Abstract
In recent years, the supermarket industry has become increasingly competitive. One outcome has been the proliferation of a variety of pricing formats, and considerable debate among academics and practitioners about how these formats affect consumers' store choice behavior. This paper advances the idea that consumer shopping behavior (as defined by average size of the shopping basket and the frequency of store visits) is an important determinant of the store choice decision when stores offer different price formats. A recent Wall Street Journal article that summarized the result of Bruno’s management switching the chain from EDLP to HILO illustrates the importance of this issue: “The company’s price-conscious customers, used to shopping for a fixed basket of goods, stayed away in droves.” Thus, the audience for this paper includes practitioners and academics who wish to understand store choices or predict how a change in price format might affect store profitability and the mix of clientele that shop there.

This paper attempts to understand the relationship between grocery shopping behavior, retail price format, and store choice by posing and answering the following questions. First, after controlling for other factors (e.g., distance to the store, prior experience in the store, advertised specials), do consumer expectations about prices for a basket of grocery products (“expected basket attractiveness”) influence the store choice decision? This is a fairly straightforward test of the effect of price expectations on store choice. Second, are different pricing formats (EDLP or HILO) more or less attractive to different types of shoppers? To adequately answer the second question, we must link consumers’ category purchase decisions, which collectively define the market basket, and the store choice decision.

We study the research questions using two complementary approaches. First, we develop a stylized theory of consumer shopping behavior under price uncertainty. The principal features and results from the stylized model can be summarized as follows. Shoppers are defined (in a relative sense) as either large or small basket shoppers. Thus, we abstract from the vicissitudes of individual shopping trips and focus on meaningful differences across shoppers in terms of the expected basket size per trip. The shoppers make category purchase incidence decisions and can choose to shop in either an EDLP or a HILO store. Large basket shoppers are shoppers who have a relatively high probability of purchase for any given category, and as such they are more captive to prices across many different categories. The first two propositions summarize the price responsiveness of shoppers. In particular, the large basket shoppers are less responsive to price in their individual category purchase incidence decisions; this makes them more responsive to the expected basket price in their store choice decisions. This key structural implication of the model highlights an asymmetry between response at the category level and response at the store level. The result is quite intuitive; a (large basket) shopper with less ability to respond to prices in individual product categories will be more sensitive to the expected cost of the overall portfolio (the market basket) than choosing a store. The final proposition derives the price at which a given shopper will be indifferent between an EDLP and a HILO store. The key insight is that as a shopper increases his or her tendency to become a large basket shopper, the EDLP store can increase its (constant) price closer and closer to the average price in the HILO store. Conversely, as the shopper becomes more of small basket shopper, the EDLP store must lower its price closer to the deal price in the HILO store. Thus, we have the interesting result that small basket shoppers prefer HILO stores, even at higher average prices.

The empirical testing mirrors the development of the consumer theory. We test the implications of the propositions using a market basket scanner panel database. The database includes two years of shopping data for 1,042 households in two separate market areas. We first use household-level grocery expenditures to model the probability that a household is a large or small basket shopper. Subsequently, we estimate purchase incidence and store choice models. We find that after controlling for important factors such as household distance to the store, previous experience in the store, and advertised specials, price expectations for the basket influence store choice. Furthermore, EDLP stores get a greater than expected share of business from large basket shoppers; HILO stores get a greater than expected share from small basket shoppers. Consistent with the implications of the propositions, large basket shoppers are relatively price inelastic in their category purchase incidence decisions and price elastic in their store choice decisions.

(Shopping Behavior; Choice Models; Pricing; Market Basket)
1. Introduction

Supermarket retailers are actively engaged in formulating pricing strategies. In fact, retail pricing strategy is consistently viewed by practitioners as the one of the “top five priorities in retail management” (Progressive Grocer 1995). Supermarket managers, in particular, stress that “nothing is more important in business than getting the pricing strategy right” (Supermarket Business 1993). Some retailers position themselves on the basis of “Low Prices, Everyday” across a wide assortment of product categories, while others offer temporary deep discounts in a smaller group of categories. The former strategy is commonly known as “EDLP,” the latter as “HILO.”

Thus, the pricing activity of the retailer is quite complex: It involves a strategic choice (EDLP or HILO) and the setting of specific prices across a broad range of products. Similarly, the shopping behavior of the consumer is also complex. Shoppers purchase products from a wide range of categories and often visit multiple stores. Prior research (Dickson and Sawyer 1990) suggests that while consumers may have relatively poor knowledge of individual product prices, they can make accurate distinctions about price levels in different stores (Alba et al. 1994).

It is in this setting that we seek answers to two questions that have yet to be fully addressed by either academic researchers or the trade. First, does pricing activity influence store choice? Second, is there evidence of any systematic relationship between the store’s price format and type of shopper it attracts? That is, are some types of consumers (as defined by shopping behavior and demographics) more apt to prefer HILO and others EDLP? This research sheds light on both issues. First, we show that price expectations influence store choice. Second, we establish a link between purchase incidence and store choice and show that consumers’ preferences for store price format—EDLP or HILO—depend upon their shopping behavior.

1.1. Price Format and Store Choice

Even though practitioners are very interested in price format and academic researchers (e.g., see Blattberg and Neslin 1990, Neslin 1994) consider it a topic worthy of research, few studies investigate the relationship between store price format and consumer behavior. In part, the research effort has been hindered by a lack of data on the appropriate unit of analysis (i.e., the total basket of purchases made by consumers on their shopping trips). Not surprisingly, single category studies by Kumar and Leone (1988) and Bucklin and Lattin (1992) offer mixed evidence of any direct effect of category-level marketing activity on consumers’ store choice decisions. Using aggregate-level sales data, Kumar and Leone find that store switching patterns are influenced by merchandizing activity in disposable diapers; Bucklin and Lattin (1992) analyze household-level scanner panel data and find no direct store choice effect for liquid laundry detergent. Mulhern and Leone (1990) present an event study of a discrete change in store price format at a single retailer. They found that sales increased when the store switched from EDLP to HILO.

Hoch, Dreze, and Purk (1994) report the impact of price increases and decreases on consumer sales response. Relatively inelastic response to price increases and decreases led the researchers to conclude that an EDLP format might be undesirable as a strategy for attracting new customers. However, this study focused on individual categories in isolation, and did not consider the relationship between price format, shopping behavior, and price-based expected utility for a basket of goods. Lal and Rao (1997) suggest that EDLP and HILO are equilibrium retailer positioning strategies when consumers have heterogeneous transport (store visit) costs. Ho, Tang, and Bell (1997) show that consumers derive an “option value” from price variability, specifically, that rational, cost-minimizing shoppers visit a store more frequently and purchase smaller average quantities per trip as price fluctuation at the store increases.

1.2. Shopping Behavior and Segmentation by Expected Basket Size

Grocery shopping behavior has three unique characteristics that suggest a relationship between shopping behavior and preference for different price formats: (1) Consumers typically shop for multiple items on a given trip; (2) for most of these items, they are usually unable to determine actual prices before visiting the store;¹

¹ Only a small subset of items is advertised, and this subset of items typically differs from store to store.
and (3) grocery shopping is repetitive—while individual trips may differ somewhat, most consumers settle into specific shopping patterns with respect to the average basket size per trip and frequency of shopping. Together, these three factors suggest that (1) store choices (if influenced at all by pricing) will be influenced by prices for a “basket” of multiple items, (2) price expectations (rather than actual prices) will be the mechanism for this influence, and (3) it may be useful to segment consumers according to fundamental differences in shopping behavior.

