Bell and Lattin 1998, Hoch, Dreze and Purk 1994, Lal and Rao 1997).

Supermarket price formats – like most elements of retail formats – are subject to a wide range of variation. Among the available formats, the literature and the trade press have emphasized Every Day Low Pricing (EDLP) and Promotional Pricing (PROMO or HILO) as two clear and opposing positioning choices. The stylized view is that EDLP stores offer lower average prices but HILO stores offer frequent price specials on individual items. Industry studies have reported that the implementation of EDLP and HILO pricing formats is more of a continuum than a dichotomy (e.g., Information Resources, Inc. 1993). Few supermarkets run "pure EDLP programs" (Hoch, Dreze and Purk 1994) and, indeed, EDLP stores also engage in frequent price promotions. Nevertheless, two clear distinctions do emerge in practice: (1) mean prices are lower in EDLP stores than HILO stores and (2) price promotions are deeper (implying greater price variance) in HILO stores than EDLP stores (Information Resources, Inc. 1993). We also provide confirming evidence of these two distinctions for the stores in our data set. While there are clearly variations in the implementation of EDLP and HILO price formats, these positioning strategies have become a stable feature of the supermarket retailing landscape and the subject of a considerable body of industry and academic study.

Background on EDLP-HILO Competition

Recent conceptual thinking and theoretical work on retail positioning strategy holds that supermarkets with different price formats compete for business from most, if not all, segments of consumers (Corstjens and Corstjens 1994). This is accomplished via multidimensional strategies which offer specific elements that appeal to different segments. Interestingly, this view of retail strategy (i.e., a store can appeal to shoppers from multiple segments) stands in some contrast to the conventional view of product line strategy in which different products are designed and targeted to appeal to different segments of buyers.

Also using the notion of multidimensional strategy, Lal and Rao (1997) develop a theoretical model of EDLP-HILO format competition. In their model, both EDLP and HILO stores compete for the business of shoppers in each of two segments: cherry-pickers and time-constrained consumers. Their argument proceeds as follows: The EDLP store appeals to time constrained shoppers by offering an attractive price proposition for one-stop shopping; it also appeals to the cherry pickers by offering lower prices on items not contemporaneously promoted in a HILO store. On the other hand, the HILO store appeals to time constrained shoppers by offering greater convenience and higher service levels; it also appeals to cherry pickers with its advertised price specials. Thus, in Lal and Rao's conceptualization the two price formats compete directly with each other, but in different ways in two different segments. An important feature to note in the Lal and Rao framework is the implication that shoppers in each segment switch between the EDLP and HILO format stores in order to fulfill their shopping needs over time.

The idea that both HILO and EDLP stores compete in this fashion has strong intuitive appeal, as it is generally acknowledged that most stores' customer bases include consumers with different demographic and shopping behavior characteristics. This notion is also supported by recent empirical research on supermarket shopping behavior and store choice. This research suggests that shoppers may alternate visits to different store types according to their purchase plans. For example, Kahn and Schmittlein (1992) link consumers' tendencies to use coupons to whether or not the shopping trip was major or minor. More recently, Bell and Lattin (1998) link consumer preference for shopping in EDLP versus HILO stores to the expected dollar size of the household's shopping basket. Bell, Ho, and Tang (1998) show how EDLP and HILO stores present shoppers with a trade-off between the fixed and variable costs of shopping. For example, their findings show that HILO stores can offer lower total costs for small baskets while EDLP stores offer lower total costs for large baskets. One collective implication of these findings is that shoppers will vary their choice of store format depending upon how much they expect to spend on the shopping trip (e.g., major versus minor).

Switching Between HILO and EDLP Stores

Taken together, the conceptual, theoretical, and empirical arguments above strongly suggest that supermarkets with different price formats compete vigorously with each other – provided, of course, that they are located in the same trading area. One implication of this type of competitive structure is that we should observe a high level of switching among EDLP and HILO stores located nearby. In the empirical analysis below, we present a store switching matrix for households shopping among five competing stores in a trading area in a large midwestern city in the U.S. Two of these stores are from different chains advertising an EDLP format, one is from a chain advertising a HILO format, and the remaining two are from the same chain and follow a high-tier HILO strategy (i.e., they follow a promotional pricing strategy, but charge somewhat higher average prices than the HILO store). Thus, the setting provides a good natural experiment for investigating retail price format competition. As we detail below, we find a large extent of switching between the two EDLP stores and among the three HILO stores. However, we find only a limited degree of switching between stores with different price formats.

This empirical finding suggests that supermarkets with different price formats may not induce extensive store-format switching among shoppers. It further suggests that consumers may not perceive stores with different price formats as close substitutes for each other (i.e., when shoppers do switch, they choose a store of the same format). The implications of this type of market structure for a retailer are quite important: the strategic positioning of a store following a particular price format may circumscribe the boundaries of its market much more so than might be expected on the basis of trading areas and product assortment. If many shoppers are unwilling to switch formats, then promotional tools may be ineffective in reaching large portions of the market, even when the store is conveniently located to them.

Objectives and Overview

The objective of our paper is to pursue this (somewhat unexpected) finding by modeling the store choice decisions of supermarket shoppers and studying segmentation in this behavior. We begin by developing a conceptual framework for the factors influencing choice of price format. We argue that the selection of a preferred store format is driven by shoppers' demographic and behavioral characteristics. Because these characteristics are largely stable over time (i.e., they tend to change slowly, if at all), we also expect to observe a great deal of stability in *store format choice behavior*. Our intended contribution is to develop this reasoning in more depth and provide empirical evidence to support hypotheses that logically stem from it.

Our conceptual framework holds that households will fall into a store choice segment on the basis of their stable characteristics. Early empirical research on store shopping behavior patterns also supports the notion that many aspects of store choice behavior will be stable. For example, Cunningham (1962) found stable patterns of store loyalty in his study of Chicago Tribune panel data. This was later supported by Farley (1968), who found that a factor structure relating demographics, buying activity, and store switching behavior was also stable over time. Regularity in the patterns of store choice has also been revealed by applications of the Dirichlet model to the supermarket setting (e.g., Uncles, Ehrenberg, and Hammond 1995). Thus, there is considerable evidence from earlier literature for the notion that store choice behavior may be both stable and strongly linked to demographic factors. Building upon this, we address the question of whether or not choice of *supermarket price format* may also share many of these stable characteristics.

