DAVID R. BELL, TECK-HUA HO, AND CHRISTOPHER S. TANG*

The authors develop and test a new model of store choice behavior whose basic premise is that each shopper is more likely to visit the store with the lowest total shopping cost. The total shopping cost is composed of fixed and variable costs. The fixed cost is independent of, whereas the variable cost depends on, the shopping list (i.e., the products and their respective quantities to be purchased). Besides travel distance, the fixed cost includes a shopper's inherent preference for the store and historic store loyalty. The variable cost is a weighted sum of the quantities of items on the shopping list multiplied by their expected prices at the store. The article has three objectives: (1) to model and estimate the relative importance of fixed and variable shopping costs, (2) to investigate customer segmentation in response to shopping costs, and (3) to introduce a new measure (the basket size threshold) that defines competition between stores from a shopping cost perspective. The model controls for two important phenomena: Consumer shopping lists might differ from the collection of goods ultimately bought, and shoppers might develop category-specific store loyalty.

Determining Where to Shop: Fixed and Variable Costs of Shopping

Industry research suggests that location explains up to 70% of the variance in people's supermarket choice decisions (*Progressive Grocer* 1995). At the same time, retailers devote considerable time and effort to setting prices and formulating promotion strategies to increase store traffic. This raises an important question in retail strategy: To what extent can a store's pricing and other marketing activities be used to increase the store's traffic and influence the mix of clientele that shops there? To address this question, we take the perspective of a shopper and decompose the total cost of each shopping trip (for any shopper–store pair) into fixed and variable components. The fixed cost is independent of, whereas the variable cost depends on, the shopping list (or basket). Consequently, unlike the fixed cost, the variable cost varies from trip to trip, because the household has a different shopping list for each trip. The basic premise of this article is that each shopper is more likely to visit the store that imposes the lowest total shopping cost for each trip.

Using the shopping cost framework, we analyze the underlying factors that affect store choice. The empirical analysis uses household-level scanner panel data from a market in which the stores adopt different pricing and positioning strategies. The five stores include two "Every Day Low Price" (EDLP) stores from different chains, a promotional pricing (HILO) store, and two "high-tier" HILO stores from the same chain. We label these stores E1, E2, H1, HH1, and HH2, respectively. The household-level scanner panel data contain demographic information and the residential location of each household. The household information enables us to analyze how certain factors (such as travel distance between each household and each store) affect the fixed cost of shopping. It is reasonable to expect that the fixed costs will vary across stores and shoppers because EDLP stores typically offer lower levels of service than the HILO stores do (e.g., Lal and Rao 1997). In addition, we examine how a store's pricing format affects the variable cost of shopping. By analyzing the underlying factors that affect the total shopping cost, our model provides insights into actions that stores can take to increase patronage or alter their mix of clientele.

Previous empirical research on store choice and sales tends to focus on either the fixed or the variable cost of shopping, but not both. The retail site selection models (e.g.,

*David R. Bell is an assistant professor and Christopher S. Tang is a professor, The Anderson School, UCLA (e-mail: david.bell@ and christopher.tang@anderson.ucla.edu). Teck-Hua Ho is an associate professor, The Wharton School, University of Pennsylvania (e-mail: hoteck@wharton.upenn.edu). The authors thank Doug Honold and Tara Merrill of Information Resources Inc. for providing them with the data. The comments of the editor and four anonymous *JMR* reviewers were extremely helpful, and the authors are grateful for their advice and suggestions. The authors also thank Randy Bucklin, Jeongwen Chiang, and seminar participants at the Hong Kong Institute of Science and Technology, The Wharton School, and Dallas INFORMS for their comments. This research is partially supported by the UCLA committee on research grant 92. All authors contributed equally.

Brown 1989; Craig, Ghosh, and McLafferty 1984; Huff 1964) focus on the fixed cost of shopping and assume that shoppers are influenced primarily by store location and the associated travel costs. For example, in Huff's gravitational model of site selection, the utility of a store is inversely proportional to a nonlinear specification of the distance between the store and the household. Thus, the retail site selection models do not capture the effect of retail pricing format on store choice.

The stream of research on how retail pricing format affects store choice and sales typically focuses on the variable costs of shopping. For example, Mulhern and Leone (1990) examine the effect of a change in price format at one supermarket. They find that moving from an EDLP to a HILO format led to an increase in store sales. Hoch, Dreze, and Purk (1994) show that, across a broad range of product categories, consumer price response is relatively inelastic. One implication of this study is that, in a mature market, an EDLP strategy might not be effective for attracting new customers. In a subsequent study, Hoch and colleagues (1995) find that consumer characteristics explain almost 70% of the variance in category-level price elasticities. This suggests that retailers should tailor store pricing strategies in accordance with the demographic characteristics of their immediate markets. Kahn and Schmittlein (1992) find that the likelihood of purchasing an item on sale or with a store coupon is related to whether the shopping trip is a major or fill-in trip.[1] In particular, purchases made in the presence of sales are more associated with fill-in trips than with major trips. Conversely, purchases when coupons are available are more associated with major trips than with fill-in trips. Finally, Bell and Lattin (1998) demonstrate that there is a systematic relationship between a household's shopping behavior and store preference. In particular, EDLP stores attract more large-basket shoppers; HILO stores attract more small-basket shoppers. The impact of the fixed cost of shopping on the store choice has been neglected in this stream of research.

This article contributes to the current literature in three ways. First, we develop a new econometric store choice model at the individual household level. Our model is the first to partition explicitly the *ex ante* expected total cost of shopping into two components: fixed and variable costs. In capturing the fixed cost, we quantify the travel cost of each household for each store choice alternative and the household's inherent (i.e., category-independent) loyalty toward each store. The variable cost is captured by modeling the household's expected expenditure at a given store. That is, we assess the household's shopping list prior to the store visit and compute what the household is likely to pay at each store. Because the *ex ante* shopping list is unobservable, a model of conditional purchase incidence is used to assess the probability that an *ex post* purchased item was on the *ex ante* shopping list. A household may develop a category-specific store loyalty when it buys an item from the same store repeatedly. This habitual behavior may provide some implicit value to the shopper. Specifically, the shopper's search cost for the item may decrease, and his or her ability to recognize a deal occasion for the item at the store may increase. Consequently, we use category-specific store loyalty to adjust the variable cost.

Second, we use a latent class approach (e.g., Kamakura and Russell 1989) to investigate consumer segmentation in response to the fixed and variable costs of shopping. Although the latent class approach has been used extensively to study brand choice and purchase incidence (e.g., Bucklin and Gupta 1992; Kamakura and Russell 1989), its application to store choice is new. The latent class analysis enables us to determine what types of shoppers (as defined by demographics and shopping behavior) prefer what kinds of stores (as defined by positioning and pricing strategies). It also offers the potential to test some recent models of store pricing and positioning equilibria (e.g., Lal and Rao 1997; Lattin and Ortmeyer 1991) that suggest that consumer heterogeneity sustains the existence of different pricing formats. By identifying whether distinct customer segments exist, our work provides an empirical test of these models.

Third, we use the estimated parameters to determine the "basket size threshold," which can be used to measure the relative competitiveness of a given pair of stores. The basic idea is that there exists a threshold level of basket size beyond which one store is preferable to the other and below which the reverse is true. The threshold is determined by solving for the basket size, such that two stores impose an identical total cost on the consumer. In our subsequent discussion, we use the measure to analyze the competitive vulnerability of a store and explain why some stores are more attractive (to some segments) than others.

Our model assumes that a household has a linear expected disutility over the total shopping cost.[2] We use a logit model setup and estimate our store choice model using a market basket database provided by Information Resources Inc. (IRI). The database comes from a large metropolitan area in the United States and contains two years of data (June 1991 to June 1993). It contains information on the grocery shopping behavior of 520 households at five different supermarkets. We use information from 943 stock keeping units (SKUs) that are common to two or more of the stores in our database. Common SKUs are used to ensure that direct price comparisons are meaningful. These 943 SKUs account for 71% of the total number of purchases in the product categories studied at all stores during the course of two years.

The article is organized as follows: We present our model of shopping costs and the consumer's store selection process; we then describe the database and present the substantive findings. Finally, we discuss the implications of our findings and suggest future research directions.

THE MODEL

We investigate store visit behaviors of H households (indexed by h = 1, ..., H) visiting S stores (indexed by s = 1, ..., S) over a time horizon of D days (indexed by d = 1, ..., D). If a store visit is made on day d, household h must decide which store s to visit. We assume that the household's decision to make a shopping trip is exogenous and that store choice is driven primarily by the total cost of shopping for that particular shopping visit. We seek to understand how the total cost of shopping differs across stores for

[1] Major and fill-in trips were classified on the basis of the amount spent on the trip (see also Kahn and Schmittlein 1989).

[2] In a previous version of this article, we relaxed the linearity assumption and examined two nonlinear specifications. We found, however, the linear model to be quite robust.

different consumers and how this heterogeneity in shopping cost delineates store competition.

We assume that the total cost of shopping for each store visit consists of a fixed and a variable (i.e., shopping list-related) cost. The fixed cost for a household shopping at a store depends on the household's inherent preference for and loyalty toward the store, as well as the distance the household must travel to reach the store. The fixed cost for a given store–household pair is shopping list independent. (Fixed costs clearly vary across stores for the same household because of differences in distance and inherent preference.) Unlike the fixed cost, the variable cost can vary from trip to trip because the household might have a different shopping list for each trip. It is a weighted sum of the quantities of items on the shopping list, multiplied by their expected prices at the store.

In the following sections, we describe the shopping process, the components of the total cost, the store choice decision, and how the basket size threshold is computed.

The Shopping Process

We assume that shoppers adopt the following decision-making process to determine which store to visit:

1. Formulate a shopping list by either writing it down on a piece of paper or constructing and remembering it mentally. The shopping list contains the items to be purchased and their respective quantities.
2. Evaluate the total cost of shopping at each store. As is indicated previously, the total cost has both fixed and variable components. As we note subsequently, this total cost is typically different than the total dollar expenditure incurred in purchasing the products on the shopping list.
3. Select the store that yields the lowest total cost of shopping.

This decision-making process is analogous to those discussed in previous analytical and empirical models of store choice (e.g., Bell and Lattin 1998; Simester 1995). These models posit a shopper who evaluates the total shopping cost of each shopping trip for each store systematically. In real life, shoppers might use a different decision-making process and select stores on the basis of other factors. For example, some shoppers might select a store on the basis of product assortment. Our model, though simplified, represents a way to integrate some prior retail site selection models and research in retail pricing format. It provides a useful benchmark on which other extensions can be built. The evaluation of the fixed and variable costs for a household to shop at a store depends on three major household factors: shopping list, knowledge of store prices, and habitual behavior with respect to store visits.

Shopping list. We assume that each shopping trip made by household h on day d is accompanied by a shopping list that specifies the planned requirements $r_h^d(i)$ for each product i (i = 1, ..., N_h^d), where N_h^d corresponds to the total number of products on the shopping list. The planned requirement $r_h^d(i)$ represents the quantity of product i that household h intends to buy prior to the store visit on day d. Therefore, the shopping list is denoted by $R_h^d = [r_h^d(1), ..., r_h^d(i), ..., r_h^d(N_h^d)]$.

The shopping list acts as short-term memory for the shopper. Prior research (Bettman 1979; Morwitz and Block 1996) suggests that many shoppers plan grocery purchases prior to making a shopping trip. A recent study (Rickard 1995) suggests that more than half of supermarket shoppers write down the planned purchase items on shopping lists prior to visiting a store. The availability of "shopping list" writing pads for sale supports such a notion.