There are two studies that indicate shopping behavior may play a role in determining store choice. Kahn and Schmittlein (1992) show that consumer tendencies to use coupons is greater on “major” shopping trips. Bucklin and Lattin (1991) show that purchase incidence decisions and consumer response to category marketing activity vary according to the consumer’s prior state (“planned” or “opportunistic”). In the spirit of this work, we claim and show that across-household variance in the choice of store (format) is strongly linked to heterogeneity in households’ shopping behavior. We operationalize a measure that characterizes a household’s shopping behavior by first recognizing that household-level shopping behavior can be completely described by three variables: expected basket size per trip, shopping frequency, and total aggregate consumption (across all trips). (Note that if you know any two of these things, the third follows automatically.) These variables are highly correlated—households with larger expected basket sizes shop less frequently. (In our data, the correlation between these variables is −0.70.) This implies that most of the across-household variance in shopping behavior can be accounted for by controlling for just one of the three variables.

In this paper, we choose expected basket size per shopping trip as the primary segmentation variable and subsequently account for the residual variance by including shopping frequency in the model. (Given the high correlation between the two variables, it should be possible to obtain similar findings by focusing on shopping frequency first and expected basket size second.) We focus on expected basket size first because we believe it is the most meaningful primary segmentation variable of the three available. It is observable by the individual retailer (each checkout receipt is a “basket”), it relates directly to purchase incidence and in-store behavior (i.e., is managerially relevant), and leads to clear distinctions between households. Conversely, most retailers do not have access to information on the shopping frequency of all their customers. Thus our unit of analysis is the individual shopper, as defined by a single summary measure: expected basket size per trip. All propositions and proofs are based on this measure.

1.3. Overview and Contribution
The intuition underlying our model of store choice derives from a simple thought experiment (which we elaborate upon and formalize in §2). Imagine for the moment that there are two types of shopper: a large basket shopper and a small basket shopper. Both types of shopper consume the same aggregate quantity of groceries over some fixed time horizon. The total bundle of grocery requirements consists of grocery needs over multiple product categories, and the shoppers can choose between two types of store: an EDLP store and a HILO store. The EDLP store adopts a constant Every Day Low Price for every product category. The promotional pricing (HILO) store sometimes prices a category at the regular price and other times at a deep discount. The pricing is such that the constant category price in the EDLP store is always lower than the expected price in the HILO store.

On a given trip, the large basket shopper purchases in many product categories, and therefore fulfills a relatively large percentage of his total grocery needs on a single visit. This implies that he is captive to pricing across a wide array of categories and is relatively lacking in flexibility to take advantage of occasional price deals. This type of shopper will (all else equal) prefer EDLP due to the lower expected basket price. The

---

3 This assumption is for ease of exposition and is not part of our model development.
4 Pure versions of HILO and EDLP seldom exist in practice and EDLP/HILO is best thought of as a continuum. It is, however, usually possible to differentiate stores according to their relative tendency to operate one format or the other. We elaborate in on this in §3.
small basket shopper, on the other hand, can benefit from price variation in the store, even at higher average prices. She divides her total consumption needs into many smaller baskets—buying more product categories when the prices are relatively low, and deferring purchase when prices are high. Deferring purchase in high-priced categories is not problematic as she will soon return to the store. Our example illustrates an important intuition: The shopping behavior of the consumer determines his ability to profit from variance in store price format, *ceteris parabus*. Therefore, if one segments the market into large and small basket shoppers, it should be the case that they prefer EDLP and HILO, respectively (all other things being equal).

To summarize, this paper makes three new contributions: (1) a demonstration that basket-level price expectations influence store choice, (2) a rationale for why large basket shoppers prefer EDLP and small basket shoppers prefer HILO, and (3) a theory of how consumers’ market basket purchase incidence decisions relate to and influence the store choice decision. We outline a theory of store choice behavior for large and small basket shoppers, develop propositions, and test their implications using market basket scanner panel data. Recall that we focus on differences across households (i.e., segmentation across shoppers) and we do not seek to make a contribution to the understanding of intertemporal variation in a given household’s shopping patterns.

Our model has the following structural implications. Large basket shoppers are relatively less price elastic in their individual category purchase incidence decisions; this makes them more price elastic (with respect to expected basket prices) in their store choice decisions. This idea can be represented in a simple $2 \times 2$ matrix:

<table>
<thead>
<tr>
<th>Shopper Type</th>
<th>Category</th>
<th>Store</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Basket</td>
<td>Elastic</td>
<td>Inelastic</td>
</tr>
<tr>
<td>Large Basket</td>
<td>Inelastic</td>
<td>Elastic</td>
</tr>
</tbody>
</table>

Interestingly, our key finding has some parallels to earlier literature. For example, Krishnamurthi and Raj (1991) show that brand loyal households are relatively inelastic in brand choice, yet relatively elastic in purchase quantity decisions. This phenomenon has also been observed by Bucklin, Gupta, and Siddarth (1998).

The implications of our propositions are strongly supported by an extensive set of empirical analyses that cover 1,042 shoppers in two distinct metro markets with EDLP and HILO stores. Our work also has some clear insights for retailers. For example, if an EDLP format is attractive to large basket shoppers (and HILO is more attractive to small basket shoppers), this has obvious implications for whom different types of retailers should target. In §6 we expand our analysis to link consumer tendencies to be large and small basket shoppers to household demographics. Therefore, our analyses help retailers understand why different types of shoppers prefer the formats they do, what the important characteristics of these shoppers are, and how retailers can appeal to each class of shopper.

The paper proceeds as follows. In §2 we present the store choice theory and link it to a utility model based on the nested logit. This utility model forms the basis of our econometric work in §4 and §5. Section 3 describes the dataset, §4 presents the empirical model specifications, §5 reports the findings, and §6 develops implications for retailers and concludes the paper.

### 2. Theory Development

#### 2.1. Shopping Behavior: Large and Small Basket Shoppers

Previous research has looked at differences in shopping behavior *within* an individual household over time. Kahn and Schmittlein (1989) point out that shopping behavior may be different depending upon whether the household is taking a major trip to the store or just a fill-in trip (these trips are defined by whether the total expenditure on groceries is above or below the modal expenditure on groceries, respectively). Kahn and Schmittlein (1992) show that the use of coupons is more prevalent when the household takes a major shopping trip.

Our research proceeds in the same spirit. That is, we also argue that underlying shopping behavior plays an
important role in the response patterns exhibited by the consumer. Our focus, however, is on characterizing the variation in shopping behavior across households (rather than across shopping occasions). Our primary rationale lies in our focus on store choice: We are more interested in explaining store preference across households rather than in being able to predict which store a household will choose on a particular shopping trip. We also note that while it is relatively easy to classify shopping trips after the fact according to the total amount spent, it is very difficult to predict a priori whether a household will take a (relatively) large or small trip on a given occasion (see Bell 1995).

Our intuition suggests that small basket shoppers—i.e., those who satisfy a relatively small proportion of their total demand for groceries on each shopping trip—have greater flexibility with respect to taking advantage of price variation over time. These are the households that, all other things remaining equal, will prefer to shop in a HILO store because they can get a lower overall average price through opportunistic purchase behavior. We argue that for these shoppers the expected basket attractiveness will be higher in a store whose format offers greater variability in price, even at higher average prices. Conversely, the large basket shopper (who must purchase a relatively large proportion of their total grocery demand on each trip) will prefer to shop in an EDLP store where the overall expected price across a wide variety of categories is lower.