Recall that Lal and Rao (1997) identified two segments in their model: time-constrained consumers and cherry pickers, both of whom cross-shop the HILO and EDLP price formats. Building upon this as a point of departure, we propose that behavior may be better characterized by *three* segments of shoppers, two of whom do *not* cross-shop formats. Like Lal and Rao, we expect cherry pickers to patronize both EDLP and HILO stores. We depart from Lal and Rao in proposing that time constrained shoppers split into two groups: those seeking high service and those not. We expect that the service seekers will patronize chiefly HILO stores while the non-service seekers will patronize chiefly EDLP stores. Thus, in our framework, cherry pickers would be the only shoppers who switch regularly between EDLP and HILO stores. If they make up only a fraction of the market, the overall pattern of store competition will be primarily intra-format instead of inter-format.[1]

[1]Like Lal and Rao (1997), we implicitly assume that supermarkets with HILO price formats will also provide higher service levels than EDLP stores. Hoch et al (1994) found that HILO stores tend to be more profitable than EDLP stores (over 35 percent higher profits in their analyses), so there is more room for such stores to compete on non-price characteristics such as service levels. Hoch et al also observe that EDLP stores tend to reduce operating costs relative to

Empirically, we seek to determine whether our proposed three-segment framework provides a significantly better representation of the market than existing frameworks calling for fewer segments. We also see to determine the nature of store competition within each segment and how closely it conforms to our conceptual framework. To do this, we develop a multinomial logit model of store choice and estimate it with latent classes. Latent class models have been widely applied to brand choice behavior (e.g., Kamakura and Russell 1989, Bucklin and Gupta 1992, Gupta and Chintagunta 1994), but have only recently begun to be applied to the store choice decision (Bell, Ho, and Tang 1998). For example, a latent segment representation will permit us to assess whether the store competition within each segment is primarily intra-format, inter-format or both.

In the sections that follow, we first present our conceptual framework for store format choice. We then specify our logit model of consumer store choice based on household demographics, behavioral characteristics and marketing-mix variables. Next, we describe the key features of the data used in this study. This is followed by our empirical findings from the switching matrix and from estimating the store choice model with and without latent segments. We conclude by summarizing, reviewing the implications for retail strategy, noting the limitations of this work, and describing opportunities for further research.

Conceptual Development

We open the discussion of our conceptual framework by returning to the paper by Lal and Rao (1997) – one of the recent papers to explicitly address EDLP versus HILO competition. The authors build their model on the basis of two segments of consumers: time constrained shoppers and cherry pickers. Stores compete on price, service and convenience (e.g., time costs), and can attract both segments of consumers by means of a multidimensional strategy. The time constrained shoppers have high opportunity costs for shopping and may choose either an EDLP (if they do not have time to look for price specials) or a HILO store (if they place a high value on service). The cherry-pickers have lower opportunity costs for shopping, are more price-sensitive, and may choose either format depending on the price differences across categories in any given week.

Our framework reconceptualizes Lal and Rao's two-segment view of shopping behavior into

HILO stores, which would be consistent with the provision of lower service levels.

three segments. In Lal and Rao's framework, HILO-EDLP switching by a time constrained shopper would stem from longitudinal changes in preferences for supermarket service.[2] While these may indeed occur within shoppers over time, we expect that preference for service levels also will be quite heterogeneous across shoppers. We note that it was the *cross-sectional* differences in the opportunity cost of time that gave rise to the initial split between cherry pickers and time constrained shoppers. Our argument is that if shoppers also differ in their preferences for service levels, then the time constrained segment should naturally divide into two further sub-groups: one that patronizes lower-service EDLP stores and one that patronizes higher-service HILO stores.

Following the foregoing logic, we therefore expect to observe three segments of store choice behavior: (1) cherry pickers, (2) time constrained shoppers with high service needs (service seekers), and (3) time constrained shoppers with low service needs (time constraineds). We assume that cherry pickers pay more attention to price than quality, and therefore have low service needs. While cherry pickers will engage in inter-format switching, the other two segments are expected to be format loyal and their store switching therefore should be primarily intra-format. Thus, we predict that low service time-constrained shoppers will prefer the EDLP format, high service time-constrained shoppers will prefer the HILO format, and cherry pickers will switch between formats. *In our conceptualization, the overall degree of inter-format competition in a market therefore depends upon the size of the cherry picker segment relative to the other two.*

We now discuss the factors that we hypothesize to be predictive of whether a shopper is a cherry picker, a service seeker, or a time-constrained shopper. (We emphasize that service seekers are also time constrained, but truncate the label for expositional ease.) In this study, we consider three sets of factors: (1) demographics, (2) long-run shopping behavior, and (3) marketing activity in each store. By placing these factors into a latent class analysis of store choice, we can evaluate our predictions based on how well they align, by segment, with the predicted store format selection. Our discussion below is organized around each of these three sets of factors.

Demographic Factors

Previous research has established a number of empirical links between demographic variables and

[2] We define a high service supermarket as one offering higher average quality and price for the store's merchandise versus competition.

store choice and shopping behavior. Gupta and Chintagunta (1994) used demographics to determine consumers' segment memberships in a logit mixture model of brand choice. They found that income and household size significantly affected segment membership probabilities, with low-income, larger households belonging to more price- and promotion-sensitive segments. Hoch et al (1995) showed that both consumers' demographic characteristics and the store's competitive environment were predictive of store-level price elasticity. In particular, they documented that store-level elasticities were related to the following demographic characteristics of store patrons: non-white ethnicity (positive), family size (positive), education (negative), and home value (negative). Finally, Bell, Ho and Tang (1998) reported that segments of shoppers differing in their sensitivity to the fixed and variable costs of shopping had different demographic profiles. In particular, they identified larger, younger families with lower per capita incomes as incurring high fixed costs of shopping (and therefore preferring predominantly EDLP stores).

Based on the above findings, we identify both household income (INC) and family size (SIZE) as likely predictors of store format selection behavior and offer the following hypotheses:

H1 Income (INC) will be negatively related to the choice of EDLP stores and positively related to the choice of HILO stores.

H2 Family Size (SIZE) will be positively related to the choice of EDLP stores and negatively related to the choice of HILO stores.

With respect to characterizing the patterns of segmentation, one key construct is opportunity cost of time. While this is not directly measured, we use income and family size as proxies. Higher income households have a higher opportunity cost of time than low income households (i.e., time spent shopping is traded off against other activities at a different rate). Similarly, large families place time demands on shoppers that make the opportunity cost of shopping higher than for small families. Thus, the cherry picker segment (i.e., the only one that is not time-constrained) should be characterized by low incomes and small family sizes. Combining H1 and H2 with this reasoning suggests that households in the cherry picker segment will do more shopping at EDLP stores than HILO stores while continuing to switch between the two price formats.