Unfortunately, we are unable to observe consumers' shopping lists directly. Because of factors such as unplanned purchases and stockouts, the list of items actually bought at the store is likely to be different from the shopping list developed prior to the store visit. We distinguish between the *shopping list* (items for which purchase is planned prior to the store visit) and the *purchased list* (items actually bought) as follows: We denote the purchased list by $Q_h^d = [q_h^d(1), ..., q_h^d(i), ..., q_h^d(M_h^d)]$, where M_h^d corresponds to the total number of purchased products, and $q_h^d(i)$ is the purchased quantity of item i. As we show subsequently, we derive the (unobserved) shopping list, R_h^d, from the purchased list, Q_h^d, using a conditional purchase incidence model.

Price knowledge. We assume that, prior to the store visit, consumers do not know the actual prices in each store for each product on their shopping list. Rather, they have knowledge of the price distribution (e.g., Assunçao and Meyer 1993; Dickson and Sawyer 1990; Ho, Tang, and Bell 1996; Lal and Rao 1997).[3] A considerable body of work in marketing (e.g., Kalwani et al. 1990; Kalyanaram and Winer 1995; Lattin and Bucklin 1989; Winer 1986) supports the notion that consumers develop knowledge for product prices and that this knowledge influences choice behavior. In addition, consumer use of price knowledge is central to analytical models of store choice (e.g., Lal and Matutes 1994; Lal and Rao 1997; Simester 1995). In this article, we also rely on the assumption that consumers develop some prior knowledge about the pricing environment in different stores. Specifically, we assume that each household h knows the price distribution of each product i at each store s. This knowledge has been acquired through previous visits to the store and exposure to advertising activity on the part of the store. Let $Ap_d^s(i)$ be the actual price for product i at store s on day d. Let $\mu_s(i)$ be the mean price, which we assume is known by all households.[4] The behavioral implication of our modeling assumption is that consumers have some sense of relative mean price levels in different stores. This seems reasonable when we consider that most shoppers acquire such holistic knowledge through exposure to television and newspaper advertising and that the shopping experience is a repetitive and frequent activity. Our assumption is also consistent with recent work by Alba and colleagues (1994), who

[3]This assumption may appear strong. Shoppers implicitly are assumed to have high search costs and not to look for deals and promotions during each shopping trip. Search cost clearly varies across the population. However, our model is still a good approximation for customers who look for deals in basket shopping. Because the customers shop for a basket of products, it is unlikely that a store will be promoting all the products that are on the consumers' shopping list. Thus, as long as the basket size is not small, we should expect to find some products that are on sale in one store, whereas others are on sale in a competing store. Therefore, the benefit of search diminishes, and the deal effects somewhat cancel out, in basket shopping.

[4]A less restrictive (but less parsimonious) assumption is that there is some heterogeneity in price knowledge; that is, $\mu_s(i)$ differs across households. A reasonable way to allow for this heterogeneity is to have the price knowledge of each shopper depend only on prices derived from individual past purchases rather than on store weekly prices (e.g., sample mean and variance). Thus, our model implicitly assumes that customers sample price information frequently. (Note that our households are screened to make sure they take at least one trip per month.)

performed experimental manipulation checks on basket price perceptions. They show that consumers retained strong impressions about relative price levels across supermarkets and that these impressions were consistent with actual price levels at the stores.

Habitual store visit behavior. A household can develop habitual store visit behavior that generates two kinds of store loyalty: category-independent and category-specific. A household's category-independent store loyalty captures its habitual preference for a store, independent of the shopping list. (The category-independent loyalty is analogous to brand loyalty as discussed in brand choice literature. Similar to the brand choice measure, category-independent store loyalty can be specified by using the actual store choice decisions in the initial months of data.) This category-independent loyalty tends to lower the fixed cost of shopping at the store.

Conversely, a household's category-specific store loyalty (e.g., buys Pampers disposable diapers from Wal-Mart and Crystal Geyser mineral water from Trader Joe's) depends on the shopping list and thus can vary from trip to trip. The category-specific loyalty tends to reduce the variable cost of shopping, because it reduces the search cost and increases the shopper's ability to recognize deal occasions for the item at the store. In the next section, we describe how both kinds of store loyalty reduce the total cost of shopping.

Total Cost of Shopping

The total cost associated with a specific store visit consists of two components: a fixed and a variable, shopping list-related cost. Using the price knowledge for product i in store s as a basis, household h is able to compute the full cost of purchasing the items on the shopping list R_h^d on day d at store s. Specifically, the total cost, $TC_h^d(s)$, for household h to shop at store s on day d is given by

$$(1) \qquad TC_h^d(s) = F_h(s) + V_h^d(s),$$

where $F_h(s)$ and $V_h^d(s)$ are the fixed and variable costs of shopping on day d by household h at store s.

The fixed cost $F_h(s)$ is partitioned further into factors that reflect the inherent cost associated with visiting a store s (e.g., service, assortment), category-independent store loyalty, and travel cost. Specifically,

$$(2) \qquad F_h(s) = \alpha_s + \theta \times SL_h(s) + \phi \times D_h(s)^{\omega}.$$

Note that the inherent cost, α_s, is the same for all shoppers. Household h's category-independent loyalty toward store s, denoted by $SL_h(s)$, varies from shopper to shopper and is initialized using the first three months of data. The category-independent store loyalty sensitivity parameter is θ. If θ is negative, category-independent store loyalty lowers the fixed cost of shopping. The geometrical distance between household h and store s is $D_h(s)$, and ϕ and ω measure the consumer's distance sensitivity, or willingness to travel. As in the gravitational retail site selection models, the effect of distance enters nonlinearly. In the latent class analysis, we allow for the possibility that members of different market segments have different parameter values (i.e., we estimate $\alpha_s|_g$, $\theta|_g$, $\phi|_g$, and $\omega|_g$, for segments g = 1, ..., G).

The expected variable cost, $V_h^d(s)$, is the sum of the products of the planned purchase quantities of items on the shopping list on the day of shopping and their expected prices at the store:

$$(3) \qquad V_h^d(s) = \sum_{i=1}^{N_h^d} r_h^d(i) \times \mu_s(i).$$

Recall that we do not observe the shopping list, R_h^d. Instead, the list of purchased items, Q_h^d, is recorded. Consequently, we must develop a model that relates the purchased list Q_h^d to the *ex ante* shopping list, R_h^d.

There are several reasons Q_h^d might be different than R_h^d. First, shoppers often make additional and unplanned purchases when they are in a store. These purchases can account for as much as one-half to two-thirds of all purchases made by consumers in supermarkets (Bowman 1987; Bucklin and Lattin 1991; Park, Iyer, and Smith 1989). Second, shoppers might forget to buy some products on R_h^d or omit others because of high prices or stockouts. This will cause some products that are on the shopping list to disappear from the purchased list. Third, shoppers might alter the planned purchase quantity for some product i on the shopping list, so that $q_h^d(i) \neq r_h^d(i)$, because of the store's promotional activities. A comprehensive theory of shopping lists should capture all three factors.

Our data set will not allow us to model the second factor. In light of the evidence that the list of purchased items is often greater than the shopping list, we expect the first factor to dominate the second. Also, the effect of the third factor (quantity adjustment) on store choice appears to be modest.[5] Thus, we model only the major factor—the effect of unplanned purchases. Given that unplanned purchases are prevalent, we are likely to inflate the variable cost if we simply assume that the shopping list is the same as the purchased list.[6]

Let $\pi_h^d(i)$ be the probability that an *ex post* purchased item i is on the *ex ante* shopping list. We want $\pi_h^d(i)$ to be equal to 1 for item i if it is on the shopping list and 0 otherwise. We posit the following conditional purchase incidence model:

$$(4) \quad \pi_h^d(i) = \frac{\exp\left\{\delta^I \times I_h^d(i) + \delta_P \times \left[ap_s(i) - \mu_s(i)\right] + \delta_C \times C_h(i)\right\}}{1 + \exp\left\{\delta^I \times I_h^d(i) + \delta_P \times \left[ap_s(i) - \mu_s(i)\right] + \delta_C \times C_h(i)\right\}},$$

where $I_h^d(i)$ is household h's inventory for product i, and $C_h(i)$ is the household's consumption rate for product i. Because item i is more likely to be on the shopping list if its inventory is low, it is not on sale, and the consumption rate is high, we expect $\delta_I < 0$, $\delta_P > 0$, and $\delta_C > 0$. (In the estimation, we scale $\pi_h^d(i)$ so that when the inventory is the lowest and the price differential and consumption rate are the highest, we have the item on the *ex ante* shopping list with probability = 1.)

Our rationale for this formulation follows Bucklin and Lattin (1991), who show that consumers are much more re-

[5]In a previous version of this article, we considered a model in which shoppers modify their planned purchase quantity in response to discrepancies between actual and expected prices. The store choice model fit improved slightly, and it also appeared that quantity adjustment occurs in only a few categories.

[6]Note, however, that our market basket database only covers a subset of the items carried by a supermarket. Given this, our empirical models actually may underestimate the variable cost, depending on the extent of unplanned shopping and the actual size of the basket.

sponsive to in-store prices when they are in an unplanned purchase mode. Therefore, we expect instances of unplanned shopping will be indicated by purchases that take place when the actual price [$ap_s^d(i)$] is lower than the expected price [$\mu_s(i)$]. We should see $\delta_P > 0$, so that the probability of the item being on the shopping list decreases with the consumer's tendency to engage in unplanned shopping. The formulation also suggests that planned shopping is likely to be driven by need, so that conditions of low inventory and high consumption rate imply a higher likelihood that a purchased item is on the shopping list. Thus, the variable cost is as follows:

$$(5)\qquad V_h^d(s) = \sum_{i=1}^{M_h^d} \pi_h^d(i) \times q_h^d(i) \times \mu_s(i).$$

Equation 5 has the simple interpretation that, in choosing a store, consumers consider both the likelihood that a certain amount of product will be bought and the price they expect to pay. In addition to this "shopping list" effect on the perception of variable costs, we also conjecture a category-specific store loyalty effect. Category-specific store loyalty arises when, for a given shopper, the proportion of category purchases in a store exceeds the proportion of store visits received by the store. To capture this effect, we use the shopper's loyalty toward buying the category containing product i from store s to lower store s's unit price, as follows:

$$(6)\ V_h^d(s) = \sum_{i=1}^{M_h^d}\left[\pi_h^d(i) \times q_h^d(i)\right] \times \left[\frac{e^{\psi \times CL_h(i,s)}}{1 + e^{\psi \times CL_h(i,s)}} \times \mu_s(i)\right],$$

where $CL_h(i, s)$ is the category-specific loyalty of household h buying in the category containing product i from store s. If $\psi < 0$, the term [$e^{\psi \times CL_h(i,s)}/1 + e^{\psi \times CL_h(i,s)}$] decreases as $CL_h(i, s)$ increases. Therefore, the perceived price of product i at store s decreases as category-specific loyalty toward the store increases.[7] The rationale is that category-specific store loyalty reduces price implicitly because it reduces the time and cost required for the shopper to search for the product in the store. It also increases the ability of the shopper to recognize deal occasions for the product. In summary, the total cost of shopping is given by

$$(7)\qquad TC_h^d(s) = \alpha_s + \theta \times SL_h(s) + \phi \times D_h(s)^{\omega} + \sum_{i=1}^{M_h^d}\left[\pi_h^d(i) \times q_h^d(i)\right] \times \left[\frac{e^{\psi \times CL_h(i,s)}}{1 + e^{\psi \times CL_h(i,s)}} \times \mu_s(i)\right].$$

Store Selection

To estimate consumer response parameters, we relate the shopping cost to consumer utility for store s by assuming a linear utility function. Let β be the response parameter that captures the consumer's sensitivity toward relative costs across stores. The utility of shopping at store s by household h on day d is[8]

$$(8)\qquad U_h^d(s) = -\beta \times TC_h^d(s) + \epsilon_h^d(s).$$

We assume that $\epsilon_h^d(s)$ in Equation 8 are independent and identically distributed double exponential random errors.[9] We can use Equations 1, 2, and 6 to obtain the expression for the deterministic utility for household h, which shops at store s on day d, as follows:

$$(9)\quad E\left[U_h^d(s)\right] = -\alpha_s' - \theta' \times SL_h(s) - \phi' \times D_h(s)^{\omega} - \beta \times \sum_{i=1}^{N_h^d}\left[\pi_h^d(s) \times r_h^d(i)\right]\left[\frac{e^{\psi \times CL_h(i,s)}}{1 + e^{\psi \times CL_h(i,s)}} \times \mu_s(i)\right],$$

where $\alpha_s' = \beta \times \alpha_s$, $\theta' = \beta \times \theta$, and $\phi' = \beta \times \phi$.