From an empirical point of view, the question remains as to how we should capture the differences between these different types of shoppers. Our approach is to build an index, based on observable characteristics of shopping behavior, that reflects the household’s potential flexibility for responding to price variation in the store environment. We now briefly sketch out the ideas underlying this index; the formal derivation of the index is described in §4.1.

Figure 1 shows a plot of grocery shopping frequency (average number of trips per year) and total grocery expenditures (average dollars spent per year) for 548 household panelists from a suburban metropolitan market (for a more detailed description of the data used in this research, see §3). Note that there is considerable variation across households in shopping frequency, ranging from fewer than 50 trips per year to more than 400 trips. As we noted in §1.2 one possible approach to building our index would be simply to use the average shopping frequency of the household: The higher the frequency of shopping, the greater the flexibility of the household to take advantage of price variation (since each trip accounts for a smaller proportion of the total demand for groceries). The problem with this approach, however, is that it does not control for the substantial heterogeneity across households in annual grocery expenditures (among those households who shop on average 100 times per year, there are some who spend less than $2,000 and others who spend more than $6,000).

We therefore propose an index based on a combination of total expenditures and shopping frequency. The diagonal line in Figure 1 describes a set of households whose average expenditures per trip—i.e., the expected basket size—are the same (in this case, $20). Everyone above the diagonal line has an expected basket size of less than $20. As one moves toward the upper left corner of the graph (i.e., as the distance from the diagonal line increases), the average basket size decreases. A household in the upper left, with a relatively small expected basket size, should have the opportunity to buy opportunistically, picking up an item or
two in response to a sale or delaying the purchase of an item because it is not on sale (and the household will not run out of it before the next shopping trip). In the lower right corner of the quadrant are the households with the least flexibility in their shopping.

Our actual index (described in more detail in §4.1) is an estimate of the probability that the household will spend $20 or less on a given shopping trip. It is a continuous measure between zero and one, designed to discriminate between households that we designate “large basket shoppers” (those in the lower right, who on average buy large baskets and shop relatively infrequently, denoted LBS) and those we designate “small basket shoppers” (those in the upper left, who on average buy small baskets and shop relatively frequently, denoted SBS). Note that the index is constructed a priori (i.e., the probabilities are not estimated as part of the model). Our goal is to test if the proposed index does in fact improve our ability to predict choice of different stores according to price format.

Figure 2 gives a conceptual representation of the predictions from our model. The figure shows five shoppers (A, B, C, D, and E) who we expect will differ in their relative store preference. (Note that households A, B, and C all have the same expected basket size per trip.) Relative to household A, we expect D to have a higher preference for HILO and E to have a higher preference for EDLP (due to the effect of expected basket size on store preference). A comparison of the relative preferences of A, B, and C is based on a frequency argument, which we consider in §5.3 in an extension to the main results. We expect C to have a higher preference for HILO and B to have a higher preference for EDLP. This is because, all else equal, a household who shops more often can gain from exposure to price variation. Figures 1 and 2 reiterate our choice of expected basket size as the primary segmentation variable. A greater number of shoppers lie to the left or right of the expected basket size line than lie on the line.

2.2. Choice Behavior of Large and Small Basket Shoppers

In the introduction, we argued that the small basket shopper, with the flexibility to take advantage of occasional price deals, may prefer to shop in a HILO store even when the expected prices are lower in the EDLP store. We now offer an illustration to help formalize our intuition regarding the relationship between shopping behavior and preference for store price format. To keep things simple, we will assume that consumers choose between two different stores. Each store carries N product categories (each consisting of a single brand). The stores differ with respect to their pricing format. One charges a fixed everyday low price $E$ for its products, and the other engages in a promotional strategy in which it charges a low discount price $L$ with probability $p$ and a high regular price $H$ the remainder of the time. We denote the first store EDLP and the second HILO. In order to reflect what we observe in the real world (i.e., that EDLP stores generally have a lower average price than their HILO competitors), we assume that $E < \pi L + (1 - \pi)H$. While the consumer is aware of the prices charged by the stores, he/she does not know exactly when the discount price will be available in the HILO store.

We model the probability that the consumer makes a category purchase using the logit model. Thus, we assume that the utility for a purchase in category $c$ (relative to the no-purchase alternative) is given by:

$$U_c = V_c + \epsilon_c,$$  (1)
where $V_c$ represents the deterministic component of utility (i.e., that which can be modeled explicitly) and $\epsilon_c$ is a double-exponentially distributed error term. For the purposes of this illustration, we will write $V_c$ as a simple linear function of price:

$$V_c = V_c^* - \gamma_c \text{Price}_c. \quad (2)$$

Note that it is straightforward to include other factors such as product category inventory that might also influence category choice behavior (and we do this in §4$^6$) but for the purposes of this illustration we limit our focus to the effect of price. Following the standard assumptions of the logit model, the probability of category purchase is given by:

$$p_c = \frac{\exp(V_c)}{1 + \exp(V_c)}. \quad (3)$$

We now compare two shoppers who differ with respect to their expected basket size. The large basket shopper has a greater probability of purchasing from the product category on any given trip. We capture this difference in purchase propensity via a single model parameter: the deterministic utility, net of price, $V^*_c$ (given in Equation (2)). Thus, the models describing purchase incidence for the two types of shoppers are exactly the same except that we have $V^*_\text{LBS} > V^*_\text{SBS}$. $V^*$ should be thought of as a household-level parameter that determines the probability of purchase incidence.

The abstraction from the vicissitudes of individual trips allows us to focus on meaningful differences across shoppers in expected basket size per trip. We define $n_{\text{LBS}}$ and $n_{\text{SBS}}$ as the number of trips per year taken by the large and small basket shopper, respectively. Note that the values of $n$ and $V^*$ can be chosen so that large basket shoppers and small basket shoppers purchase the same total quantity of groceries over the course of a year. In Figure 2, these are all shoppers who lie on a horizontal line that crosses the expected basket size line (e.g., shoppers D, A, and E). For the moment, we focus on expected basket size per trip; however, in §5.3 we provide results for the case in which expected basket size is the same for two shoppers, but they differ in shopping frequency (e.g., shoppers A, B, and C in Figure 2).

Our model of category purchase incidence has the following implications for the price responsiveness of different types of shoppers:

**Proposition 1.** The large basket shopper is less responsive (as measured by category purchase elasticity) than the small basket shopper to changes in category prices.

**Proof.** The result follows from the fact that the category price elasticity is given by $-\gamma_c (1 - P_c)$. The larger the probability of category purchase incidence $p_c$ as determined by $V^*_c$, the smaller the elasticity (in absolute value).$^7$

So far, we have described the process by which a category purchase is made once the consumer has entered the store. The decision of which store to visit will depend upon the expected outcome of the purchase incidence decisions across all categories in the consumer’s basket. All other things remaining equal (e.g., distance to the store, shopping convenience, etc.), the consumer is more likely to shop in the store offering the greater expected basket attractiveness.

A useful property of the logit model (Ben Akiva and Lerman 1985, McFadden 1984) is that it provides a closed form expression for the expected maximum utility of a choice. Once the shopper has chosen to visit a particular store, the expected maximum utility of the category purchase decision (described by the logit model in Equations (1), (2), and (3)) is given by:

$$CA_c = \ln[1 + \exp(V_c)], \quad (4)$$

where $CA_c$ denotes the attractiveness of category $c$ and where $V_c$ is a function of the prevailing price in category $c$ in the chosen store (Equation (2)). Note that this expression encompasses both purchase and nonpurchase outcomes. It is therefore interpretable as the expected maximum utility, once in the store, of making a category purchase decision in that store. Assuming that category purchase decisions are independent across categories,$^8$ we can sum across categories to get

$^6$ For example, $V^*_c = \gamma_b + \gamma_1 \text{CR} - \gamma_2 \text{INV} - \gamma_3 \text{PRICE}$, where CR and INV are consumption rate and inventory, respectively.