On the other hand, the households in the two time-constrained segments should have higher incomes and larger family sizes than the cherry-picker segment. We can also characterize expected differences in income and family size between the high service and low service time-constrained shoppers. Shoppers seeking high retail service are likely to prefer high-quality/high price brands and should therefore have the highest level of income. Time constrained shoppers with low service preferences might be expected to have an intermediate income level. Similarly, we expect the largest families to be in the low service time-constrained segment (with income held constant, their larger shopping baskets should make price matter more than quality) while the families in the service seeking segment should be intermediate in size.

Shopping Behavior Factors

We discuss two ways in which consumers' weekly shopping patterns may differ from each other, as shown in Table 1.

[Table 1 about here]

In the first comparison, consumer A makes 5 shopping trips per week, and spends $50 each time, while consumer B makes only 3 trips of the same size. These two consumers spend the same dollar amount on average but differ only in their shopping frequency rate. In the second comparison, consumers C and D visit the store with the same frequency but differ only in their average trip spending. In reality, consumers may differ along both dimensions at the same time. We incorporate both variables in our models, and estimate the effects of each of them on consumers' store choices.

The first variable, consumers' shopping frequency rate, controls for the number of opportunities for the shopper to find special deals in the categories of interest. Consistent with Hoch et al (1995), we expect that the higher the shopping frequency (RATE), the greater the number of such opportunities, and the more likely consumers are to visit HILO stores to take advantage of their frequent grocery specials. This leads to our next hypothesis:

H3 Shopping frequency (RATE) is negatively related to the choice of EDLP stores and positively related to the choice of HILO stores.

In terms of segmentation, we expect higher shopping frequencies to go hand-in-hand with cherry picking behavior. Conversely, we expect households in the time constrained segments to manage their shopping so as to take the fewest trips possible. Service seeking shoppers visiting HILO stores, however, are expected to take a somewhat larger number of trips than the time-constrained shoppers who visit EDLP stores.

The second variable, the average amount spent per trip (AVE), is a measure of the panelist's average shopping basket size. As noted above, previous work has identified this factor as a major determinant of store choice. Kahn and Schmittlein (1992) found that there were significantly more fill-in trips than major trips at non-preferred stores, whereas major shopping trips had a tendency to occur at preferred stores. They argued that this helps explain differences in elasticities for promotional variables such as features and coupons. While the former were more effective on fill-in trips, the latter were used more often on major trips.

Bell and Lattin (1998) demonstrated that consumers' expected size of their shopping basket was predictive of their choice of EDLP versus HILO stores. Consumers expect EDLP stores to have a lower price level across all categories and a lower price variance over time. Given this expectation, large basket shoppers, who have a higher probability of purchase for any given category, prefer the format with a lower average basket price (i.e., EDLP). Conversely, small basket shoppers, who have a lower probability of category purchase, prefer the store with more price variation (i.e., HILO). Following this result, we state

H4 Average spending per trip (AVE) is positively related to the choice of EDLP stores and negatively related to the choice of HILO stores.

With respect to segmentation, we expect that basket size should be largest for time constrained shoppers patronizing primarily EDLP stores and smallest for service seeking shoppers patronizing primarily HILO stores. Cherry pickers, who shop at both HILO and EDLP stores, might be expected to have intermediate basket sizes (because they have more time to visit different stores than time-constrained shoppers and divide their spending accordingly).

Marketing Variables

We estimate the effect of two time-varying marketing variables that we expect to influence store choice. The first is the store's price level (PRICE), measured by calculating the average price of a basket of SKUs which were offered by all stores in any given week. The second is the number of features (FEAT) run by each store in any given week. Note that we do not include in-store displays, as those are revealed only to the shopper once inside the store. With these variables we state two final hypotheses:

H5 The basket price level (PRICE) is negatively related to the likelihood of choosing any store.

H6 The number of features (FEAT) is positively related to the likelihood of choosing any store.

In the segment-level model, we expect that the effect of price will be strongest for the cherry picker households, the most price sensitive of the three segments. As a key input to the cherry picker's store choice decision, we also expect feature advertising to have a strong effect on store choice in that segment. On the other hand, households in the two time-constrained segments, with little time to consult feature ads, should show small or insignificant effects for the FEAT variable.

In the table below, we summarize our expectations regarding the alignment of the six predictor variables with the three-segment conceptualization of store choice behavior.

Factor or Variable	*Cherry-Picker Segment*	*Service-Seeking Time-Constrained Segment*	*Time-Constrained Segment*
Income (INC)	Low	High	Intermediate
Family Size (SIZE)	Small	Intermediate	Large
Shopping Rate (RATE)	High	Intermediate	Low
Average Basket Size (AVE)	Intermediate	Small	Large
Weekly Price Level (PRICE)	−−	−	−
Feature Ad Activity (FEAT)	++	+	0

Note that the demographic variables, INC and SIZE, delineate the cherry pickers from the other two segments while the behavioral variables, RATE and AVE, delineate the time-constrained segment from the other two. Thus, the predicted influence of these variables permits us to characterize the anticipated nature of the three segments of shoppers. This also provides another justification for expecting three segments to better represent store choice behavior than two segments.

Model Specification

We use a multinomial logit choice model to explain consumers' store choices given that a shopping trip has taken place. We then extend this model to include latent segments. To operationalize the store choice model, we formulate a store-specific and time varying utility function for each household and each store. To control for the effect of travel time on store choice, we include a variable measuring each household's distance to the store (DIST) and expect it to be negatively signed. In addition to DIST, we include the three sets of predictor variables above, demographic (INC and SIZE), behavioral (RATE and AVE), and marketing (PRICE and FEAT).

In contrast to much of the tradition in logit modeling, we do not include a store loyalty variable in our model specification. Since it is designed to capture heterogeneity in revealed preference, a loyalty variable would effectively replace all of the stable predictor variables in our model, thereby making it impossible to reveal their effects on store choice. Consequently, the decision to include or not include a loyalty variable depends upon the objectives of the study. We argue that loyalty variables need to be omitted when the objective is to understand factors influencing preference (or preference segmentation). When study objectives center on studying factors other than preference, loyalty variables can help produce parsimonious segmentation schemes and identify segmentation in response to marketing variables (e.g., Bucklin and Gupta 1992). In the analysis presented in Bell, Ho, and Tang (1998), for example, a store loyalty variable is included in their model. Because our study is aimed at *identifying* determinants of stable behavior in store choice (i.e., aspects of loyalty itself), we do not include the variable.