On the basis of the deterministic utility $E[U_h^d(s)]$, household h will shop at store S_h^d on day d, where S_h^d is a random variable that has a probability distribution given by

$$(10)\ \text{Prob}\left\{S_h^d = s \,\middle|\, E\left[U_h^d(s)\right], s = 1, ..., S\right\} = \frac{e^{E\left[U_h^d(s)\right]}}{\sum_{q=1}^{S} e^{E\left[U_h^d(q)\right]}}.$$

Equation 10 completes the specification of the cost of shopping and the store choice process.

Basket Size Threshold

Equations 1 and 6 suggest that the expected total cost of shopping on day d for household h with a shopping list R_h^d at store s is given by

$$(11)\qquad TC_h^d(s) = F_h(s) + \sum_{i=1}^{M_h^d}\left[\pi_h^d(i) \times q_h^d\right] \times \left[\frac{e^{\psi \times CL_h(i,s)}}{1 + e^{\psi \times CL_h(i,s)}} \times \mu_s(i)\right] = F_h(s) + \sum_{i=1}^{M_h^d} r_h^d(i) \times m_{h,s}(i),$$

where $r_h^d(i) = \pi_h^d(i) \times q_h^d(i)$, and $m_{h,s}(i) = [e^{\psi \times CL_h(i,s)}/1 + e^{\psi \times CL_h(i,s)}] \times \mu_s(i)$.

To find how the basket size threshold influences store choice and can be used to understand competition, we consider two stores, s_1 and s_2. Clearly, if a store has both lower fixed and variable costs, the shopper will prefer that store. We consider the interesting case in which one store, say s_1, has a higher fixed cost and a lower variable cost; that is, $F_h(s_1) > F_h(s_2)$, and $E[V_h^d(s_1)] < E[V_h^d(s_2)]$. Figure 1 depicts the relationship between the basket size threshold and the expected total cost, incurred at stores s_1 and s_2, respectively. (See the Appendix for the calculation of the threshold.)

Figure 1 can be interpreted as follows: First, the figure portrays the case in which store s_2 has a lower fixed cost and higher variable cost than store s_1. The relative expected total cost levels imply that there is a critical basket size level, $r^*_{h,12}(y)$ [$r^*_{h,12}(y) > 0$], such that store s_1 will be preferred

[7]In the estimation, we scale the weight so that the household's perceived price is the same as the expected price if category-specific loyalty is zero.

[8]Adding a constant term to Equation 8 is unnecessary because there is already a constant term α_s in the total cost equation. In addition, because α_s is specified as a component of the total cost term, a more negative α_s means the customer has a higher utility (or lower cost) for store s.

[9]Although the error terms for different stores are likely to be correlated, the inclusion of store loyalty in the model will cause the IIA assumption to hold at the level of the individual. We thank an anonymous reviewer for this insight. See also Currim (1982).

to store s_2 if the shopper's basket size $r_{h,12}(y)$ is beyond $r^*_{h,12}(y)$. (If store s_1 were to have lower fixed and lower variable costs, then $r^*_{h,12}(y) < 0$ and store s_1 would be strictly preferred to s_2.) Second, the slope of the expected total cost curve represents the unit price of the standard basket at each store. In Figure 1, store s_2 has a lower fixed cost (perhaps because of better service, convenience, and so forth), but it has a higher unit price for the standard basket. This is the sort of relationship we expect to find when store s_2 practices HILO and store s_1 offers EDLP.

The basket size threshold can be computed at the market segment level as well. We take the average of the segment members' threshold level to obtain the segment level threshold. In a subsequent section, we compute $r^*_{sq|g}(y)$ for all store pairs, s and q, $s \neq q$, and for each segment, $g = 1, \ldots, G$, and discuss how these thresholds can provide important insights into store competition.

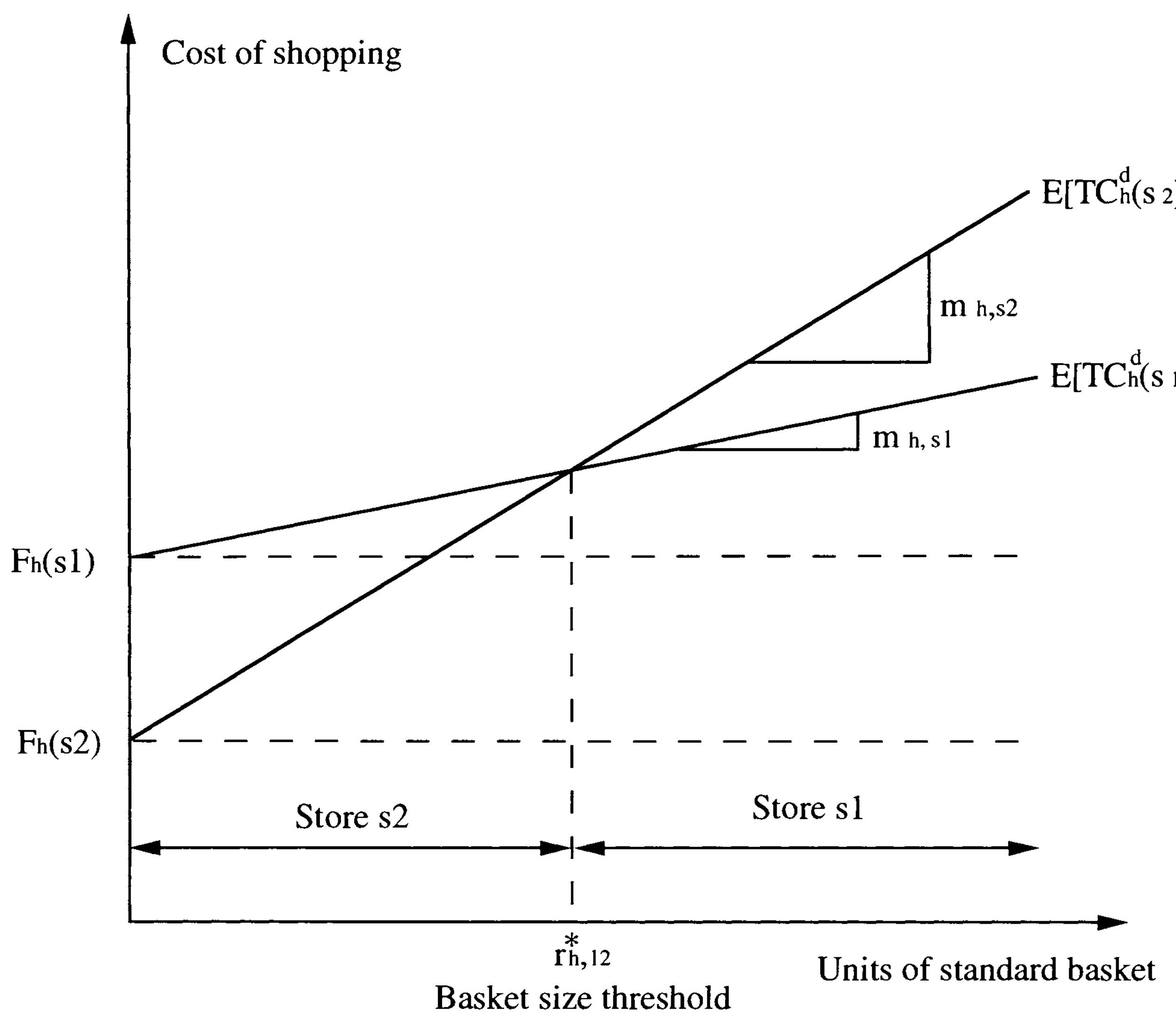
DATA SELECTION

To calibrate our model, we obtained an IRI database for two years of data (June 1991–June 1993) that contain shopping basket purchase histories for 520 households at five supermarkets. These data include

- •merchandizing information for each SKU at each store (weekly pricing and promotion information),
- •purchase histories for multiple product categories for each household, and
- •demographic information (e.g., family size, household income) and five-digit zip codes for all panelists.

Stores. To conceal the identity of the stores in the data set, we label them as E1, E2, H1, HH1, and HH2. E1 and E2 are from different chains and explicitly advertise as EDLP stores. H1 is a HILO store from a third chain; HH1 and HH2 are a higher tier of HILO store and are from the same chain.

Figure 1
THE BASKET SIZE THRESHOLD



To ensure that the revealed pricing strategies of the stores were consistent with the price positioning, we computed the unit price of a 12-category basket for each of the supermarkets (Bell 1995). Table 1 shows that the mean and the variance of the basket prices are consistent with the price positioning of the stores (i.e., EDLP stores have lower average prices and lower variance, though H1 appears to have less variability than E1).

Product categories. A total of 24 product categories were potentially available for use in the analysis (for a description, see Bell 1995). From this set, we choose a smaller subset for model calibration and analysis. We begin by securing a broad representation (e.g., household products, food and nonfood) and apply additional criteria to ensure that selected categories are bought frequently, contain SKUs that are common to at least two of the five stores (to enable meaningful price-based comparisons), and are responsive to price variation. To do this, we refer to a recent paper by Ho, Tang, and Bell (1997) that used the same market basket database. The combination of these three criteria led us to select the 12 product categories listed in Table 2. Table 2 reports the number of purchases made by the 520 households during three consecutive time periods: "initialization" (first 3 months), "calibration" (next 18 months), and "validation" (final 3 months). The analysis uses 943 SKUs; the 68,808 calibration purchases made by the 520 households on 30,012 shopping trips account for more than 71% of all purchases made for all SKUs and all 12 categories. The focus on only 943 frequently bought SKUs might undermine the importance of variable cost, because the union of all shopping baskets is likely to contain more SKUs. Thus, our results should be taken as a conservative estimate of the relative importance of the variable (shopping list-related) cost. We also determine that the chosen SKUs account for 16% of the total expenditures made by the panelists. In our model, each SKU is treated as a separate product. To ensure that prices of products in the same category are directly comparable, we scale each price according to the standard unit for that category.[10]

Households. In Equation 9, there are three household-level independent variables: $D_h(s)$, $SL_h(s)$, and $CL_h(i, s)$. We use information on household and store locations to represent distance between each household and each store. Distance traveled and willingness to travel are critical determinants of the fixed cost of shopping (see Equation 2). To compute the travel distance, $D_h(s)$, between each household

[10]*Standard units* are defined by the data supplier, IRI. In the bacon category, the 16-ounce size is considered a standard unit of product. To make the prices of all bacon SKUs comparable, we compute per unit prices for each bacon SKU, with 16-ounce as the appropriate unit. A complete description of our procedure is available on request.

Table 1

STORE PRICING[a] AND ADVERTISING

		Trips		*Pricing*		*Advertising*
Price Tier	*Store*	*Total*	*%*	$\overline{X}$	$\hat{\sigma}$	*Explicit EDLP?*
Lowest	E1	6562	21.5	24.86	1.35	YES
	E2	8184	26.8	24.94	1.21	YES
Middle	H1	9944	32.6	27.59	1.27	NO
Highest[b]	HH1	2764	9.0	30.13	1.76	NO
	HH2	3096	10.1	30.95	1.80	NO

[a]Based on basket prices for highest market share SKUs.
[b]Stores HH1 and HH2 are from the same chain.