$^7$ We have not explicitly addressed stockpiling in our analysis. We would expect that the ability to stockpile would provide the same type of flexibility as buying smaller baskets on a more frequent basis.

$^8$ Lattin et al. (1997) find that with a relatively small number of exceptions (e.g., pairs of categories such as detergent and fabric softener), the assumption that category purchase decisions are made independently is reasonable.
an overall measure of the basket attractiveness offered by the particular store:

\[
BA = \sum_c \ln[1 + \exp(V_c)].
\] (5)

Note that category weighting is taken care of in the term \(V_c\). When a household purchases from a category only infrequently, \(V_c\) will have a large negative intercept term, driving the value of \(\ln[1 + \exp(V_c)]\) nearer to zero.

In making the decision of which store to visit, the consumer must assess the attractiveness of the basket before entering the store. The expression will therefore be a function of the shopper’s expectations regarding the prevailing prices. In the case of the EDLP store, in our stylized example, the product is always available at price \(E\) so the expected basket attractiveness is given by:

\[
EBA(EDLP) = \sum_c \ln[1 + \exp(V_c^* - \gamma_c E)].
\] (6)

In the HILO store, the consumer does not know the prices prevailing in each category before entering the store. He/she expects to encounter a low price \(L\) with probability \(\pi\) or a high price \(H\) with probability \((1 - \pi)\); thus, the expected basket attractiveness is given by:

\[
EBA(HILO) = \sum_c \pi \cdot \ln[1 + \exp(V_c^* - \gamma_c L)] + \sum_c (1 - \pi) \cdot \ln[1 + \exp(V_c^* - \gamma_c H)].
\] (7)

It is this measure of expected basket attractiveness\(^9\) that we include in our model of store choice to capture the appeal of different pricing formats to different types of shoppers.

In Proposition 1, we considered the difference in price responsiveness across different types of shoppers at the category level. We can now investigate the price responsiveness of different types of shoppers to changes at the level of the market basket.

**Proposition 2.** The large basket shopper is more responsive (as measured by store choice elasticity) than the small basket shopper to changes in basket prices.

\(^9\) In §4, we discuss the details of operationalizing the EBA measure in a real-world environment much more complex than our highly stylized environment.

**Proof.** See Appendix A.

Taken together, Propositions 1 and 2 suggest that a shopper who is less price elastic in his or her individual category purchase decisions will be more price elastic when it comes to the store choice decision. As noted in §1, this type of phenomenon has been documented in the literature (Krishnamurthi and Raj 1991, Bucklin, Gupta, and Siddarth 1998) for other combinations of choice decisions.

In Equations (6) and (7), we can affect the type of shopper by manipulating the parameter \(V^*\); Small values correspond to small basket shoppers (i.e., those who have a relatively low probability of purchasing the category on any given store visit) and large values correspond to large basket shoppers. We now show that our measure exhibits the following important property: Large basket shoppers have a higher expected basket attractiveness in EDLP stores, while small basket shoppers have a higher expected basket attractiveness in HILO stores. We first present a simple numerical example, then formalize the result in Proposition 3.

We begin by choosing a set of prices \((L, H, \text{ and } E)\), deal probability \((\pi)\) and a value of \(V^*\) such that the consumer is indifferent between the two price formats. We then show that as \(V^*\) increases, \(EBA(EDLP) > EBA(HILO)\), implying that as a household becomes more of a large basket shopper, its preference for EDLP increases. Let \(L = 1.00, H = 2.00, \text{ and } \pi = 0.5.\) For \(V^* = 0\), it is easy to show that at an everyday low price of \(E = 1.4016\), the consumer is indifferent between HILO and EDLP (in both cases, \(EBA = 0.2201\)). If we increase \(V^*\) to 1.0, \(EBA(EDLP) = 0.5124\) and \(EBA(HILO) = 0.5032.\) If we decrease \(V^*\) to \(-1.0, EBA(EDLP) = 0.0867\) and \(EBA(HILO) = 0.0878.\)

We now state this result more formally.

**Proposition 3.** All other things remaining equal, the more the consumer tends toward being a large basket shopper, the greater his or her preference for the EDLP pricing format.

**Proof.** See Appendix A.

The results of Propositions 1, 2, and 3 have testable implications. First, after controlling for other factors (e.g., location, advertised specials, etc.), our measure of expected basket attractiveness, which controls for...
across-shopper differences in responsiveness and across-store differences in pricing, should be an important predictor of store choice. Second, after controlling for variation across households in category usage and preference structure, EDLP stores (characterized by lower average prices and lower in-store price variability) should receive a higher share of visits from large basket shoppers; for HILO stores (characterized by lower deal prices and higher in-store price variation), the opposite should hold. The two questions posed in the Introduction can now be restated as hypotheses:

Hypothesis 1. In markets with HILO and EDLP price formats and variation in consumers’ shopping trip behavior, a price-based measure of expected basket attractiveness will help explain store choice, and

Hypothesis 2. Variation in this price-based measure of expected basket attractiveness will have a systematic effect on store choice. Large basket shoppers will prefer EDLP; small basket shoppers will prefer HILO.

There is a plausible alternative explanation that runs counter to the intuition developed in this section. Previous research (e.g., Kahn and Schmittlein 1989) suggests that the smallest baskets are purchased by consumers on “fill-in” trips, which are taken with the sole objective of purchasing items that have run out at home. (Note, however, that in defining the fill-in trip, Kahn and Schmittlein focus on within-shopper behavior; we model differences across shoppers.) In contrast to our picture of opportunistic purchase behavior, this notion of a fill-in trip suggests a shopper who is for the most part price inelastic at the category level, and not especially responsive to deals. If small basket shoppers are in fact taking a larger proportion of these inelastic fill-in trips, then we should not expect to see the relationship between EBA and store choice proposed above.

3. Data
To carry out our empirical tests, we needed to build a database to capture shopping behavior not in a single category (like most of the datasets analyzed in the literature), but across a variety of categories in the market basket. With assistance from IRI, we built a database of individual scanner panel data across two dozen categories, 12 of which have been processed for the purposes of this study. To increase the likelihood that categories in the database feature frequently in consumers’ baskets, a special effort was made to select high penetration, frequently purchased items. Some thought was also given to the consumer’s likely portfolio on a given shopping trip (e.g., dairy products, paper products, cleaning items, etc.). In selecting product categories we also relied upon previous research by Fader and Lodish (1990) that identified differences among product categories with respect to purchase frequency, category penetration, and promotional activity.

For the purposes of this study, we selected a subset of 12 categories and subcategories, taking care to ensure variety and that the selected items were among the more frequently purchased product categories. They are: Bacon, Butter, Ground Coffee, Saltine Crackers, Hot Dogs, Ice Cream, Liquid Detergent, Margarine, Paper Towels, Soft Drinks (Colas), Sugar, and Bathroom Tissue. Category statistics (number of unique brand indexes, total volume, and store sales for individual brands and sizes) are given in Appendix B.