Turning to other variables in the model, note that DIST, PRICE and FEAT are store-specific. This means that their effects can be assessed by estimating only one parameter for each variable in the logit model. On the other hand, INC, SIZE, RATE and AVE are household-specific variables that do not vary over time or across stores. Following Greene (1996), we can incorporate such variables in the logit model by creating one dummy variable for each alternative minus one (to avoid singularity). Given the five store alternatives in our data, we use four dummy variables for each explanatory variable, and estimate one parameter for each.

The store utility function in the logit model is given below:

$$U_t^h(s) = \alpha_s + \beta_1 \cdot DIST_t^h(s) + \beta_{2s} \cdot INC^h + \beta_{3s} \cdot SIZE^h$$
$$+ \beta_{4s} \cdot AVE^h + \beta_{5s} \cdot RATE^h + \beta_6 \cdot PRICE_t(s) + \beta_7 \cdot FEAT_t(s) \quad (1)$$

where $U_t^h(s)$ is the utility that household h assigns to store s on shopping occasion t. The variables in equation (1) are defined as follows:

$DIST_t^h(s)$	=	distance, in miles, of household h to store s,
$RATE^h$	=	shopping frequency rate for h, computed as store visits per week,
INC^h	=	income, in dollars, for household h (divided by 10,000),
$SIZE^h$	=	family size for household h,
AVE^h	=	average size, in dollars, of shopping basket purchased by h,
$PRICE_t(s)$	=	price in store s, week t, for a basket of SKUs in all stores, and
$FEAT_t(s)$	=	number of SKUs featured by each store in week t.

Note that with the parameters for store 1 set equal to zero to avoid singularity, we estimate four parameters $\{\beta_{vs}\}$ for each of the demographic and shopping behavior variables (where the subscript s goes from 2 to 5). The variables DIST, PRICE and FEAT have only one parameter each.[3]

The probability that household h chooses to visit store s at trip t is then given by the standard multinomial logit formulation (Guadagni and Little 1983):

$$P_t^h(s) = \frac{\exp(U_t^h(s))}{\sum_r \exp(U_t^h(r))}. \quad (2)$$

In this model, which we label the aggregate store choice model, all households share the same vector of parameters for the store choice utility function (equation 1). Next, we specify a latent segments store choice model for our data. This model fits segment-specific parameters for the utility model

[3]The location information for stores and households in our dataset is limited to zipcode. The variable for distance is computed as the number of miles between the household and the store as measured from the centroid of the store's zipcode and the centroid of the household's zipcode. Though the distance measure is somewhat crude, it is nevertheless highly predictive of store selection in the choice model.

in equation (1). For example, a two-segment model would call for the estimation of two separate vectors of coefficients, one pertaining to each segment. A three-segment model entails three vectors of coefficients, and so on. (For a more detailed discussion see, for example, Kamakura and Russell 1989). In the multi-segment model, the probability that household h chooses store s on shopping occasion t is given by

$$P_t^h(s) = \sum_{j=1}^{J} P_t^h(s|j) \cdot P^h(j) \tag{3}$$

where $P_t^h(s|j)$ is the probability of choosing store s, conditional on membership in segment j, and $P^h(j)$ is the prior probability for each segment j. Note that the parameters describing the store choice probabilities, $P_t^h(s|j)$, will be specific to each segment j. All parameters are estimated by maximum likelihood methods.

We have two objectives for the latent segment model. The first objective is to verify that store choice segments can be represented by means of the specific demographic and behavioral characteristics of their members, as hypothesized. This will shed light on our empirical conjecture that store choice behavior for competing price formats is driven by slow-moving demographic and behavioral factors and is therefore likely to be more stable than recognized heretofore in the literature. Our second objective is to determine whether the hypothesized three-segment structure of store format choice provides a better representation of store choice behavior than a two-segment structure. We will assess this using model fit criteria.

Before proceeding, we note that our modeling approach to incorporating demographics in the latent segmentation differs from an alternative approach develop for use in brand choice (Gupta and Chintagunta 1994). In our approach, the demographic and behavioral variables contribute to the store utility, $U_t^h(s)$, rather than to the prior probability of segment membership, $P^h(j)$. Our approach therefore permits the model to predict store choice and to identify segments based on both stable as well as time varying factors. This joint formulation enables the model to capture the direct effect of stable (i.e., demographic) characteristics on store choice. Unlike brand choice where these variables have small explanatory power, the strong effects expected in store choice support their inclusion into the store utility function. Our approach also avoids the need to determine, *a priori*, which variables affect segment membership and which variables affect store choice. As a result, the segments are identified not only on the basis of their members' sensitivity to distance and marketing variables, but also by the contribution of each household's demographic and behavioral characteristics to the utility of each store. In sum, we feel that this approach is better suited to the problem of predicting and understanding store choice than the Gupta and Chintagunta (1994) approach, which was developed for use in brand choice.

Data

We estimate our model using a portion of a dataset provided by Information Resources, Inc., known as the Stanford market basket data. The selected dataset comprises scanner panel data from 34,591 shopping trips, 488 panelists, and 24 categories for five supermarkets located in a suburban area of a major midwestern city in the U.S. The data cover the period from 1991 to 1993 and we use one year of the data to fit our model.

We refer to the five stores in the data set as follows (proprietary restrictions prevent the release of store names). The two EDLP stores are EDLP1 and EDLP2, two "high-tier" promotion stores are HIHI1 and HIHI2, and the third promotional store is HILO. The HILO store has a lower average basket price than the two HIHI stores and is therefore positioned somewhat between the EDLP price format and the two HIHIs. HIHI1 and HIHI2 are from the same chain; EDLP1, EDLP2, and HILO each belong to different chains.

To quantify the nature of the competing retail price formats in this market, we calculated the mean and the standard deviation of the average price of a product basket comprising those SKUs that were on the shelves of all stores in any given week. We report these in Table 2. The computations confirm that the HILO stores are more expensive and have greater price variability than the EDLP stores and that the HIHI stores are positioned somewhat more expensively than the HILO store.

[Tables 2 and 3 about here]

In Table 3, we present some additional descriptive statistics about the five stores. For example, panelists spent a larger dollar amount and purchased a larger number of categories when visiting

EDLP versus HILO stores. Note that the small number of categories purchased per trip is due to the fact that they are counted over the available basket sample of 24 categories while each store, of course, actually stocks hundreds of categories.

Empirical Analysis and Estimation Results

We begin our empirical analysis by first presenting the store switching matrix for the five stores in the data sets, computed over the calibration period (please see Table 4). We present the matrix in row-conditional form, so that the relative switching proportions can be readily discerned. The repeat-visit rates for each store are given by the diagonal elements and the switching rates between stores are given by the off-diagonal elements. In Table 5, we report the distances, in miles, between each pair of stores. (The distances are based on reported zipcodes and reflect the centroid-to-centroid distance between zipcodes. This means that stores located in the same zipcode will show zero distance between them.)