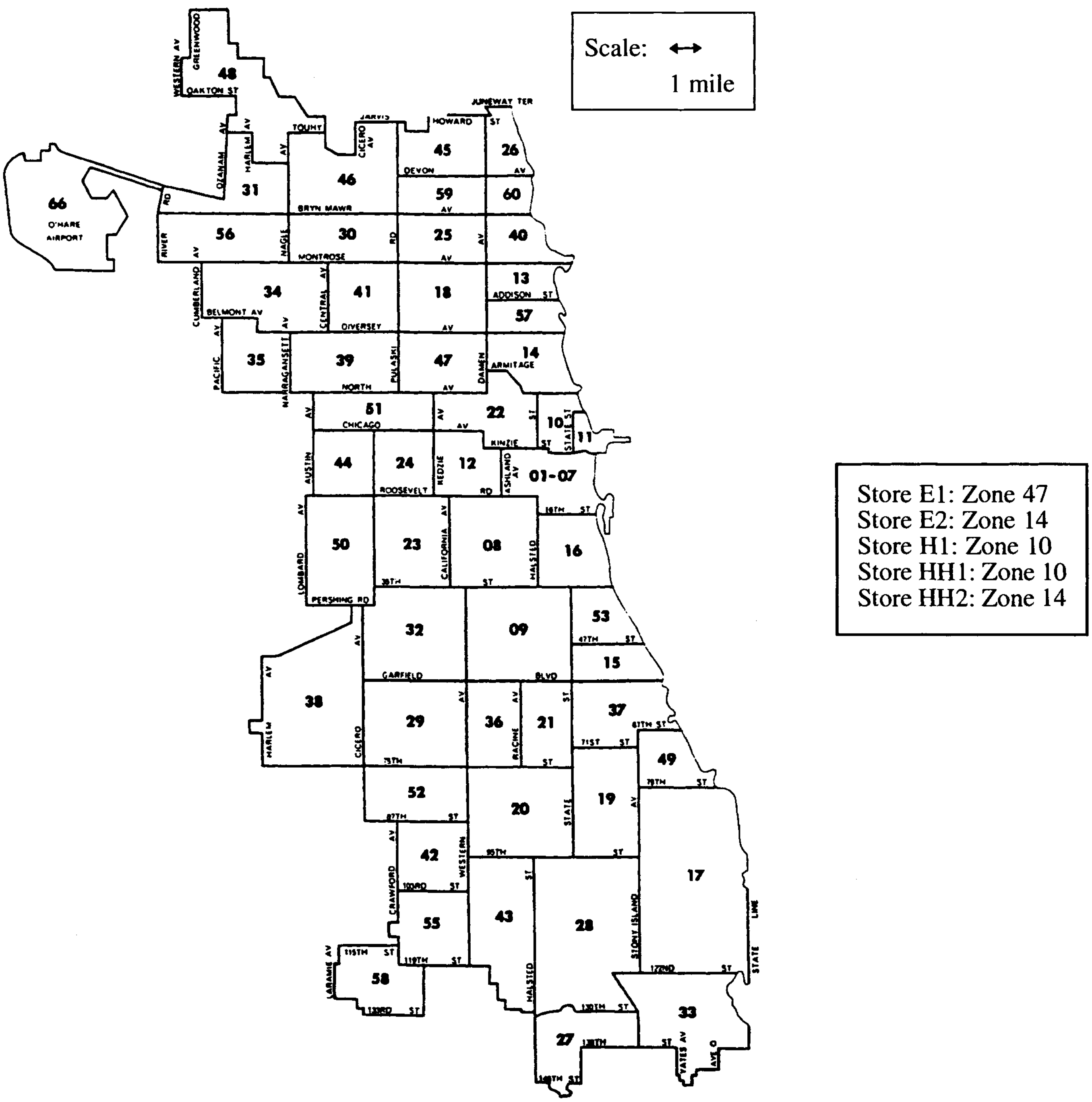
Table 2

PRODUCT CATEGORIES USED IN ANALYSIS

		Number of Purchases		
Category	*Number of SKUs*	*Initialization*	*Calibration*	*Validation*
Canned catfood	151	1873	10,292	1600
Paper towels	64	1396	8557	1301
Flavored soda	109	1275	7918	1173
Bathroom tissue	24	1205	7317	1129
Regular cereal	52	1014	6058	859
Cola/cola-flavored soda	26	1036	5982	935
Yogurt	70	851	5657	928
Margarine	61	673	5035	698
Potato chips	102	647	3827	543
Ice cream	93	595	3121	489
Frozen pizza	130	313	2671	388
Liquid detergents	61	379	2373	330
Totals	943	11,257	68,808	10,373

h and each store s, we proceed as follows: We first obtain five-digit zip code information for each of the households and call each of the supermarkets to determine their exact location. Figure 2 is a map of the geographic area in which the stores and the households are located. (Note that stores E2 and HH2 share the same zip code, as do H1 and HH1.) Because we have no additional information about the locations of the households, we assume that households in the same zip code are distributed uniformly over the area of the zip code. Each zip code is approximated by a rectangle and defined by the coordinates of the corners of the rectangle. To compute the expected distance between each household and each store, we take the integral of the distance between respective coordinates of the rectangle (or household) and

Figure 2
LOCATION OF STORES



the store. The variable $D_h(s)$ corresponds to this estimated expected distance.[11]

Category-independent and category-specific store loyalty [$SL_h(s)$ and $CL_h(i, s)$] are initialized using actual store visits in the first three months. They are computed in the same way that brand loyalty often is computed (e.g., Bucklin and Lattin 1992). Specifically, $SL_h(S) = [n_h(s) + (1/S)/(N_h + 1)]$ and $CL_h(i,s) = [n_h(i,s) + (1/S)/(N_h(i) + 1)]$, where $n_h(s)$, S, N_h, $n_h(i, s)$, and $N_h(i)$ are the number of trips made by household h to store s, the total number of stores, the number of trips made by household h, the number of purchase instances for household h in product category i in store s, and the total number of purchases made by household h in category i, respectively.

ESTIMATION AND RESULTS

Using the expected utility $E[U_h^d(s)]$ given in Equation 9 and the store choice probability given in Equation 10 as a basis, we now describe the process for estimating the model parameters and report our results. Nonlinear parameters (i.e., ω) are estimated using Fader, Lattin, and Little's (1992) iterative Taylor Series procedure.

Estimation Procedure and Strategy

Procedure. We estimate the model parameters using maximum likelihood. For each household h on a given day d, we define an indicator variable T_h^d that denotes whether or not a store visit occurs:

$$T_h^d = \begin{cases} 1, & \text{if household h shops on day d,} \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, in the single-segment model the likelihood function is simply as follows:

$$(12)\quad L = \Pi_{h=1}^{H}\Pi_{\{d:T_h^d = 1\}}\text{Prob}\left(S_h^d = s\right)$$

$$= \Pi_{h=1}^{H}\Pi_{\{d:T_h^d = 1\}}\frac{e^{E\left[U_h^d(s)\right]}}{\sum_{q=1}^{S} e^{E\left[U_h^d(q)\right]}}.$$

To allow for heterogeneity in the fixed and variable cost parameters, we subsequently relax the assumption of a single segment and estimate parameters conditional on segment membership. That is, we estimate segment sizes, τ_g, for segments g = 1, ..., G. In this case, the likelihood function can be written as follows:

$$(13)\quad L = \Pi_{h=1}^{H}\left\{\sum_{g=1}^{G}\tau_g\Pi_{\{d:T_h^d = 1\}}\frac{e^{E\left[U_h^d(s|g)\right]}}{\sum_{q=1}^{S} e^{E\left[U_h^d(q|g)\right]}}\right\},$$

[11]The construction of the appropriate distance measures is quite tedious and requires the computation of more than 300 integrals in closed form. The details are available from the authors on request. In a previous draft of the article, we assumed that each store and panelist was located at the centroid of its respective zip code. This revised approach, though more involved, leads to a larger magnitude coefficient and a higher level of significance for distance. We thank an anonymous reviewer for suggesting the idea.

Table 3

ESTIMATION STRATEGIES FOR MODELS

	Step 1	*Step 2*
PFC Model		
Model 1.1	Store-fixed costs (α_s)	Latent class
FFC Models		
Model 1.2	1.1 plus distance (α_s, φ)	
Model 1.3	1.1 plus category-independent loyalty (α_s, θ)	
Model 1.4	1.1 plus distance and store loyalty (α_s, φ, θ)	
Model 1.5	1.4 plus nonlinear distance effect (α_s, φ, θ, ω)	
TC Models		
Model 1.6	1.5 plus basket costs (α_s, φ, θ, ω, β)	
Model 1.7	1.6 plus list probability (α_s, φ, θ, ω, β, δ_I, δ_P, δ_C)	
Model 1.8	1.7 plus category-specific loyalty (α_s, φ, θ, ω, β, δ_I, δ_P, δ_C, ψ)	

where the τ_gs are not estimated directly but rather as logit functions (e.g., Kamakura and Russell 1989).

Estimation strategy. Our estimation scheme consists of two steps. In step 1, we estimate three classes of nested, single-segment models. The first class of models estimates only the inherent costs associated with individual stores (i.e., the α_ss). We refer to this model as a Partial Fixed Cost (PFC) model. This simple model (in which the parameter estimates do nothing more than represent store shares of visits) is our first benchmark. The second class of models captures the incremental effects of distance, category-independent store loyalty, and nonlinear response to distance. These are the Full Fixed Cost (FFC) models. The third class of models contains the Total Cost (TC) models, of which there are two versions: one captures household sensitivity to variable costs, and the other adds to this the effect of category-specific store loyalty. The model formulations and estimation strategy for the full series of nested models are summarized in Table 3.

In step 2, each model then is reestimated according to the latent class formulation of Equation 13. Specifically, we reestimate each single-segment model, continuing until we obtain the model with the best Bayesian Information Criterion (BIC)-adjusted in-sample fit and the best out-of-sample validation log-likelihood. We confirm that the best single-segment model is also the preferred latent class specification (with four distinct segments). In the multisegment analysis, the household-level posterior probabilities of segment membership are computed and related to observable demographics (e.g., family size, income) and shopping behavior variables (e.g., average number of shopping trips). The posterior probability that household h is a member of segment g is given by

$$(14)\quad \text{post}_h^g = \frac{L\left(X_h \mid g\right)\tau_g}{\sum_{j=1}^{G} L\left(X_h \mid j\right)\tau_j},$$

where $L(X_h \mid g)$ is the likelihood function value for household h, given membership in segment g, and τ_g is the prior probability of membership in segment g.

Empirical Results

Shopping costs and store choice. Table 4 lists the log-likelihood, BIC-adjusted log-likelihood, a goodness-of-fit index, $\bar{\rho}^2$ (Ben-Akiva and Lerman 1985; Hardie, Johnson,

Table 4

MODEL FITS FOR SINGLE-SEGMENT LINEAR UTILITY STORE CHOICE MODELS

	Parameters	*Log-Likelihood*	*BIC*	$\bar{\rho}^2$	*Validation*	*Validation Hits*
PFC Model						
Model 1.1 (Store Intercepts)	4	−45,309.0[a]	−45,782.9[b]	.0619[c]	−7,264.0[d]	.313
FFC Models						
Model 1.2 (1.1 plus ϕ)	5	−38,514.0	−38,539.8	.2025	−6,088.1	.543
Model 1.3 (1.1 plus θ)	5	−15,794.1	−15,819.9	.6729	−2,662.8	.819
Model 1.4 (1.1 plus ϕ, θ)	6	−15,504.6	−15,535.5	.6789	−2,609.8	.819
Model 1.5 (1.1 plus ϕ, θ, ω)	7	−15,444.8	−15,480.9	.6801	−2,597.6	.823
TC Models						
Model 1.6 (1.5 plus β)	8	−15,042.2	−15,083.4	.6884	−2,547.6	.828
Model 1.7 (1.5 plus β, δ_I, δ_P, δ_C)	11	−14,986.4	−15,043.1	.6895	−2,528.4	.826
Model 1.8 (1.5 plus β, δ_I, δ_P, δ_C, ψ)	12	−14,927.2	−14,989.0	.6907	−2,522.1	.828

[a]Number of calibration observations is 68,808 purchases, 30,012 trips.
[b]BIC = LL − (k/2) × ln(n), where k and n are number of parameters and observations, respectively.
[c]$\bar{\rho}^2 = 1 - (LL - k)/LL_0$, where k is number of parameters in this model, LL_0 = −48,302.5.
[d]Number of validation observations is 10,373 purchases, 4747 trips.

and Fader 1993), the validation log-likelihood, and the validation hit ratio for all single-segment models.