Data Scope. The data are drawn from two separate metro markets (hereafter referred to as Market A and Market B) in a large U.S. city, and cover a two-year period from June 1991 to June 1993. Market A covers 494 households in an urban environment. We have 72,833 unique transactions that result in 68,418 shopping trips over a period of 104 weeks. Three supermarkets are tracked in this market, but 90% of the volume (in terms of dollars spent and trips taken) is accounted for by two stores. The stores are from three competing chains and no store explicitly advertises as an EDLP format. However, their actual pricing patterns suggest that one store is EDLP relative to the other two. Market B has 548 households in a suburban market; we have records of 88,945 transactions for a

Transaction records for different categories that occur in the same store on the same day for the same household are combined into the same shopping trip.
total of 81,106 unique shopping trips over the same 104-week period. Two of the five stores in this market explicitly advertise as operating an EDLP format. For the purposes of model calibration and validation, we separate the data into three periods: initialization (first six months), calibration (next year), and validation (last six months).

The two markets provide us with somewhat different conditions under which to test our theory. Perhaps the most straightforward test is in Market B, where two of the five stores explicitly advertise as EDLP. Thus, not only are the pricing patterns clearly different from the other stores (i.e., the average basket price and variability over time are lower), but also the differentiation on pricing format is clearly emphasized in store advertising. In Market A, none of the stores advertise as EDLP; however, one store does exhibit different pricing such that the average basket price is lower and the price variability is lower. From our theory, we expect households to exhibit different preferences for the three stores depending upon their shopping behavior. Thus, even in the absence of an explicit EDLP alternative, Market A constitutes a test of our theory from the perspective of three stores with different amounts of price variation in the environment.

### Store Pricing

Table 1 lists the total number of visits to each of the eight stores. It notes whether or not the stores advertise as EDLP and shows the average price (over the entire two-year period) for a comparable basket of items in each store. There are many different ways to define the “basket” for this comparison. We construct the basket using the largest share brand-size alternative in each product category. While an imperfect measure, it has the advantage of standardization across stores (i.e., each store’s basket contains the same set of SKUs), and the largest share SKUs are likely to be the most salient for the greatest number of shoppers. The purpose of computing these basket prices is simply to check first whether there are differences across stores, and second, whether the differences correspond to the advertised position of the store.

Confidentiality prevents us from identifying the stores by name, so we use the following codes. In Market A, “RE1” denotes the “relative EDLP” store; “RH1” and “RH2” denote the “relative HILO” stores. Even though RE1 does not advertise as an EDLP format, there is a clear relative difference in the basket price across the three stores. In Market B, the bona fide EDLP stores are “E1” and “E2”; “H1” is the HILO store and “HH1” and “HH2” are two higher-tier HILO stores from the same chain. In Market B, differences between the price formats are more pronounced—there are three “tiers” of pricing. E1 and E2 (the two explicitly EDLP stores) are in the lowest price tier, H1 is in the next tier, and HH1 and HH2 define the highest tier. For both markets, the basket price differences between stores from different price tiers are significantly different from zero.

### Other Panelist Information

Our dataset also contains information on important household-level demographics, including family size, family member ages, and income. These variables will be utilized when we seek to explain differences between large and small basket shoppers. We also obtained information on the households’ and stores’ five-digit zip codes and constructed proxy measures for the distance to each store for each of the 1,042 households in our dataset.

### 4. Model Specification and Measures

In this section, we complete the specification of our model and formalize the measures used in different...
components of the model. In §4.1, we formally develop a measure to describe the extent to which a customer exhibits the characteristics of a large basket shopper or a small basket shopper. In §4.2, we operationalize our measure of expected basket attractiveness such that the relative attractiveness of the basket in a given store (EDLP or HILO) depends upon the type of shopper. The measure is derived from a nested logit model of brand choice and purchase incidence, which is conditional upon shopper type. In §4.3, we specify a model of store choice behavior, in which the probability of store choice is a function of expected basket attractiveness. While our models are not complicated, there are many parameters and variables to track. Table 2 provides a glossary of notation, model variables, parameters, and the estimation steps.

4.1. Shopper Type
In §2, we articulated a rationale for differentiating among shoppers a priori on the basis of expected basket size. We begin by assuming that there are two different types of shoppers, large basket shoppers and small basket shoppers, each described by a different set of model parameters. Our goal is to come up with a single measure per household that will capture the propensity of that household to purchase small (or large) baskets when visiting the grocery store.

We take the following approach to estimate this propensity. We begin by assuming that the household’s distribution of grocery expenditures across trips is given by a lognormal distribution, i.e.,

\[ Y_i^h \sim \log \text{Normal} \left( \mu_y^h, \sigma_y^h \right), \]  

where \( Y_i^h \) represents the total basket expenditures by household \( h \) on store visit \( t \). We then estimate the parameters of the distribution \( (\mu_y^h, \sigma_y^h) \) by maximum likelihood (see Appendix), and estimate the propensity of the household to exhibit the characteristics of a small basket shopper by the proportion of trips in which the amount spent is less than some cutoff \( x \). Thus,

\[ p_{SBS}^h = \int_0^x f(Y_i^h) dY_i^h, \]  

where \( f \) is the lognormal distribution. The value of \( x \) in Equation (9) relates to the slope of the line in Figure 1: It is the expected basket size that divides those who are more likely to be small basket shoppers from those who are more likely to be large basket shoppers. For our data, we found that a value of \( x = 20 \) gave us the result that approximately half of the households in our sample had a value of \( p_{SBS}^h < 0.5 \). A cutoff of $20 is also consistent with the notion of a “small trip” as defined in recent research on “How Consumers Shop” by Progressive Grocer (1992). Figure 3, which plots shopping frequency versus grocery expenditures for small basket shoppers (i.e., those shoppers with \( p_{SBS}^h > 0.5 \)) and large basket shoppers suggests that we have achieved reasonably good discrimination (the correlation across households between \( p_{SBS}^h \) and expected basket size is \(-0.81\)).

Note that the choice of a single cutoff is not particularly restrictive in that we still obtain an index that varies continuously between 0 and 1. The greater each household’s departure from the cutoff (i.e., the extent to which the household’s average shopping trip is that much greater than or less than $20), the greater the certainty with which we can assign them to either of the two shopper types. Moving the cutoff merely shifts the average level of the index and does not strongly influence the results.11

4.2. Expected Basket Attractiveness
Because we have assumed that the consumer assesses the attractiveness of the basket before entering the store, our measure must be based on the price expectations of the consumer rather than on the prices prevailing in the store at the time of the actual visit.12 In the general case, if we let \( M \) denote the marketing variables in the model (i.e., price, feature, and display activity across all brands in all categories), then what we want to create is the following measure:

\[ \text{EBA}_{ij}^h = \int_M \sum_c \ln \left( 1 + \exp \left( V_{ij}^h - \xi_j(M) \right) \right) \xi_j(M) dM. \]  

Note that EBA is conditional on shopper type, \( j \in \{ \text{LBS, SBS} \} \). That means that we have a different measure for each shopper type in the model, constructed using the

11 We reestimated the model results with \( x = 30 \) and the substantive results were unchanged.

12 When building our model of store choice, however, we do control for the effect of feature advertising. For a more detailed analysis of advertised prices and store choice, see Bodapati (1996).
Table 2: Glossary of Notation, Variables, Parameters, and Estimation Steps

<table>
<thead>
<tr>
<th>Indexes, Data Periods, and Number of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
</tr>
<tr>
<td>$h$</td>
</tr>
<tr>
<td>$i$</td>
</tr>
<tr>
<td>$j$</td>
</tr>
<tr>
<td>$k$</td>
</tr>
<tr>
<td>$s$</td>
</tr>
<tr>
<td>$t$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Market A</th>
<th>Market B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initialization</td>
<td>17,127 20,708</td>
</tr>
<tr>
<td>Calibration</td>
<td>33,452 41,762</td>
</tr>
<tr>
<td>Validation</td>
<td>17,839 18,636</td>
</tr>
<tr>
<td>Panelists</td>
<td>494 548</td>
</tr>
</tbody>
</table>