Looking first at the switching matrix in Table 4, we observe more switching within formats than between formats. On any given shopping trip, shoppers have probabilities of choosing a different store from the one they shopped last time. The store that is switched to, however, is much more likely to share the same price format than to have a different format. This result holds regardless of the distance among the stores – i.e., shoppers appear more willing to switch to a store of the same format even though it is not the next closest alternative. For example, EDLP2 customers switch primarily to EDLP1 even though the HILO stores are closer to EDLP2 than EDLP1. Similarly, HIHI2 customers switch primarily to HILO even though the closest alternative to HIHI2 is EDLP2. These patterns suggest that price format preferences could have even stronger influences on store choice than locational convenience.

[Tables 4 and 5 about here]

Estimation Results for the Aggregate Store Choice Model

We first briefly describe the results from fitting the aggregate store choice model, in which no segmentation is incorporated. Table 6 reports the values of the likelihood function, the number of parameters, the Bayesian Information Criterion (BIC)[4] and the degree of improvement over an intercept-only null model (as given by the U^2 statistic) for each of a series of nested specifications.[5]

[Table 6 about here]

The first model reported in Table 6 relates store choice only to distance (DIST). This should be considered the null or benchmark model against which the additional explanatory power of demographic, behavioral, and marketing mix variables can be compared. The second model adds the demographics, income (INC) and family size (SIZE), and produces a significant improvement in fit. All coefficients are correctly signed and statistically significant. The third model adds basket spending, AVE, and shopping frequency, RATE. Again, all parameters are statistically significant and the U^2 jumps to 28%. The last model adds two marketing variables: (1) the weekly average price of a basket of all those SKUs which were on the shelves of all stores in any given week and (2) the number of features run by each store in any given week. As shown in Table 6, the improvements in the log likelihood function and the value of U^2 are marginal. Nevertheless, the parameter for the PRICE variable is negative, and that for the FEAT variable is positive, both significant at the 0.05 level. In sum, the results for the aggregate store choice model show that the distance, demographic, behavioral, and marketing variables all contribute significantly to the prediction of store choice behavior.

Latent Segments Store Choice Model

We now turn to the estimation results for the latent segment store choice model. We wish to investigate whether the segments are related to store price formats, and if so, what the characteristics of the segment members might be. For example, if two segments are identified, each cross-shopping EDLP and HILO stores, then Lal and Rao's segmentation scheme of time-constrained shoppers and cherry pickers is supported. On the other hand, our conceptual framework posits that at least three distinct latent segments should be found, one HILO format loyal, one EDLP format loyal, and one

[4]The Bayesian Information Criterion (Schwarz 1978) selects the best model formulation based on a penalized likelihood function value. It is widely used in choice model selection and has excellent performance as a selection criterion (Rust et al 1995).

[5]The parameters for each specification are available from the authors upon request.

that switches between both formats.

The results of several runs of the model, from one to five segments, are reported in Table 7. The table shows large improvements in the values of the log likelihood function, BIC and U^2 with the introduction of a second segment and then a third segment. The fit improvements, while statistically significant, begin to narrow with the fourth and fifth segments. Given that we have not included a loyalty variable in our logit specification, and that our household-specific variables explain only part of the loyalty to a specific store, we believe that the model providing the best BIC value may well include more than six segments.[6] While statistical segmentation is virtually costless, segmentation strategy is not, and we must be careful to focus our attention on those segments that are both conceptually interpretable and economically meaningful.

[Table 7 and Figure 1 about here]

To illustrate this point, we present in Figure 1 a graph of the improvements in the U^2 statistic as a function of the number of segments in the latent class model. Its interpretation is analogous to that of a Scree plot in factor analysis. The "elbow" occurs between three and four segments, with the fit improvements offered by the addition of a third segment outpacing the fit improvements offered by subsequent segments. Thus, an important result from the fitting of the latent segment models is *the convincing rejection of a two-segment solution in favor of a three- (or more) segment solution.* Based on the elbow in Figure 1, the interpretability of model parameters, and a desire for parsimony and managerial relevance, we will discuss the results obtained from the three-segment solution.

[Table 8 about here]

Table 8 reports the parameter estimates for the three-segment store choice model. The segmentation results reveal substantial heterogeneity in response to the store-specific model variables. Note that DIST is not significant for segment 2, but strongly significant in segments 1 and 3. Also note

[6] Due to the large number of store visits made by each panelist, the model has much more information to work with than in the case of brand choice, leading to the identification of a much greater number of segments.

that the FEAT variable is not significant in segment 3. For the parameters of the household-specific demographic and shopping behavior variables, we also report, for each segment, the marginal effect of each variable on the store choice probability alongside the coefficient estimate.[7]

To aid in the interpretation of the demographic and shopping behavior patterns for each segment, we assign the households in our sample to the segment for which their posterior membership probability is the highest and compute segment means for the variables.[8] The resulting average variable values for each segment are reported in Table 9 along with the segment mean probabilities of choosing each store and the segment mean distances to each store.

[Table 9 about here]

Given the parameter estimates (Table 8), the segments' posterior average variable values (Table 9), and the segments' store choice probabilities (Table 9), we can now characterize the nature of the shopping behavior in each of the three segments. We begin by noting that, for each segment, the marginal coefficients of the demographic and behavioral variables generally have higher magnitude whenever they refer to the stores with the higher probabilities of being chosen. The signs of such coefficients reflect the relative preferences of the shoppers within each segment, not across segments. For example, the RATE marginal coefficient in segment 3 (time-constrained shoppers) is negative for EDLP2 but positive for EDLP1. This result means that shoppers preferring EDLP1 visit the store more frequently than shoppers preferring EDLP2 (though both groups shop less frequently than households in the other two segments).

Segment One: "Cherry Pickers" (40 percent of households). This segment is the only one of the three with a robust pattern of switching between store formats. Though EDLP1 dominates (69.2%),

---

[7]The sign of the coefficient of each variable on the probability of choosing any one store is not necessarily the same as the sign of the corresponding response to that variable. Following the procedures in Greene (1996), we compute the marginal effect on the store s choice probability of a one-unit change of the underlying variable X_v, according to the formula $\frac{\delta P(s)}{\delta X_v} = (\beta_{vs} - \bar{\beta}) \cdot P(s)$. These values have the same meaning as the betas of the classical linear regression model and are reported in the columns of Table 8 labeled "marginal."