Observe from Table 4 that Model 1.2, which includes the household-level measure of distance $D_h(s)$, fits substantially better than the benchmark PFC model (Model 1.1). From Model 1.3, it is also clear that category-independent store loyalty accounts for a substantial amount of variance. Moving from Model 1.4 to 1.5, we find that the inclusion of the power of the distance, ω, improves the fit. This suggests that the fixed cost is nonlinear in distance, which is consistent with gravitational models of retail site selection (Huff 1964). Although the fixed cost component is critical for explaining store choice, it is evident from Models 1.6 through 1.8 that the variable cost (through the inclusion of the parameters β, δ_I, δ_P, δ_C, and ψ) also plays a statistically significant role. Thus, both the fixed and variable costs are necessary to provide a comprehensive theory of store choice. As in the best fitting model (Model 1.8), there is a residual effect of category-specific store loyalty, over and above the effect of (category-independent) store loyalty itself. This effect highlights the importance of category-based competition among retailers, alluded to by some single-category studies (e.g., Bucklin and Lattin 1992).

Table 5 presents the estimated response parameters and associated t-ratios for the best fitting single-segment model (TC Model 1.8). Note that the estimates of the inherent costs (α_s) are close to zero and negative for the highest-tier stores (HH1 and HH2). This is consistent with Lal and Rao's (1997) study, in that HILO stores are perceived as having higher levels of service (i.e., they impose lower inherent costs on shoppers). Next, notice that the estimates of ϕ and θ are signed correctly and significantly different from zero. Stores that are farther away impose higher costs; stores that customers have some historical tendency to visit impose lower costs. The loyalty variable has a nice behavioral interpretation in the context of our shopping cost framework. Habitual visits to one store facilitate the development of familiarity with the store's characteristics (service, parking, location of products, and so forth) and thus implicitly reduce the fixed cost of subsequent visits to that same store.

Table 5

PARAMETER ESTIMATES FOR SINGLE-SEGMENT LINEAR UTILITY TOTAL COST MODEL

Variable	*Parameter*	*t-ratio*
Fixed Costs		
α_1 (Store E1)	.000[a]	—
α_2 (Store E2)	.129	4.302
α_3 (Store H1)	.090	2.320
α_4 (Store HH1)	−.134	−2.824
α_5 (Store HH2)	−.163	−3.850
ϕ (Distance)	1.705	10.783
θ (Loyalty)	−4.162	−108.116
ω (Power)	.409	−4.863[b]
Variable Costs		
β (Basket Cost)	.798	30.070
δ_I (Inventory)	−.032	−.074
δ_P [$ap^d_s(i) - \mu_s(i)$]	.374	10.321
δ_C (Consumption rate)	2.442	6.017
ψ (Loyalty weight)	−.144	−10.186
⇒ Minimum weight[c]	.928	

[a]Normalized for identification. When α_s are more negative (cost is lower), utility is higher.
[b]t-test is for H_0: $\omega = 1$ (distance effect is linear).
[c]$2 \times \{\exp[\psi \times CL_h(i,s)]/\{1 + \exp[\psi \times CL_h(i,s)]\}\}$, with $CL_h(i, s) = 1.0$.

The impact of variable cost is significant, albeit small, compared with that of the fixed cost. To preserve the model structure, we report β as our estimated β divided by a scaling factor that makes both the maximum incidence probability $[\pi^d_h(i)]$ and category-specific loyalty weight equal to 1. Notice that β is highly significant, which confirms that variable cost plays a role in store choice. Thus, store location is not the only factor that explains store choice; pricing format matters.

The incidence probability parameters (δ_I, δ_P, δ_C) have the correct signs, and the price discount and consumption rate response parameters (δ_P, δ_C) are significantly different from zero. These results suggest that an *ex post* purchased item is more likely to be on the *ex ante* shopping list if it is bought at regular prices and consumed heavily.

The parameter ψ requires some explanation. Category-specific store loyalty, $CL_h(i, s)$, is bound between zero and one, and an increase in category-specific store loyalty should lead to a reduction in the *ex ante* variable cost of shopping because of an implicit reduction in the price evaluation of that store. Note that ψ is negative and significantly different from zero, which confirms the idea of implicit cost reduction associated with category-specific loyalty. The value of $\psi = -.144$ implies that a consumer who is completely loyal to a store for a particular category would only assess 93% of the price of that category against that store when evaluating which store to visit.[12]

Segmentation in response to shopping costs. Table 6 presents the model fits for the latent class analysis. (We also estimated Gupta and Chintagunta's [1994] model, in which demographic variables are included directly into the prior probabilities. The four-segment model with the most significant demographic predictors [family size and income] yielded a BIC of –12,070.5, compared with –12,047.1 for Model 4.8, so we did not pursue this formulation further.) Total Cost Model 4.8 is the best formulation. It fits significantly better than the corresponding three-segment solution and has a better BIC-adjusted fit than the five-segment solution, whose value is –12,088.1.

Table 7 shows the parameter estimates from the four-segment TC model, Model 4.8. The multiple-segment model enables us to examine the underlying reasons that different segments prefer different stores. The interpretation of the parameters is similar to that of the single-segment case.[13] In Table 7, the range of the parameters $\alpha_{s|g}$ for segments 1 and 3 is relatively small (–.107 to .480 for segment 1 and –2.032 to .000 for segment 3). This observation implies that, collectively, households in segments 1 and 3 do not have strong inherent store preferences. Conversely, the range of the parameters $\alpha_{s|g}$ for segments 2 and 4 is quite large (.000 to 6.414 for segment 2 and –5.717 to .000 for segment 4). Only E1 and E2 will appeal to segment 2; H1, HH1, and HH2 will appeal to segment 4. However, the values of θ suggest that segments 1 and 3 are influenced more strongly by category-independent habitual behavior than segments 2 and 4 are.

The sensitivity to distance (ϕ) varies considerably across segments. Segment 4, the HILO segment, is the most averse to distance, whereas segment 2, the EDLP segment, is the most willing to travel. The range for sensitivity to variable cost (β) is relatively small (1.254 to 2.122); however, the EDLP segment is the most sensitive. Thus, the relative roles of fixed and variable cost vary considerably across segments.

The parameters for the list probability (δ_I, δ_P, δ_C) are signed correctly, with the exception of δ_{I3} (δ_{I3} = .035, t-ratio = .480). The parameter estimates and data imply that the average probability that a purchased item was on the *ex ante* list is highest for segment 1 (.69) and lowest for seg-

[12]When $CL_h(i, s) = 0$, Equation 9 implies that the corresponding weight would be equal to a half. To facilitate exposition, we multiply the implied weight by 2, so that $CL_h(i, s) = 0$ will lead to an implicit weight of 1. In other words, if consumers have no category-specific store loyalty, they evaluate the price of product i at the store by imputing the full expected cost.

[13]Although it may facilitate exposition, it is not strictly valid to compare coefficients across segments if error variances differ. We computed within-segment error (or residual) variance and concluded there were no significant differences across segments. Alternatively, the tests recommended by Swait and Louviere (1993) could be applied. We are grateful to an anonymous reviewer for this observation.

Table 6

MODEL FITS FOR MULTIPLE-SEGMENT LINEAR UTILITY STORE CHOICE MODELS

	Parameters	*Log-Likelihood*	*BIC*	*Validation*
Two Segments				
PFC Model 2.1 (Store intercepts)	9	–28,843.3	–28,889.7	–4,844.7
FFC Model 2.2 (2.1 plus ϕ)	11	–27,224.6	–27,281.3	–4,556.6
FFC Model 2.3 (2.1 plus θ)	11	–14,390.8	–14,447.5	–2,446.6
FFC Model 2.4 (2.1 plus ϕ, θ)	13	–13,881.0	–13,948.0	–2,377.4
FFC Model 2.5 (2.1 plus ϕ, θ, ω^a)	13	–13,768.8	–13,835.8	–2,351.5
TC Model 2.6 (2.5 plus β)	15	–13,339.6	–13,416.9	–2,311.4
TC Model 2.7 (2.5 plus β, δ_I, δ_P, δ_C)	21	–13,241.2	–13,349.4	–2,300.4
TC Model 2.8 (2.5 plus β, δ_I, δ_P, δ_C, ψ)	23	**–13,073.3**	**–13,191.8**	**–2,279.3**
Three Segments				
PFC Model 3.1 (Store intercepts)	14	–23,936.3	–24,008.5	–4,188.5
FFC Model 3.2 (3.1 plus ϕ)	17	–20,584.4	–20,672.0	–3,651.7
FFC Model 3.3 (3.1 plus θ)	17	–13,273.4	–13,361.0	–2,360.5
FFC Model 3.4 (3.1 plus ϕ, θ)	20	–13,137.8	–13,240.9	–2,374.8
FFC Model 3.5 (3.1 plus ϕ, θ, ω^a)	20	–13,004.8	–13,107.9	–2,422.2
TC Model 3.6 (3.5 plus β)	23	–12,506.3	–12,624.9	–2,253.8
TC Model 3.7 (3.5 plus β, δ_I, δ_P, δ_C)	32	–12,450.1	–12,615.0	–2,251.1
TC Model 3.8 (2.5 plus β, δ_I, δ_P, δ_C, ψ)	35	**–12,394.2**	**–12,574.6**	**–2,234.3**
Four Segments				
PFC Model 4.1 (Store intercepts)	19	–22,512.7	–22,610.6	–4,009.7
FFC Model 4.2 (4.1 plus ϕ)	23	–18,811.0	–18,929.6	–3,419.3
FFC Model 4.3 (4.1 plus θ)	23	–12,907.6	–13,026.2	–2,300.3
FFC Model 4.4 (4.1 plus ϕ, θ)	27	–12,667.3	–12,806.5	–2,277.0
FFC Model 4.5 (4.1 plus ϕ, θ, ω^a)	27	–12,598.3	–12,737.5	–2,260.3
TC Model 4.6 (4.5 plus β)	31	–12,190.6	–12,350.4	–2,224.2
TC Model 4.7 (4.5 plus β, δ_I, δ_P, δ_C)	43	–11,973.5	–12,200.3	–2,212.9
TC Model 4.8 (4.5 plus β, δ_I, δ_P, δ_C, ψ)	47	**–11,805.4**	**–12,047.7**	**–2,194.7**

[a]In multisegment models, separate ω parameters are not identifiable. For this reason, ω is fixed at the estimated value for the single-segment model.