Models, Parameters, and Estimation Steps

<table>
<thead>
<tr>
<th>Component</th>
<th>Variables</th>
<th>Parameters</th>
<th>Data Used</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Shopper Type</td>
<td>$Y_{ht}$</td>
<td>$\mu_h$, $\sigma_h^2$</td>
<td>Initialization</td>
<td>$\rho_{ht}^{\text{SBS}}$</td>
</tr>
<tr>
<td>2. Brand Choice</td>
<td>$BLOY_{hi}$, $SLOY_{ji}$</td>
<td>$\alpha_{hi}$, $\kappa_{hi}$, $\beta_{hi}$</td>
<td>Calibration</td>
<td>$\mathcal{O}_{ht}$</td>
</tr>
<tr>
<td></td>
<td>$LBP_{ht}$, $LSP_{ht}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\text{PRICE}_{cst}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\text{FEAT}_{cst}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\text{DISP}_{cst}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Purchase Incidence</td>
<td>$CR_{cjt}$, $INV_{cjt}$</td>
<td>$\gamma_{cjt}$</td>
<td>Calibration</td>
<td>$EBA_{cst}$</td>
</tr>
<tr>
<td>4. Store choice</td>
<td>$DIST_{st}$, $CONSID_{st}$</td>
<td>$\psi_{st}$</td>
<td>Calibration</td>
<td>$EBA_{st}$, $ADV_{st}$</td>
</tr>
</tbody>
</table>

Figure 3: Total Trips and Aggregate Expenditures by Shopper Type

- Large Basket Shoppers
- Small Basket Shoppers

4.3. Store Choice

Our model of store choice is built around our proposed measure of expected basket attractiveness, which in turn is based upon a calibrated nested logit model of brand choice and purchase incidence. Because these models are relatively standard, we present them first without much discussion. We then describe our model of store choice in somewhat more detail.

The key question we need to address is how the specification of brand choice and purchase incidence should differ across type of shopper. Because our theory is driven only by the differences across shoppers in purchase incidence, we make the following modeling decisions. For brand choice, we use the same set of model parameters to describe the purchase behavior of calibrated model parameters for that shopper type. The problem, of course, is that unlike our simple one-brand, two-price example in §2, the true joint distribution of marketing activity—denoted $g(M)$—is exceedingly complex. We therefore use a point-mass approximation of $g(M)$, using each visit by the household during the initialization period as an equally weighted observation on the joint distribution of $M$. If we let the subscript $w$ denote store visits made by the household during the initialization period, then:

$$EBA_{ht} = \frac{1}{n^h} \sum_w \sum_c \ln(1 + \exp(V_{cst1j}(M_w)))$$

where $M_w$ denotes the marketing activity in the store during visit $w$, and $n^h$ is the number of store trips made by household $h$ during the initialization period. In this formulation, we have assumed that the consumer is aware of the distribution of marketing activity in all stores, even though he/she may frequent one store more often (or even exclusively). Note that $EBA$ continues to be subscripted by time, because $V_{cst1j}$ is a function of variables other than marketing activity (e.g., the inventory level of the household) that vary over time.
both large and small basket shoppers. For purchase incidence, we use a model parameterization conditional on shopper type (in other words, we use one set of parameters for large basket shoppers and a different set for small basket shoppers).

**Brand Choice.** For brand choice the deterministic component of household \( h \)'s utility for brand \( i \) and size \( k \)\(^{13} \) in category \( c \) in store \( s \) on trip \( t \) is given by:

\[
U_{ikst}^h = \alpha_{hc} + \kappa_{kc} + \beta_{ikc}^h X_{ikst}^h. \tag{12}
\]

The elements of the matrix \( X_{ikst}^h \) are:

- \( \text{BLOY}^h_i = \text{loyalty of household } h \text{ to brand } i \),
- \( \text{LBP}^h_i = 1 \text{ if } i \text{ was the last brand purchased, } 0 \text{ otherwise} \),
- \( \text{SLOY}^h_k = \text{loyalty of household } h \text{ to size } k \),
- \( \text{LSP}^h_{it} = 1 \text{ if } k \text{ was the last size purchased, } 0 \text{ otherwise} \),
- \( \text{PRICE}_{ikst}^h = \text{actual shelf price of brand-size } i, k \text{ in store } s \text{ at time } t \),
- \( \text{FEAT}_{ikst} = 1 \text{ if brand-size } i, k \text{ appeared in a feature ad at trip } t, 0 \text{ otherwise, and} \)
- \( \text{DISP}_{ikst} = 1 \text{ if brand-size } i, k \text{ was specially displayed at trip } t, 0 \text{ otherwise.} \)

The loyalty variables are based on the share of the brand or size during the initialization period (Bucklin and Lattin 1992). The probability of brand choice is given by the familiar multinomial logit framework:

\[
p_{ik}^h(i, k) = \frac{\exp(U_{ikst}^h)}{\sum_{lk} \exp(U_{lkst}^h)}. \tag{13}
\]

**Purchase Incidence.** For purchase incidence the deterministic component of utility conditional on shopper type \( j \) is given by:

\[
V_{cstij}^h = \gamma_{0c} + \gamma_{1c} \text{CR}_c^h - \gamma_{2c} \text{INV}_c^h + \gamma_{3c} \text{CA}_c^h, \tag{14}
\]

for category \( c \) in store \( s \) at time \( t \), where:

- \( \text{CR}_c^h = \text{a measure of the household’s consumption rate in category } c \),
- \( \text{INV}_c^h = \text{a measure of household inventory at time } t \) in category \( c \), and

\(^{13}\text{In accounting for multiple brand-size combinations we adopt the parsimonious formulation developed by Fader and Hardie (1996).}\)

\( \text{CA}_c^h = \ln \sum_{ik} \exp(U_{ikst}^h) \), category \( c \)'s attractiveness.

The term \( \text{CA}_c^h \) is the expected maximum utility from brand choice; the variables \( \text{CR}_c^h \) and \( \text{INV}_c^h \) capture differences in purchase propensity across households over time.

The probability of purchase incidence, conditional on shopper type, is given by the binomial logit:

\[
p_{cst}^h(\text{inc} | j) = \frac{\exp(V_{cstij}^h \text{inc} | j)}{1 + \exp(V_{cstij}^h \text{inc} | j)}. \tag{15}
\]

Note that because \( V_{cstij}^h \) is a function of idiosyncratic household variables such as consumption rate and inventory, there is heterogeneity in \( p_{cst}^h(\text{inc} | j) \) within shopper type (i.e., not all small basket shoppers have the same probability of buying margarine on any given store visit). Following the theoretical development in §2, we expect that the parameters describing the category choice will result in a higher probability of purchase incidence for large basket shoppers than for small basket shoppers.

Finally, the unconditional probability of purchase incidence is given by:

\[
p_{cst}^h(\text{inc}) = p_{cst}^h(\text{inc} \mid \text{LBS}) \cdot p_{\text{LBS}}^h + p_{cst}^h(\text{inc} \mid \text{SBS}) \cdot p_{\text{SBS}}^h, \tag{16}
\]

where \( p_{\text{LBS}}^h \) and \( p_{\text{SBS}}^h \) are the probabilities determined a priori using the procedure described in §4.1.