[8]The posterior membership probability is the Bayesian product of the prior membership probability $P^h(j)$ times the likelihood of observing their choices given membership in segment j. It is computed according to the following formula: $Post^h(j|s) = \frac{P^h(j) \cdot L^h(s|j)}{\sum_{j=1}^{J} P^h(j) \cdot L^h(s|j)}$.

both HILO (12.8%) and HIHI2 (15.9%) receive significant patronage. The group is characterized by the lowest income levels, smallest family size, moderate basket size, and the highest shopping frequency. An interesting aspect of the results for this segment is that the most preferred store (EDLP1) is, on average, the farthest away. The presence of inter-format switching together with the pattern of demographic and shopping behavior characteristics fit with the predicted traits of the cherry picker segment. Lastly, we note that, as expected, price and feature are both predictive of store choices for the cherry pickers, though the t-statistic for feature is marginal.

Segment Two: "Service" (22 percent of households). This segment – the smallest of the three – includes only HILO-type shoppers, with HILO (47.3%) and HIHI1 (48.5%) equally dividing the patronage of the shoppers here. Members of this segment have the lowest average spending per trip, the highest incomes, and intermediate family size and frequency rate, all of which are consistent with the predictions for the service-seeking segment. An analysis of the parameter signs within this segment (Table 8) suggests that larger households with lower shopping frequency favor visits to HILO while HIHI1 tends to attract shoppers with the opposite characteristics. This finding corroborates the slight expected differences between HILO and HIHI stores. HILO's price format positions it midway between the EDLP and HIHI stores and consequently it is likely to attract shoppers with a profile that leans towards that of the EDLP customer.

Segment Three: "Time-Constrained Consumers" (38 percent of households). This segment, at 37.9% of the sample, includes primarily EDLP shoppers. The EDLP2 store dominates (61.9% within-segment share), but there is substantial patronage of EDLP1 (22.1%). The households in this segment have the largest family size, largest average basket size, lowest shopping frequency, and intermediate income level. These predictions are all consistent with the anticipated characteristics of time-constrained consumers who are not seeking high service. The signs of the marginal coefficients are as expected for EDLP2 (i.e., the store most likely to be selected), and are opposite for EDLP1. This result indicates a slight difference in the customer base of the two stores, and suggests that EDLP2 is more attractive to time-constrained households than EDLP1.

An across-segment interpretation of the parameters is meaningful only for those pertaining to the three variables that vary on each shopping occasion (i.e., Distance, Price and Feature). From Table 8 we note that DIST is negative and significant in segment one and segment three, but not significant in segment two. This is likely due to the fact that the two stores patronized by households in segment two (i.e., HILO and HIHI1) are located in the same neighborhood (see Table 3).

The price variable is negatively related to the choice of any store. Segments one and three, however, are more price sensitive than segment two. This fits with expectations given the higher income profile of segment two. Note also that EDLP stores are patronized only by segments one and three. This indicates that EDLP shoppers are sensitive to the overall basket price for a store (as captured by PRICE) while HILO shoppers are less so. This result suggests that basket price promotion may be more effective for EDLP stores than HILO stores and corroborates the findings in Bell and Lattin (1998). Nevertheless, PRICE is still significant for shoppers in segment two, potentially reflecting the fact that the promotional activities of the HILO stores do not go entirely unnoticed by the service-seeking shopper.

Lastly, FEAT is positively related to the choice of any store in segments one and two, though the significance level of the coefficient drops to the 0.10 level on the two-tail test. The parameter is not significantly different from zero in segment three. This pattern is again in line with expectations regarding the shopping behavior for each segment. The coefficient approaches significance only for those segments in which HILO stores are patronized significantly. The insignificance of FEAT in segment three fits with the smaller number of features run by EDLP stores, the time constraints on the shoppers in segment three, and the low information value of a feature ad from an EDLP store (due to low price variance).

In sum, the latent class analysis generates a detailed picture of the composition of the customer base for each store and store format. Our model offers confirming evidence for our hypothesis that panelists' patterns of store switching will be more intra-format than inter-format. Sixty percent of the households in our sample are classified into format loyal segments (22 percent plus 38 percent). Moreover, households in segment one have an approximately 70/30 preference for EDLP versus the HILO stores, further reducing the extent of inter-format switching in this market.

The latent class results also offer strong support for the view that segmentation in store choice behavior is better characterized by three segments than two segments. Though the three-segment solution presented here does not necessarily outperform higher-order segmentation solutions, it offers a very good fit compared with these representations. More importantly, the substantive predictions derived from reconceptualizing the Lal and Rao framework into three segments are very well supported by the pattern of our results.

Conclusion

The objective of this paper has been to enhance our understanding of store choice behavior in markets where supermarkets compete using different price formats. Our modeling results support the notion that store choice and the choice of the store format depend on stable demographic and behavioral characteristics of the consumers in a given market area. Marketing variables, as captured here, have a significant but lesser impact on store choice. These results challenge the view of multidimensional retail positioning strategy which holds that store-level marketing activity can be effective in attracting most types of consumers to a store. Rather, it appears for these data that the store's price format may limit its appeal to those shopper segments with particular demographic and behavioral traits.

Our model has also pointed out the usefulness of segmenting the market on the basis of these same demographic and shopping behavior variables (as opposed to consumers' sensitivity to store-level marketing activity alone). Our latent class analysis revealed the following three well-defined segments:

1. Cherry-pickers. Households with low opportunity costs of shopping (low income, small family size, high shopping frequency) who patronize both EDLP and HILO formats in search of the lowest prices.

2. Service-seeking shoppers. Households who patronize HILO stores seeking high retail service levels (high income, small basket size, high shopping frequency).

3. Time-constrained shoppers. Households with high opportunity costs of shopping (large family size, large basket size, low shopping frequency) who patronize primarily EDLP stores.

An interesting result from the latent class analysis is that each segment patronizes a different combination of stores. In segments two and three, these stores share the same price format while in segment one they have different formats. The segmentation helps understand the patterns found in the store switching matrix (i.e., that most store switching occurs between stores with the same price format).

Our model can also reveal the relative positioning of each store in each market segment. These results are important for retail management. For example, the segmentation results in Table 9 show that EDLP1 is strongly positioned in the cherry-picker segment, but more weakly positioned in the time-constrained segment. It is therefore more dependent upon the lower income, cherry-picking shoppers than EDLP2. Consequently, the shoppers in EDLP1 may be a less attractive long-term clientele than the shoppers in EDLP2. If store patronage is stable, as our results suggest, this raises the concern that EDLP1 may already be at a long-term competitive disadvantage to EDLP2.