Table 7

PARAMETER ESTIMATES FOR FOUR-SEGMENT LINEAR UTILITY TOTAL COST MODEL

	Segment 1		*Segment 2*		*Segment 3*		*Segment 4*	
Variable	*Parameter*	*t–ratio*	*Parameter*	*t–ratio*	*Parameter*	*t–ratio*	*Parameter*	*t–ratio*
Fixed Costs[a]								
α_1 (Store E1)	.000[b]	—	.000[b]	—	.000[b]	—	.000[b]	—
α_2 (Store E2)	.480	5.933	.124	2.853	−1.090	−9.981	−2.154	−4.106
α_3 (Store H1)	.076	.821	4.683	11.753	−.633	−3.599	−5.687	−11.294
α_4 (Store HH1)	−.107	−.894	6.414	6.416	−.562	−3.043	−5.717	−11.326
α_5 (Store HH2)	.414	3.268	4.824	10.252	−2.032	−17.172	−4.741	−9.390
ϕ (Distance)	1.370	9.005	.056	.240	1.003	3.953	3.738	18.299
θ (Category-independent loyalty)	−6.347	−51.416	−1.394	−22.426	−3.281	−29.974	−1.881	−21.067
ω (Power)	.409							
Variable Costs								
β (Basket cost)	1.254	21.811	2.122	32.701	1.710	28.048	1.541	11.789
δ_I (Inventory)	−.082	−.937	−.255	−4.723	.035	.480	−.032	−.480
$\delta_P[ap^d_s(i) - \mu_s(i)]$	.491	5.092	.316	2.693	.310	3.972	.419	5.046
δ_C (Consumption rate)	18.857	3.882	.000	.007	5.181	2.790	3.031	1.265
ψ (Category-specific loyalty)	−.675	−12.108	−.019	−.165	−.495	−9.746	−.544	−6.273
⇒ Minimum Weight[c]	.675		1.000		.758		.735	
Segment size	.446		.114		.083		.357	

[a]A negative intercept means a more preferred (lower fixed cost) store.
[b]Normalized for identification.
[c]$2 \times \left(\exp[\psi \times CL_h(i,s)]/\{1 + \exp[\psi \times CL_h(i,s)]\}\right)$, $CL_h(i, s) = 1.0$.

Table 8

SEGMENTS BY LOYALTY

Impact of Loyalty on Shopping Cost	*Category-Independent Loyalty*	
Category-Specific Loyalty	*High*	*Low*
High	Segment 1 (45%)	Segment 4 (36%)
Low	Segment 3 (8%)	Segment 2 (11%)

ment 2 (.48). Segment 1 members also are the most influenced by both store- and category-specific loyalty. This suggests that their shopping trips are likely to be the most purposeful, consistent with their having the highest list probability.

The parameters θ and ψ measure the impact of category-independent and category-specific store loyalty on the fixed and variable costs of shopping, respectively. The ranges for both are quite large. This implies substantial variance in the way households perceive habitual behavior as a means to reduce fixed and variable costs. The issue of habitual behavior is particularly interesting with respect to store choice. Casual empiricism suggests that store choices are substantially more stable than brand choices, which makes it worthwhile to understand the impact of habit on shopping costs. (In our data, we find that only 21% of households ever visit more than two supermarkets.) Furthermore, the habits are mutually reinforcing: Increased patronage of a particular store should, at the margin, increase category-specific purchases in that store; increases in category-specific loyalty across a range of categories should increase store loyalty.

As was expected, we find that the majority of households (45%) are highly responsive to both types of loyalty when evaluating shopping costs. In Table 8, using a two-by-two matrix, we sort (in a relative sense) segments according to their loyalty sensitivity.

As is shown in Table 7, segment 1 consists of 45% of the households. Relative to other segments, segment 1 has the most negative and significant values of both θ and ψ, which implies that its members are most likely to value habitual behavior and recognize the implicit benefit of both types of loyalty in reducing shopping costs. An alternative way to interpret this is that segment 1 members perceive a high cost of switching to another store or the habitual purchasing of certain items to other stores. To the best of our knowledge, our article is the first to quantify this implicit switching cost. Thus, segment 1 members are the least likely to have their store choice decisions influenced by drops in the expected prices of product categories.

Segment 2 households (11% of the total) have the least negative values of θ and ψ ($\theta_2 = -1.394$, t-ratio = −22.426; $\psi_2 = -.000$, t-ratio = .007) and are the least likely to be influenced by habitual behavior. Because $\psi_2 = 0$, segment 2 is not responsive to category-specific store loyalty and imputes the full expected cost of each category when deciding which store to visit. Thus, segment 2 members are the most likely to have their store choice decisions influenced by drops in the expected prices of product categories.

Segment 3 households (8% of the total) recognize the value of category-independent store loyalty on fixed cost reduction ($\theta_3 = -3.281$, t-ratio = −29.974). They perceive category-specific store loyalty as having a moderate influence on reducing variable cost ($\psi_3 = -.495$, t-ratio = −9.746). Conversely, segment 4 households (36% of the total) put relatively little emphasis on habitual behavior as a means of fixed cost reduction ($\theta_4 = -1.881$, t-ratio = −21.067) but perceive category-specific store loyalty as leading to a reduction in variable cost ($\psi_4 = -.544$, t-ratio = −6.273). One final observation pertains to the marginal distributions in Table 8. A high proportion of households (81%) are very loyal to stores for certain categories; a much smaller proportion (53%) are highly store loyal. This finding underscores the importance of category-based competition for supermarket retailers.

To investigate further why different segments prefer different stores, we assign households to market segments according to the posterior probability of segment membership (Equation 14). This enables us to examine the differences among segments in terms of distance between households and stores, demographic profiles, and shopping patterns. Table 9 presents the mean values of each of these classes of information for each segment. For each variable in Table 9 (e.g., distance to store E1, E2), we compute a one-way ANOVA across the market segments. In cases in which the F-statistic for the ANOVA is significant, we indicate pairwise differences between segments using letter designations. For example, segment 3 and 4 households are significantly closer to store H1 than segment 1 or 2 households are. Segment 2 households have the greatest distance from H1.

By examining the mean values reported in Table 9 and the shopping cost response parameters from Table 7, we can determine whether store choices are driven primarily by relative proximity or by other self-selection issues. Some examples follow: First, segment 4 should prefer stores H1 and HH1 because it is very sensitive to distance, $\phi_4 = 3.738$ (Table 7), and stores H1 and HH1 are the closest stores (Table 9). This speculation is confirmed because segment 4 does most of its shopping at stores H1 and HH1. Its behavior is consistent with the conventional wisdom that store choice is explained primarily by store location. However, this is not the case for segment 3. Segment 3 is relatively insensitive to distance, $\phi_3 = 1.003$ (Table 7), and is somewhat sensitive with respect to variable cost, $\beta_3 = 1.710$ (Table 7). Therefore, stores H1 and HH1 might not be the dominant stores for segment 3, even though they are the closest. This is confirmed because segment 3 spreads its shopping over different stores. Thus, store location alone cannot explain store choice adequately.

Next, we observe from Table 9 that segment 2 has the lowest per capita income and, from Table 7, that it is the most sensitive to variable cost, $\beta_2 = 2.122$. These observations suggest that segment 2 should prefer EDLP stores, and the shopping patterns in Table 9 confirm this. Specifically, segment 2 strongly prefers store E2 to HH2, even though stores E2 and HH2 are equidistant for segment 2. We also note that this segment is not sensitive to travel distance ($\phi_2 = .056$, t-ratio = .240).

Hoch and colleagues (1995) find a relatively strong relationship between demographics and store-level elasticities. We uncover a similar result; they are clear across segment

Table 9

SEGMENT MEMBER CHARACTERISTICS

	Segment 1 Mean (Standard Deviation)	*Segment 2* Mean (Standard Deviation)	*Segment 3* Mean (Standard Deviation)	*Segment 4* Mean (Standard Deviation)
Distance to:				
Store E1	1.539	1.440	1.627	1.753
	(1.082)	(1.298)	(.985)	(1.330)
Store E2	1.430[a,b]	1.762[a]	1.318[b]	1.446[a,b]
	(1.123)	(1.592)	(.645)	(1.064)
Store H1	1.440[b]	2.093[a]	.557[c]	.735[c]
	(1.797)	(2.127)	(.579)	(1.130)
Store HH1	1.723[b]	2.644[a]	.656[c]	.854[c]
	(1.797)	(1.848)	(.665)	(1.227)
Store HH2	1.430[a,b]	1.762[a]	1.318[b]	1.446[a,b]
	(1.123)	(1.592)	(.645)	(1.064)
Demographics:				
Family size	2.35[b]	2.94[a]	2.22[b]	1.64[c]
	(1.42)	(1.30)	(1.24)	(.99)
Income	34,587[a,b]	33,606[b]	42,453[a]	31,427[b]
	(22,578)	(21,016)	(28,432)	(25,058)
Head male age*	57[b,c]	52[c]	66[a]	63[a,b]
	(16.96)	(14.66)	(17.00)	(17.34)
Head female age*	57[b]	51[c]	61[b]	68[a]
	(17.68)	(15.59)	(18.85)	(15.05)
Per capita income	19,231[b]	13,221[c]	24,531[a]	20,054[a,b]
	(16,252)	(9423)	(20,627)	(15,689)
Number of Trips Per Household Taken to:				
Store E1	25[b]	77[a]	9[b,c]	0[c]
	(52)	(71)	(21)	(1)
Store E2	33[b]	70[a]	30[b]	2[c]
	(48)	(53)	(36)	(9)
Store H1	39[b]	1[c]	57[b]	132[a]
	(66)	(3)	(55)	(83)
Store HH1	15[b]	0[c]	24[a]	69[a]
	(42)	(1)	(56)	(68)
Store HH2	15[b]	1[c]	44[a]	24[b]
	(47)	(4)	(70)	(65)
Total	127[c]	148[b,c]	164[b]	228[a]

*Age is for head of household. Range is [24,80].

Note: Means with different letters are significantly different from each other: a > b > c, $p < .01$. Read across the rows.

differences in the demographic profiles. Furthermore, demographic profiles are related systematically to store formats. For example, larger, younger families with lower incomes per person prefer EDLP stores. With regard to shopping patterns, Ho, Tang, and Bell (1997) show that rational, cost-minimizing shoppers should increase their shopping frequency as stores increase price variability. We find strong evidence of this relationship. For example, members of the HILO segment (segment 4) average 228 trips per 18 months, whereas those in the EDLP segment (segment 2) average 148.

The pattern of results given in Tables 7 and 9 paints a striking picture that runs counter to the conventional retailing wisdom that location explains most of the variance in store choice. In particular, the average number of trips received by a particular store from a given segment is not necessarily correlated to distance. Our model and approach show that store choice is explained better by an analysis of consumer response to shopping costs, in which locational differences are captured as part of the fixed cost. Furthermore, we find that four distinct customer segments exist and their response profiles and shopping patterns are linked to their demographics. Finally, we identify interesting differences in household sensitivities to category-specific loyalty and the degree of planning prior to a shopping trip.

Basket size threshold. The third substantive objective of our article is to compute the basket size threshold and use it to understand market behavior. Recall that r^*_{sq} represents the basket size threshold (breakeven quantity) above which a shopper prefers store s and below which the shopper prefers store q (assuming that store s has higher fixed costs and lower variable costs than store q). As we noted previously, a concept based on pairwise comparisons might be useful in store choice, even when consumers have more than two alternatives. In our data set, the vast majority of consumers (79%) visit only one or two stores.

We compute $r^*_{sq|g}$ for all store pairs, s and q, $s \neq q$, and for each segment $g = 1, ..., 4$. With five stores, there are ten pairs of stores and therefore ten values of r^*_{sq}. Table 10 reports the average basket size purchased by each of our four segments (i.e., $\Sigma_{i=1}^{N} y_{i|g}$), the fixed and variable costs of shopping, and the basket size threshold r^*_{sq} for each store pair in each market segment. (For any given pair of stores [s, q], the unit variable costs are based on the common SKUs carried by both stores s and q. Because different stores carry different sets of SKUs, the variable cost of a store may vary when compared with different stores. The fixed cost, however, remains the same for a given store in each pairwise comparison.)

Table 10 shows that, in some cases, segment members strictly prefer one store over another. For example, segment 1 households strictly prefer store E1 over store H1 (this is indicated by $E1 \succ H1$), because it has both lower fixed and lower variable costs. There is no store that is dominated by another store in all four segments.