**Store Choice.** Like purchase incidence, our model of store choice is conditional upon shopper type (since it is a function of EBA and EBA itself is a conditional measure). However, unlike our model of purchase incidence (where we required a different parameterization to differentiate the purchase propensity of large basket shoppers and small basket shoppers), we have no strong priors regarding any difference in store choice across shopper types other than that due to the effect of EBA. Therefore, to isolate the effect of the measure on store choice, we use the same set of model parameters to describe store choice for large and small basket shoppers. Thus, for store choice, the deterministic component of utility for store \( s \) on trip \( t \) conditional on shopper type \( j \) is given by:
Among stores visited by the consumer, we expect the choice share during the initialization period to reflect all effects of the EBA measure. Including store loyalty will not permit us to separate out the effect of EBA. In contrast, the model specification in Equation (17) enables us to assess the extent to which EBA can explain the variation in store choice among stores visited by each household that cannot be explained by location.

We also included the variable ADV to account for the possibility that shoppers become aware of specific prices or promotional specials through feature advertising by the store. The measure is weighted according to the importance of the category (CR) and the attractiveness of the advertised item (BLOY · SLOY) to the consumer.

As with purchase incidence, the probability of store choice conditional on shopper type is given by the logit model

\[ p_i(s) = \frac{\exp(W_{i|s})}{\sum_r \exp(W_{i|r})} \]  

and the unconditional probability of store choice by

\[ p_i(s) = p_i(s|\text{LBS}) \cdot p_{\text{LBS}} + p_i(s|\text{SBS}) \cdot p_{\text{SBS}}. \]

The parameters of all models are estimated sequentially using maximum likelihood. The likelihood functions are given in Appendix C.

5. Empirical Results
We first report the results for category purchase incidence. Second, we report the substantive results for the two store choice hypotheses. Third, we discuss an extension to our model in which we control for across-household differences in aggregate consumption. We find that not only does EBA have a significant effect on the probability of store choice in Market B, but also in Market A (where differences in price format are more subtle). Thus, Hypothesis 1 is supported. We also find (consistent with Proposition 2) that large basket shoppers have higher store-level elasticities. Finally, examination of store choice patterns shows that, as expected, EDLP stores get a greater share of visits from the large basket shoppers. Thus, Hypothesis 2 is also supported.
5.1. Category Purchase Incidence
Table 3 shows the fits for two different model formulations. The first (“one-shopper”) model is a standard purchase incidence model in which the parameters are estimated across all households. The second (“two-shopper”) model allows for the fact that households are either large or small basket shoppers, and the incidence parameters are estimated conditional upon these shopper types.

Note from Table 3 that the BIC-adjusted fits favor the two-shopper models, so we proceed to compute the purchase incidence elasticities using parameters from these models.

Table 4 presents some very interesting elasticity results from the “two-shopper” model. As shown in Table 4, the pattern of elasticities is consistent with the predictions of Proposition 1. That is, elasticities are smaller for large basket shoppers than for small basket shoppers: large basket shoppers are less responsive to prices in their category purchase incidence decisions.

5.2. Store Choice
5.2.1. Hypothesis 1: EBA Influences Store Choice. Table 5 provides the calibration and validation model fits for four store choice models, estimated using data from both markets. These four models are: Model 1, a benchmark model with store intercepts; Model 2, a second benchmark that controls for distance to the store and experience in the store; Model 3, which adds the effects of advertising and expected basket attractiveness unconditional on shopper type (i.e., the EBA measure is calculated using purchase incidence parameters from the one-shopper model); and Model 4, which has expected basket attractiveness conditional on shopper type (i.e., the EBA measure is calculated using purchase incidence parameters from the two-shopper model).

The improvement in fit from Model 2 to Model 3 shows the importance of including $EBA_{it}$: It shows that, unconditional on shopper type, there is an effect of expected basket attractiveness on store choice. The improvement from Model 3 to Model 4, where $EBA_{it}$ replaces $EBA_{it}$, shows that the evaluation and appeal of stores are different for different shopper types. Table 6 reports the parameter estimates from the model with the best in-sample and validation fits, Model 4.

The pattern of results is entirely consistent across both markets: Hypothesis 1 is strongly supported. All parameters have the expected signs and are significantly different from zero. The results are compelling: price expectations matter even after controlling for the influence of other marketing (e.g., feature advertising) and $80 M.
Table 5  Store Choice Model Fits

<table>
<thead>
<tr>
<th>Parameters</th>
<th>LL¹</th>
<th>BIC</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1 (Store Intercepts)</td>
<td>2</td>
<td>−31,979.0</td>
<td>−31,989.4</td>
</tr>
<tr>
<td>Model 2 (1 plus (DIST_1), (CONSID_1))</td>
<td>4</td>
<td>−22,267.0</td>
<td>−22,287.0</td>
</tr>
<tr>
<td>Model 3 (2 plus (EBP_{0e}), (ADV_{0e}))</td>
<td>6</td>
<td>−22,131.7</td>
<td>−22,163.0</td>
</tr>
<tr>
<td>Model 4 (2 plus (EBP_{0e}), (ADV_{0e}))</td>
<td>6</td>
<td>−21,980.1</td>
<td>−22,011.4</td>
</tr>
<tr>
<td>Market B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1 (Store Intercepts)</td>
<td>4</td>
<td>−63,704.0</td>
<td>−63,725.3</td>
</tr>
<tr>
<td>Model 2 (1 plus (DIST_1), (CONSID_1))</td>
<td>6</td>
<td>−26,367.6</td>
<td>−26,404.1</td>
</tr>
<tr>
<td>Model 3 (2 plus (EBP_{0e}), (ADV_{0e}))</td>
<td>8</td>
<td>−26,149.7</td>
<td>−26,191.4</td>
</tr>
<tr>
<td>Model 4 (2 plus (EBP_{0e}), (ADV_{0e}))</td>
<td>8</td>
<td>−26,074.7</td>
<td>−26,116.4</td>
</tr>
</tbody>
</table>

¹Calibration observations are 33,452 and 41,762 for Markets A and B, respectively.
²Validation observations are 17,839 and 18,686 for Markets A and B, respectively.

Table 6  Store Choice Model Parameters from Model 4

<table>
<thead>
<tr>
<th>Variable</th>
<th>Market A</th>
<th>Market B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter</td>
<td>t statistic</td>
</tr>
<tr>
<td>Store Intercepts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\phi_{0a}) (Store RE1)¹</td>
<td>0.000</td>
<td>—</td>
</tr>
<tr>
<td>(\phi_{0b}) (Store RH1)</td>
<td>0.158</td>
<td>5.880</td>
</tr>
<tr>
<td>(\phi_{0c}) (Store RH2)</td>
<td>−0.777</td>
<td>−28.470</td>
</tr>
<tr>
<td>(\phi_{0d}) (Store E1)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(\phi_{0e}) (Store E2)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(\phi_{0f}) (Store H1)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(\phi_{0g}) (Store HH1)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(\phi_{0h}) (Store HH2)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Control Variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\psi_1) ((DIST_1))</td>
<td>−0.013</td>
<td>−2.865</td>
</tr>
<tr>
<td>(\psi_2) ((CONSID_2))</td>
<td>3.561</td>
<td>75.563</td>
</tr>
<tr>
<td>Marketing Variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\psi_3) ((EBP_{0e}))</td>
<td>4.618</td>
<td>16.148</td>
</tr>
<tr>
<td>(\psi_4) ((ADV_{0e}))</td>
<td>0.334</td>
<td>10.545</td>
</tr>
</tbody>
</table>

¹Normalized to zero for identification.

nonmarketing (e.g., distance, previous experience at the store) variables.