Another example involves the two HIHI stores. Though they are from the same chain, Table 9 reveals that HIHI1 and HIHI2 draw business very differently across the three segments. HIHI1 receives almost all of its patronage from the service-seeking segment while HIHI2 draws most of its business from the cherry-picker segment. Management at this chain therefore confronts the challenge of maintaining a consistent image across the two stores while catering to the different benefits sought by each store's clientele. Taken together, findings such as these show how the segmentation analysis can help retailers better understand their competitive strengths and weaknesses.

There are, of course, a number of limitations of our modeling approach and caveats to our findings. One limitation is generalizability. Our model was estimated using data from the trading area of a single suburban market. It should therefore be replicated in other markets, both urban and suburban, to verify that the level of explanatory power is maintained, and to estimate the same parameters in different contexts.

A second potential limitation is the reliance on 24 categories (i.e., those available in the market basket data set) to provide a good sampling basis for the rest of the store.[9] While these categories were carefully selected to provide a good, yet parsimonious, representation of the supermarket, they do not include non-scannable items from the meat, produce, seafood, and deli departments. Obtaining information on perishable and service items will be important for understanding the nature of service

[9]More detail on the composition of the data set is reported in Bell and Lattin (1998).

competition among supermarkets. We note that our analysis depends on the 24 categories to construct the price and feature variables and to verify the different price formats followed by the competing stores. Thus, our findings regarding the effects of demographic and shopping behavior factors on store choice do not depend upon on the quality of the market basket representation.

A third limitation is that our multi-segment logit framework incorporating both store- and household-specific variables is bounded by the large number of parameters to be estimated. We note that other demographic variables and shopping behavior characteristics might be added to the analysis and, if so, could potentially change the character of the revealed segmentation. Estimating more complex store choice models with latent segments currently lies beyond the range of computational tractability.

In addition to overcoming some of the limitations noted above, we see several worthwhile directions for future research. We believe that additional study of the same market basket data can yield further insights into store choice and price format competition. For example, the intra-format switching among EDLP stores in segment 3 is of interest. Though switching in pursuit of lower prices can explain the pattern in segment one, it is not immediately clear why shoppers would choose to switch among two stores pursuing an EDLP format. Second, this study has focused solely upon the factors influencing the store choice decision. Also of interest to retailers is the shopper's in-store response to marketing activity, which may increase or decrease the total dollar amount spent on a given store visit. Finally, our ability to understand store choice behavior is limited when studies are confined to scanner panel data. The introduction of new behavioral or attitude variables into the prediction of store choice is a natural extension of this work and would entail joint collection and analysis of both survey and transaction information.

REFERENCES

Bell, David R., Teck-Hua Ho and Christopher S. Tang (1998), "Determining Where To Shop: Fixed and Variable Costs of Shopping", *Journal of Marketing Research*, 35 (August), 352-369.

Bell, David R. and James M. Lattin (1998), "Shopping Behavior and Consumer Preference for Store Price Format: Why 'Large Basket' Shoppers Prefer EDLP," *Marketing Science*, 17 (1), 66-88.

Bucklin, Randolph E. and Sunil Gupta (1992), "Brand Choice, Purchase Incidence and Segmentation: An Integrated Modeling Approach", *Journal of Marketing Research*, 29 (May), 201-215.

Corstjens, Marcel and Judy Corstjens (1994), *Store Wars*, New York: John Wiley.

Cunningham, Ross M. (1962), "Consumer Loyalty to Store and Brand," *Harvard Business Review*, 40 (November-December), 127-37.

Farley, John U. (1968), "Dimensions of Supermarket Choice Patterns," *Journal of Marketing Research*, 5 (May), 206-8.

Greene, William H. (1996), *Econometric Analysis*, New York: Macmillan.

Guadagni, Peter and John D.C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data", *Marketing Science*, 2 (Summer), 203-238.

Gupta, Sachin and Pradeep K. Chintagunta (1994), "On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models", *Journal of Marketing Research*, 31 (February), 128-136.

Hoch, Stephen J., Xavier Dreze, Mary E. Purk (1994), "EDLP, Hi-Lo, and Margin Arithmetic", *Journal of Marketing*, 58 (October), 16-27.

Hoch, Stephen J., Byung-Do Kim, Alan L. Montgomery, and Peter E. Rossi (1995), "Determinants of Store-Level Price Elasticities", *Journal of Marketing Research*, 32 (February), 12-29.

Information Resources, Inc. (1993), *Managing Your Business in an EDLP Environment*, Chicago, IRI.

Kahn, Barbara E. and David C. Schmittlein (1989), "Shopping Trip Behavior: An Empirical Investigation", *Marketing Letters*, **1** (December), 55-70.

Kahn, Barbara E. and David C. Schmittlein (1992), "The Relationship Between Purchases Made on Promotion and Shopping Trip Behavior", *Journal of Retailing*, 68 (Fall), 294-315.

Kamakura, Wagner and Gary Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure", *Journal of Marketing Research*, 26 (November), 379-390.

Lal, Rajiv and Ram Rao (1997), "Supermarket Competition: The Case of Every Day Low Pricing", *Marketing Science*, 16 (Winter), 60-80.

Rust, Roland T., et. al. (1995), "Model Selection Criteria: An Investigation of Relative Accuracy, Posterior Probabilities, and Combinations of Criteria," *Management Science*, 41 (February), 322-333.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461-464.

Uncles, Mark, A.S.C. Ehrenberg and Kathy Hammond, "Patterns of Buyer Behavior: Regularities, Models, and Extensions," *Marketing Science*, 14 (3), G71-G78.