Using basket size threshold r^*_{sq}, we can determine which stores have the best market position with respect to a market segment. For example, in segment 1, we find that HILO stores (H1, HH1, HH2) dominate EDLP stores (E1, E2) for small basket sizes. This pattern is repeated when the relatively cheaper HILO store (H1) is compared with its higher-priced competitors (HH1, HH2). Similarly, in segment 2, we find that EDLP stores (E1, E2) dominate the other three stores but compete quite strongly in a head-to-head comparison. Stores also can use Table 10 to assess the marginal value of reducing variable costs (through lower pricing, reward programs, and so forth) and fixed costs (through improved service, better parking, higher quality, and so forth). There is, however, an important distinction between these two strategies of reducing fixed or variable costs. A store that reduces its fixed cost will increase patronage from shoppers who currently shop at competitor stores, at no revenue loss from its current customers. Reducing the variable cost, however, will increase patronage at the expense of a revenue loss from current customers (who now will pay lower prices for their products).

The basket size threshold also improves our ability to predict within-household variation in store choices over time. We computed household-specific hit rates using the basket size threshold and compared them with hit rates based on parameters from the four-segment FFC model (Model 4.5). The average hit rate (averaged across all households) is .834 for the basket size threshold model and .786 for Model 4.5. A simple t-test indicates that they are significantly different (t-ratio = 2.53). On a household-by-household comparison, we have 441 cases in which the hit rates are identical, 60 cases in which the threshold model gives a better hit rate, and 19 in which it gives a worse one. Thus, the basket size threshold captures both cross-sectional and longitudinal variation in store choices.

DISCUSSION

Standard retail site selection models assume that shoppers are influenced predominantly by store location and travel distance. Retail pricing models, however, suggest that shoppers respond positively to promotion and price discounts. The former is part of our fixed cost of shopping, whereas the latter is captured by our variable cost of shopping. Therefore, we integrate these two streams of literature and allow researchers to evaluate the relative impact of both factors.

A major conceptual and practical contribution of the article is the quantification of the basket size threshold for each segment for every pair of competing stores. The basic idea is that if shoppers shop for a large basket, they will prefer stores with a higher fixed cost and lower variable cost, because the fixed cost is divided across more items.[14] The expression of marketing mix variables (factors such as service influence fixed costs; pricing strategy influences variable costs) in terms of the basket size threshold enables us to study systematically how one store can gain store traffic at the expense of other stores. The concept of the basket size threshold is valuable because it provides insights into consumer behavior that are based on the analysis of total shopping costs rather than simple heuristics related to location or price responsiveness. For example, a proximity model would predict that shoppers equidistant from two stores choose each with equal probability, yet we find this is not the case. Our approach provides a better understanding of basket shopping, market segmentation, and store selection.

[14]This same idea can be found in technology choice literature, in which firms that face a large demand prefer technology that has a higher fixed and lower variable cost.

Positioning Strategy

In addition to describing market behavior, we also can use the basket size threshold to diagnose the relative competitiveness of a store. To be competitive in a market segment, a store should avoid having high fixed and high variable costs of shopping simultaneously. A store in such a position is vulnerable to losing store traffic and likely to be squeezed out of the market.

In serving multiple market segments, a store can adopt one of the following two positioning strategies: The first, which we call a *focus* strategy, is to always serve small (large) basket sizes in all market segments by having the smallest (largest) fixed and the highest (smallest) variable cost of shopping. The alternative, which we call a *diversified* strategy, is to serve a mix of basket sizes, serving large basket sizes in some segments and small basket sizes in others.

Both strategies can lead to gains in store traffic at the expense of competitors. However, the advantage of the diversified strategy is that it is more robust to changes in consumer purchasing habits. For example, assume that shoppers, under increasing pressure for time, decide to shop less often and buy bigger baskets. This change in shopping habits will hurt the store that adopts the small basket (i.e., low fixed cost, high variable cost) focus strategy but not the store with the diversified strategy. In our data, we find evidence of both strategies. Table 11 provides some interesting summary information about the average fixed costs and the difference in the variable costs imposed by a given store and its four competitors.

As was expected, EDLP stores impose higher average fixed costs. However, they impose lower average variable costs, such that a customer could expect to save 10.5% shopping at E1 or 13.7% at E2 when these stores are compared with their four competitors. E2 is the stronger of the two focused EDLP stores because it imposes, on average, lower fixed costs and yields higher average savings. Of all 16 pair-

Table 10

ESTIMATED FIXED AND VARIABLE COSTS OF SHOPPING (CONDITIONAL ON SEGMENT MEMBERSHIP)

Store Pair (s, q)	*Fixed Costs*	*Variable Costs*	Q^*_{sq}
Segment 1, average basket size: $\Sigma_{i=1}^{N} y_i \mid 1 = 6.219$			
(E2, E1)	(4.531, 3.486)	(3.542, 3.628)	12.25
(H1, E1)	(3.772, 3.486)	(3.251, 3.132)	E1 ≻ H1
(H1, E2)	(3.772, 4.531)	(3.219, 3.025)	3.93
(HH1, E1)	(2.997, 3.486)	(2.619, 2.212)	1.20
(HH1, E2)	(2.997, 4.531)	(2.549, 2.144)	3.78
(HH1, H1)	(2.997, 3.772)	(2.387, 2.135)	3.07
(HH2, E1)	(3.169, 3.486)	(3.113, 2.518)	.53
(HH2, E2)	(3.169, 4.531)	(3.021, 2.422)	2.27
(HH2, H1)	(3.169, 3.772)	(2.849, 2.397)	1.34
(HH2, HH1)	(3.169, 2.997)	(2.438, 2.394)	HH1 ≻ HH2
Segment 2, average basket size: $\Sigma_{i=1}^{N} y_i \mid 2 = 6.496$			
(E2, E1)	(.425, .314)	(3.471, 3.512)	2.71
(H1, E1)	(2.244, .314)	(2.212, 2.096)	E1 ≻ H1
(H1, E2)	(2.244, .425)	(2.222, 2.097)	E2 ≻ H1
(HH1, E1)	(3.060, .314)	(.482, .393)	E1 ≻ HH1
(HH1, E2)	(3.060, .425)	(.479, .391)	E2 ≻ HH1
(HH1, H1)	(3.060, 2.244)	(.351, .307)	H1 ≻ HH1
(HH2, E1)	(2.300, .314)	(.090, .082)	E1 ≻ HH2
(HH2, E2)	(2.300, .425)	(.067, .055)	E2 ≻ HH2
(HH2, H1)	(2.300, 2.244)	(.062, .064)	27.67
(HH2, HH1)	(2.300, 3.060)	(.009, .011)	HH2 ≻ HH1
Segment 3, average basket size: $\Sigma_{i=1}^{N} y_i \mid 3 = 4.766$			
(E2, E1)	(.555, 1.408)	(1.280, 1.333)	E2 ≻ E1
(H1, E1)	(.754, 1.408)	(1.663, 1.573)	7.25
(H1, E2)	(.754, .555)	(1.751, 1.595)	E2 ≻ H1
(HH1, E1)	(.347, 1.408)	(1.113, .930)	5.82
(HH1, E2)	(.347, .555)	(1.188, .959)	.91
(HH1, H1)	(.347, .754)	(1.809, 1.572)	1.72
(HH2, E1)	(–.094, 1.408)	(1.289, 1.064)	6.65
(HH2, E2)	(–.094, .555)	(1.894, 1.520)	1.74
(HH2, H1)	(–.094, .754)	(2.202, 1.955)	3.44
(HH2, HH1)	(–.094, .347)	(1.563, 1.603)	HH2 ≻ HH1
Segment 4, average basket size: $\Sigma_{i=1}^{N} y_i \mid 4 = 4.578$			
(E2, E1)	(2.511, 4.953)	(2.287, 2.340)	E2 ≻ E1
(H1, E1)	(–1.356, 4.953)	(2.442, 2.288)	40.77
(H1, E2)	(–1.356, 2.511)	(3.061, 2.840)	17.47
(HH1, E1)	(–1.630, 6.681)	(1.960, 1.638)	20.45
(HH1, E2)	(–1.630, 3.387)	(2.455, 2.018)	9.48
(HH1, H1)	(–1.630, –1.829)	(2.659, 2.376)	.97
(HH2, E1)	(.636, 4.953)	(1.930, 1.594)	12.85
(HH2, E2)	(.636, 2.511)	(2.170, 1.770)	4.68
(HH2, H1)	(.636, –1.356)	(2.345, 2.087)	H1 ≻ HH2
(HH2, HH1)	(.636, –1.630)	(2.055, 2.025)	HH1 ≻ HH2

Table 11
SHOPPING COSTS BY STORE

	Store				
	E1	*E2*	*H1*	*HH1*	*HH2*
Average Fixed Cost	2.54	2.01	1.35	1.19	1.50
Variable Costs (negative values indicate savings)					
Average premium	−10.5%	−13.7%	−2.8%	11.2%	12.6%
Maximum saving	−23.6%	−24.8%	−18.8%	−1.8%	−2.6%
Maximum premium	4.0%	−1.2%	8.9%	19.2%	19.9%

wise comparisons (with four stores over four segments), the worst this store does is yield a savings of 1.2%. Store H1 has a diversified positioning and is probably the least vulnerable to any dramatic change in the cost sensitivities or shopping habits of consumers. It imposes low average fixed costs (equivalent to those imposed by the focused high-tier competitors) and, on average, offers a 2.8% savings relative to the other four stores. Furthermore, it never charges a premium that exceeds 8.9% (relative to EDLP competitors) and offers savings of up to 18.8% (relative to higher-tier HILO competitors). We also observe from Table 7 that it has the top market share in three of the four segments (1, 3, and 4) and the highest overall market share, at 30%.

Limitations and Further Research

This research makes several simplifying assumptions. First, we assume that the store choice decision is driven primarily by the total cost of shopping. Although our shopping cost framework addresses many important factors, households might select a store on the basis of other factors (e.g., product assortment). We assume that the store assortments are identical and that only prices vary across stores. Second, we assume that the shopping list drives store choice. In practice, store choice also could have an impact on the composition of the shopping list. This so-called "endogeneity" issue has been shown to potentially bias parameters in a brand choice setting (Villas-Boas and Winer 1996). Third, we derive the shopping list from the purchased list using a model of conditional purchase incidence, which improves fit. It is, however, hard to judge its empirical validity because we do not observe the shopping list.

Our work can be extended in several directions. First, by measuring the shopping list, researchers can validate the proposed model of the shopping list and establish a better relationship between the *ex post* list of purchased items and the *ex ante* list of shopping items. Second, it is worthwhile to study how other factors, such as product assortment, affect store choice. Third, it would be interesting to study what kinds of product categories are more likely to generate the development of category-specific store loyalty. In this article, we estimate an aggregate sensitivity parameter (ψ); additional research should provide a more general specification. Fourth, stores might be interested to know whether private-label products generate more category-specific store loyalty than nationally branded products. This is interesting because shoppers can find nationally branded products at competing stores but cannot find the private-label products there. If the shopping list contains private-label items, shoppers must shop at the associated stores. Thus, private-label products might generate higher category-specific loyalty. Conceptually, all these new factors also can be studied in terms of their effect on consumers' fixed and variable costs of shopping (Tang, Bell, and Ho 1998).

APPENDIX: CALCULATIONS FOR BASKET SIZE THRESHOLD

The Unit Basket

The unit basket vector represents what shopper h would buy during a usual shopping trip. This standard basket is denoted by the requirement vector $[x_h(1), x_h(2), ..., x_h(N-1), x_h(N)]$, where $x_h(i)$ is household h's expected purchase quantity of product i. We compute each element, $x_h(i)$, as follows:

$$\text{(A1)} \qquad x_h(i) = \sum_{d:T_h^d = 1} \sum_{s=1}^{S} \frac{r_{hs}^d(i)}{TOTTRIP_h},$$

where $TOTTRIP_h$ is the total number of shopping trips taken by household h over all shopping days at all stores ($T_h^d = 1$ indicates that household h took a shopping trip on day d). $r_{hs}^d(i)$ represents the household's requirement for product i in store s on day d, that is, the actual quantity purchased adjusted by the probability that the item was on the consumer's shopping list. Thus, $x_h(i)$ is the average quantity of SKU i purchased by household h during a randomly drawn trip.

The corresponding realized price vector $[m_h(1), m_h(2), ..., m_h(N-1), m_h(N)]$ is computed analogously. The realized price (or prevailing market price) for SKU i, $m_h(i)$, is nothing but the volume-weighted average of all prices paid by shopper h on all trips in all stores, as follows:

$$\text{(A2)} \qquad m_h(i) = \sum_{d:T_h^d = 1} \sum_{s=1}^{S} \frac{r_{hs}^d(i) \times ap_s(i)}{TOTVOL_h(i)},$$

where $TOTVOL_h(i)$ is the total purchase volume for product i by household h, and $ap_s(i)$ is the actual price of product i at store s on shopping day d.

The computation of the basket vectors $y \mid g$ for each segment proceeds as in Equation A1. Similarly, the store-specific unit price of product i experienced by household h, $m_{h,s}(i)$, can be computed in accordance with Equation A2 but with one important difference. The summation and weighting of prices occurs only for visits that took place in store s (i.e., we replace $TOTVOL_h[i]$ with $TOTVOL_{h,s}[i]$ in Equation A2).

Calculation of the Threshold

Suppose that the shopper buys only product i. Because $E[V_h^d(s_1)] < E[V_h^d(s_2)]$, we have $m_{h,s_1}(i) < m_{h,s_2}(i)$. Then there is a threshold value of the requirement $r_h^{d*}(i)$ below which store s_1 has a higher expected total cost and above which store s_1 has a lower expected total cost. The threshold value is derived by setting $E[TC_h^d(s_1)] = E[TC_h^d(s_2)]$ and is given by

$$\text{(A3)} \qquad r_h^{d*}(i) = \frac{F_h(s_1) - F_h(s_2)}{m_{h,s_2}^d(i) - m_{h,s_1}^d(i)}.$$

We recognize that shoppers typically buy multiple products on any store visit. Therefore, to compute a basket-based, instead of product-based, threshold value, we must develop a notion of a unit basket and its associated price, the unit basket price.

Our measure of the unit basket price is akin to the idea of the *Consumer Price Index*—we determine the cost of the typical basket at prices prevailing across the entire market. A unit of the standard basket for shopper h constitutes what shopper h normally would plan to buy during a randomly drawn shopping trip. We denote a unit of the standard basket by the requirement vector $[x_h(1), x_h(2), ..., x_h(N - 1), x_h(N)]$, where $x_h(i)$ is the expected quantity of product i that household h will buy in a random shopping trip. We denote the modified expected price of a unit of the standard basket by m_h and the modified average price vector for each item as $[m_h(1), m_h(2), ..., m_h(N - 1), m_h(N)]$.[15] In this case, the expected price of a single unit of the standard basket $m_h = \Sigma_{i=1}^N x_h(i) \times m_h(i)$.

To demonstrate how we use the concept of a unit of the standard basket to capture relative basket price levels across stores, we consider the following shopping behavior. We assume that shopper h intends to buy a basket y at store s, where basket y is denoted by the quantity vector $[y_h(1), y_h(2), ..., y_h(N - 1), y_h(N)]$. Based on the prior shopping experience of shopper h at store s, the modified average price vector $[m_{h,s}(1), m_{h,s}(2), ..., m_{h,s}(N - 1), m_{h,s}(N)]$ at store s can be estimated by the average price paid by household h, weighted by the quantities purchased by household h at store s (see Equation A2). In this case, the basket size of this trip is specified by $r_h(y)$, where $r_h(y)$ represents the number of units of the standard basket that the shopper h intends to buy and where

$$\text{(A4)} \qquad r_h(y) = \frac{\sum_{i=1}^N y_h(i) \times m_h(i)}{m_h}.$$

In addition, the modified expected price of a unit of the standard basket at store s experienced by household h, denoted by $m_{h,s}$, can be expressed as

$$\text{(A5)} \qquad m_{h,s} = \frac{\sum_{i=1}^N y_h(i) \times m_{h,s}(i)}{r_{h(y)}}.$$

In this case, the expected total cost of a trip for basket size $r_h(y)$ at store s for household h is equal to $F_h(s) + r_h(y) \times m_{h,s}$. Consequently, the basket size threshold between a pair of stores s_1 and s_2 (the point at which the expected total costs incurred at s_1 and s_2 are equivalent) is as follows:

$$\text{(A6)} \qquad r_{h,12}(y)^* = \frac{F_h(s_1) - F_h(s_2)}{m_{h,s_2} - m_{h,s_1}}.$$

[15]The value of $x_h(i)$ can be estimated by averaging the quantities purchased by household h during all store visits, adjusted by the probability that the item was on the consumer's shopping list. The value of $m_h(i)$ can be estimated by calculating the average price paid, weighted by the purchase quantities during all store visits by household h. See Equations A1 and A2.

REFERENCES

Alba, Joseph W., Susan M. Broniarczyk, Terence A. Shimp, and Joel E. Urbany (1994), "The Influence of Prior Beliefs, Frequency Cues, and Magnitude Cues on Consumers' Perceptions of Comparative Price Data," *Journal of Consumer Research*, 21 (2), 219–35.

Assunçao, Joao and Robert J. Meyer (1993), "The Rational Effect of Price Promotions on Sales and Consumption," *Management Science*, 39 (5), 517–35.

Bell, David R. (1995), "Consumer Response to Retailer Pricing Strategies," doctoral dissertation, Graduate School of Business, Stanford University.

——— and James M. Lattin (1998), "Grocery Shopping Behavior and Consumer Response to Retailer Price Format: Why 'Large Basket' Shoppers Prefer EDLP," *Marketing Science*, 17 (1), 66–88.

Ben-Akiva, Moshe and Steven R. Lerman (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Bettman, James (1979), *An Information Processing Theory of Consumer Choice*. Reading, MA: Addison-Wesley.

Bowman, Russ (1987), "Sales Promotion: What's 'In Store' for 1987," *Marketing & Media Decisions*, 22 (January), 78, 80.

Brown, Stephen (1989), "Retail Location Theory: The Legacy of Harold Hotelling," *Journal of Retailing*, 65 (Winter), 450–70.

Bucklin, Randolph E. and Sunil Gupta (1992), "Brand Choice, Purchase Incidence, and Segmentation: An Integrated Approach," *Journal of Marketing Research*, 29 (May), 201–15.

——— and James M. Lattin (1991), "A Two-State Model of Purchase Incidence and Brand Choice," *Marketing Science*, 10 (Winter), 24–39.

——— and ——— (1992), "A Model of Product Category Competition Among Grocery Retailers," *Journal of Retailing*, 68 (Fall), 271–93.

Craig, Samuel C., Avijit Ghosh, and Sara McLafferty (1984) "Models of the Retail Location Process: A Review," *Journal of Retailing*, 60 (Spring), 5–36.

Currim, Imran S. (1982), "Predictive Testing of Consumer Choice Models Not Subject to Independence of Irrelevant Alternatives," *Journal of Marketing Research*, 19 (May), 208–22.

Dickson, Peter and Alan Sawyer (1990), "The Price Knowledge and Search of Supermarket Shoppers," *Journal of Marketing*, 54 (July), 42–53.

Fader, Peter S., James M. Lattin, and John D.C. Little (1992), "Estimating Nonlinear Parameters in the Multinomial Logit Model," *Marketing Science*, 11 (Fall), 372–85.

Gupta, Sachin and Pradeep K. Chintagunta (1994), "On Using Demographic Variables to Determine Segment Relationships in Logit Mixture Models," *Journal of Marketing Research*, 31 (February), 128–36.

Hardie, Bruce G.S., Eric J. Johnson, and Peter S. Fader, (1993), "Modeling Loss Aversion and Reference Dependence Effects on Brand Choice," *Marketing Science*, 12 (Fall), 378–94.

Ho, Teck-Hua, Christoper S. Tang, and David R. Bell (1997), "Rational Shopping Behavior and the Option Value of Variable Pricing," *Marketing Science*, forthcoming.

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

———, Byung-Do Kim, Alan M. Montgomery, and Peter E. Rossi (1995), "Determinants of Store-Level Price Elasticity," *Journal of Marketing Research*, 32 (February), 17–29.

Huff, David (1964), "Defining and Estimating a Trading Area," *Journal of Marketing*, 28 (July), 34–38.

Kahn, Barbara and David C. Schmittlein (1989), "Shopping Trip Behavior: An Empirical Investigation," *Marketing Letters*, 1 (1), 55–69.

——— and ——— (1992), "The Relationship Between Purchases Made on Promotion and Shopping Trip Behavior," *Journal of Retailing*, 68 (Fall), 294–315.

Kalwani, M.U., C.H. Yim, Heikki J. Rinne, and Yoshi Sugita (1990), "A Price Expectations Model of Customer Brand Choice," *Journal of Marketing Research*, 27 (August), 251–62.

Kalyanaram, G. and Russell S. Winer (1995), "Empirical Generalizations from Reference Price Research," *Marketing Science*, 14 (3:2), G161–79.

Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (November), 379–90.

Lal, Rajiv and Carmen Matutes (1994), "Retail Pricing and Advertising Strategies," *Journal of Business*, 67 (3), 345–70.

——— and Ram Rao (1997), "Supermarket Competition: The Case of Every Day Low Pricing," *Marketing Science*, 16 (1), 60–80.

Lattin, James M. and Randolph E. Bucklin (1989), "Reference Effects of Price and Promotion on Brand Choice Behavior," *Journal of Marketing Research*, 26 (August), 299–310.

——— and Gwendolyn K. Ortmeyer (1991), "A Theoretical Rationale for Every Day Low Pricing," working paper, Graduate School of Business, Stanford University.

Morwitz, Vicki and Lauren Block (1996), "An Analysis of Planned Grocery Shopping: Influences on Consumer Goal Setting and Goal Fulfillment," working paper, Stern School, New York University.

Mulhern, Francis J. and Robert P. Leone (1990), "Retail Promotional Advertising: Do the Number of Deal Items and Size of Deal Discounts Affect Store Performance?" *Journal of Business Research*, 21 (3), 179–94.

Park, C. Whan, Easwar S. Iyer, and Daniel C. Smith (1989), "The Effects of Situational Factors on In-Store Grocery Shopping Behavior: The Role of Store Environment and Time Available for Shopping," *Journal of Consumer Research*, 15 (March), 422–33.

Progressive Grocer (1995), (April).

Rickard, Leah (1995), "Shopping Lists Tell Just Half the Real Story," *Advertising Age*, (January 9), 20.

Simester, Duncan I. (1995), "Signaling Price Image Using Advertised Prices," *Marketing Science*, 14 (2), 166–88.

Swait, Joffre and Jordan Louviere (1993), "The Role of the Scale Parameter in the Estimation and Comparison of Multinational Logit Models," *Journal of Marketing Research*, 30 (August), 305–14.

Tang, Christopher S., David R. Bell, and Teck-Hua Ho (1998) "Retail Strategies, Shopping Costs, and Grocery Store Competition," working paper, Anderson School of Management, UCLA.

Villas-Boas, J. Miguel and Russell S. Winer (1996), "Endogeneity in Brand Choice Models," working paper, Haas School of Business, UC Berkeley.

Winer, Russell S. (1986), "A Reference Price Model of Brand Choice for Frequently Purchased Consumer Products," *Journal of Consumer Research*, 13 (September), 250–56.