Table 1

Illustrative Weekly Shopping Patterns

Consumer	*A*	*B*	*C*	*D*
$ *spent/day*	50	50	40	60
$ *spent/day*	50	–	–	–
$ *spent/day*	50	50	40	60
$ *spent/day*	50	–	–	–
$ *spent/day*	50	50	40	60
Shopping Frequency	5/7	3/7	3/7	3/7
Average Spending	50	50	40	60
Total Spending	250	150	120	180

Table 2

Statistics on the Price for Identical SKU Baskets

Store	*EDLP1*	*EDLP2*	*HILO*	*HIHI1*	*HIHI2*
Mean	$47.70	$47.82	$53.53	$55.69	$55.77
Standard Deviation	$0.53	$0.59	$0.73	$0.63	$0.62

Table 3

Statistics on Shopping Trips

Store	*EDLP1*	*EDLP2*	*HILO*	*HIHI1*	*HIHI2*
Total Sales	$422,866	$592,280	$527,975	$191,067	$181,706
Total Trips	15,642	19,346	31,706	12,318	9,642
Trip Share	17.8%	22.4%	34.5%	13.7%	11.7%
Average Trip Ticket	$27.04	$30.62	$16.65	$15.51	$18.85
Ticket Std. Dev.	$22.90	$24.55	$16.62	$17.41	$17.81
Average N. Categ./Trip	3.37	3.55	1.98	1.60	1.90

Table 4

Row Conditional Store Switching Matrix

(All Figures in Percent)

Store	*EDLP1*	*EDLP2*	*HILO*	*HIHI1*	*HIHI2*	*Total*
EDLP1	75.4	22.2	1.1	0.1	1.3	100
EDLP2	17.6	73.7	4.8	1.1	2.7	100
HILO	0.6	2.9	76.7	15.6	4.1	100
HIHI1	0.1	1.8	40.8	55.2	2.1	100
HIHI2	2.3	5.0	13.4	2.8	76.5	100

Table 5

Distances Between Stores (in miles)

Store	*EDLP1*	*EDLP2*	*HILO*	*HIHI1*	*HIHI2*
EDLP1	–	3.50	4.58	4.58	3.50
EDLP2		–	2.25	2.25	0
HILO			–	0	2.25
HIHI1				–	2.25
HIHI2					–

Table 6

Model Fits for Aggregate Store Choice Models

Model	*Number of Parameters*	*Log Likelihood*	*BIC*	U^2
1 - Distance only	5	-42372	-42398	0.20
2 - Distance and Demographics (DIST, INC, SIZE)	13	-39670	-39738	0.25
3 - Distance, Demographics, and Shopping Behavior (DIST, INC, SIZE, AVE, RATE)	21	-38073	-38182	0.28
4 - Full Model Specification (DIST, INC, SIZE, AVE, RATE, PRICE, FEAT)	23	-38061	-38181	0.28

Table 7

Model Fits for Latent Segment Store Choice Models

Model	*Number of Parameters*	*Log Likelihood*	*BIC*	U^2
1 Segment	23	-38061	-38181	0.28
2 Segments	47	-27374	-27620	0.48
3 Segments	71	-22718	-22089	0.57
4 Segments	95	-20835	-21331	0.61
5 Segments	119	-18800	-19421	0.64

Table 8
Parameters Estimates for the Three-Segment Store Choice Model

Parameter	*Segment 1* *beta*	marginal	*Segment 2* *beta*	marginal	*Segment 3* *beta*	marginal
DIST	-1.470		0.004		-0.385	
	(-70.5)		(0.2)		(-43.7)	
INC_1		-0.151		0.000		0.040
INC_2	-0.171	-0.002	-0.576	0.010	-0.191	-0.007
	(-5.1)		(-7.1)		(-13.9)	
INC_3	0.737	0.067	-1.019	-0.021	-0.110	0.005
	(22.6)		(-12.9)		(-4.4)	
INC_4	0.881	0.010	-0.982	-0.004	1.816	0.000
	(25.7)		(-12.4)		(18.2)	
INC_5	0.703	0.077	-0.108	0.016	-0.629	-0.038
	(24.0)		(-1.4)		(-19.0)	
$SIZE_1$		0.252		0.000		-0.033
$SIZE_2$	0.690	0.006	-1.099	0.004	0.269	0.073
	(14.9)		(-9.5)		(12.3)	
$SIZE_3$	-1.257	-0.115	-0.795	0.221	0.485	0.025
	(-18.3)		(-7.2)		(12.3)	
$SIZE_4$	-0.916	-0.008	-1.730	-0.226	0.490	0.000
	(-10.6)		(-15.3)		(8.5)	
$SIZE_5$	-1.217	-0.136	-1.210	0.001	-0.624	-0.065
	(-21.8)		(-10.6)		(-11.5)	
AVE_1		0.020		-0.000		-0.003
AVE_2	-0.066	-0.000	0.298	-0.002	0.026	0.007
	(-7.7)		(7.9)		(10.2)	
AVE_3	-0.093	-0.008	0.326	-0.025	-0.076	-0.007
	(-13.9)		(9.0)		(-12.1)	
AVE_4	-0.166	-0.002	0.439	0.029	-0.795	-0.000
	(-20.1)		(12.1)		(-13.6)	
AVE_5	-0.090	-0.010	0.322	-0.001	0.053	0.003
	(-13.3)		(9.0)		(10.6)	
$RATE_1$		-0.113		-0.000		0.071
$RATE_2$	-1.454	-0.009	2.330	-0.019	-0.521	-0.124
	(-10.0)		(7.2)		(-16.8)	
$RATE_3$	0.583	0.054	2.186	-0.438	-0.348	-0.002
	(5.3)		(7.2)		(-6.9)	
$RATE_4$	0.974	0.012	4.074	0.466	-3.860	-0.000
	(8.8)		(13.4)		(-12.1)	
$RATE_5$	0.521	0.057	2.626	-0.009	0.338	0.056
	(4.6)		(8.5)		(9.2)	
PRICE	-0.912		-0.661		-0.968	
	(-2.7)		(-2.1)		(-2.2)	
FEAT	0.030		0.027		-0.001	
	(1.8)		(1.9)		(-0.1)	

Note: t-values in parentheses. Subscript numbers correspond to stores as follows: 1-EDLP1, 2-EDLP2, 3-HILO, 4-HIHI1, and 5-HIHI2.

Table 9
Segment-Level Values for Explanatory Variables, Store Choice Probabilities, and Distances

Variable	*Segment 1* *"Cherry Pickers"*	*Segment 2* *"Service"*	*Segment 3* *"Time"*
Demographic and Shopping Behavior Means			
INC	31,943	43,325	39,744
SIZE	2.14	2.47	2.87
AVE	28.04	24.71	33.32
RATE	1.51	1.36	1.14
Effect Size of Marketing Variables			
PRICE	-0.912	-0.661	-0.968
FEAT	0.030	0.027	-0.001
Store Choice Probability Means			
EDLP1	69.2%	0.0%	22.1%
EDLP2	0.6%	2.4%	61.9%
HILO	12.8%	47.3%	7.5%
HIHI1	1.5%	48.5%	0.0%
HIHI2	15.9%	1.8%	8.5%
Mean Distances			
DIST 1	3.76	4.75	3.15
DIST 2,5	2.43	3.07	2.95
DIST 3,4	2.03	2.61	3.92
Segment Sizes	40.2%	21.9%	37.9%

Figure 1
Relative Improvements in U^2
by Number of Latent Segments