5.2.2. HYPOTHESIS 2: LARGE BASKET SHOPPERS PREFER EDLP. Hypothesis 2 is linked to Hypothesis 1, but is perhaps more intriguing. Our theory predicts that large basket shoppers will prefer to shop at EDLP and small basket shoppers will prefer to shop at HILO. We test Hypothesis 2 by generating contingency tables of Store Price Format × Shopper Type and counting the number of calibration shopping trips that occur in each cell. First, we classify the individual shoppers as being either large basket shoppers or small basket shoppers. We do this using a simple rule: The household is a small basket shopper if \(p_{\text{Shins}} > 0.50\). Tables 7 and 8 present the contingency tables and associated \(\chi^2\) statistics for Markets A and B, respectively. The observed number of trips in each cell is given first, and the expected number of trips is given in parentheses. In Market A we see that store RE1 (the relative EDLP store) receives almost 1,000 more shopping trips from large basket shoppers (observed = 8,439; expected = 7,473) than one would expect by chance. Conversely, RH1 (the relative HILO store) receives over 1,000 more trips from the small basket shoppers (observed = 8,083; expected = 6,944) than one would expect by chance alone.¹⁵ In Market B with the bona fide EDLP stores (E1 and E2), the same pattern is repeated. There is overwhelming support for Hypothesis 2.

Recall that Proposition 2 implies that the store choice

¹⁵ Store RH2 closed for 12 weeks and later reopened so it has a very small market share. Nevertheless, the results for this store are also consistent with Hypothesis 2.
Table 7 Number of Visits by Type of Shopper (Market A)\(^1\)

<table>
<thead>
<tr>
<th>Store</th>
<th>Small Basket</th>
<th>Large Basket</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>RE1</td>
<td>9,028(^2)</td>
<td>8,439</td>
<td>17,467</td>
</tr>
<tr>
<td>(9,994)(^3)</td>
<td>(7,473)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RH1</td>
<td>8,083</td>
<td>4,052</td>
<td>12,135</td>
</tr>
<tr>
<td>(6,944) (^3)</td>
<td>(5,191)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RH2</td>
<td>2,030</td>
<td>1,820</td>
<td>3,850</td>
</tr>
<tr>
<td>(2,203) (^3)</td>
<td>(1,647)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>19,141</td>
<td>14,311</td>
<td>33,452</td>
</tr>
</tbody>
</table>

\(^1\)Observed number of trips.
\(^2\)Expected number of trips.

\(^3\)\(\chi^2 = 687.3, p < 0.001\).

Table 8 Number of Visits by Type of Shopper (Market B)\(^1\)

<table>
<thead>
<tr>
<th>Store</th>
<th>Small Basket</th>
<th>Large Basket</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>3,156(^2)</td>
<td>4,121</td>
<td>7,277</td>
</tr>
<tr>
<td>(4,571)(^3)</td>
<td>(2,706)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2</td>
<td>3,881</td>
<td>5,304</td>
<td>9,185</td>
</tr>
<tr>
<td>(5,769) (^3)</td>
<td>(3,416)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H1</td>
<td>11,626</td>
<td>3,155</td>
<td>14,781</td>
</tr>
<tr>
<td>(9,284) (^3)</td>
<td>(5,497)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HH1</td>
<td>4,604</td>
<td>1,252</td>
<td>5,856</td>
</tr>
<tr>
<td>(3,678) (^3)</td>
<td>(2,178)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HH2</td>
<td>2,964</td>
<td>1,699</td>
<td>4,663</td>
</tr>
<tr>
<td>(2,929) (^3)</td>
<td>(1,734)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>26,231</td>
<td>15,531</td>
<td>41,762</td>
</tr>
</tbody>
</table>

\(^1\)\(\chi^2 = 5055.4, p < 0.001\).
\(^2\)Observed Number of Triips.
\(^3\)Expected Number of Triips.

5.3. Controlling for Differences in Shopping Frequency

In formulating our a priori measure of shopper type, we attempted to build an index that could discriminate between shoppers on the basis of expected basket size. We show that an increase in the expected basket size per trip leads to increased preference for EDLP (Proposition 3). However, as we noted in the Introduction, it is also important to control for residual differences in shopping frequency.

Returning to Figure 2, we note that shoppers A, B, and C all have the same expected basket size, but differ in their shopping frequency. As we noted in §2, we expect household C to have the highest relative preference for HILO. The intuition is simple and complements our theory. All else equal, additional trips to stores with price variation give the shopper some benefit, due the probability associated with finding items on deal (at price \(L\)). To control for the effect of across-household differences in shopping frequency (and therefore differences in aggregate consumption), we add an additional variable, SFREQ\(^h\) (shopping trips per week) to the store choice model of Equation (17):

\[
W_{dij} = \phi_s + \psi_1 \text{DIST}_{s}^h + \psi_2 \text{CONSID}_{s}^h + \psi_3 \text{EBA}_{dij}^h + \psi_4 \text{ADV}_{s}^h + \theta_{r(s)} \text{SFREQ}^h,
\]

where \(\theta_{r(s)}\) is a parameter for the price format \(r\) prevailing in store \(s\). In Market A where there are two price formats we estimate one parameter; in Market B where there are three price formats we estimate two.
parameters. We estimate the $\theta_{\phi(s)}$ parameters using the standard procedure for variables that vary across individuals, but not across choice alternatives (see Ben-Akiva and Lerman 1985, Greene 1996). We expect that, all else equal, more frequent shoppers prefer HILO. That is, we expect $\theta_{\text{HILO}}(s) > 0$, $\theta_{\text{EDLP}}(s) < 0$.

In both markets, adding $\text{SFREQ}^h$ improves the model, but leaves the other parameters almost unchanged and still highly significant. (The BIC and validation values are $-21,942.2$ and $-11,431.2$, and $-26,015.0$ and $-12,385.1$ for Markets A and B, respectively.) Furthermore, the $\theta_{\phi(s)}$ parameters are highly significant. In Table 9, we report the format choice probabilities and the marginal effects of $\text{SFREQ}^h$ on format choice. The table shows that choice probabilities for small basket shoppers favor HILO; choice probabilities for large basket shoppers favor EDLP. Furthermore, as expected, households who shop more frequently prefer HILO and those who shop less frequently prefer EDLP.

### 6. Discussion

#### 6.1. Implications for Retailers

In this paper, we have differentiated among shoppers on the basis of their “expected basket size” on a given trip. Furthermore, we have shown that this delineation of shopper types is a very important predictor of store choice. While the retailer may be interested to know that large basket shoppers will prefer EDLP (and why this is so), more information is needed for this finding to be managerially actionable. To understand who the different shoppers are, we estimated the following regression model:

$$p_{\text{SBS}} = \alpha + \beta_1 \text{SIZE}^h + \beta_2 \text{INCOME}^h + \beta_3 \text{AGE}^h,$$

(21)

where $\text{SIZE}^h$, $\text{INCOME}^h$, and $\text{AGE}^h$ are the family size, annual income, and age of the head of the household, respectively. The results are shown in Table 10. We see that small basket shoppers are older, have smaller incomes, and smaller families. The value of thinking about segmenting consumers by basket size when trying to understand preferences for EDLP and HILO is highlighted in a recent *Wall Street Journal* article. In describing the problems experienced by KKR after their $1$ billion buyout of Bruno’s, the Birmingham, Alabama, supermarket chain, the article notes: “In the summer of 1996 Bruno’s management tried to edge up prices in their lower-priced (EDLP) stores with the intent of offering frequent shopper programs to offset the price increase. The company’s price-conscious customers, used to shopping for a fixed basket of goods, stayed away in droves.”

When we decompose the EBA variable (recall that
EBA_{htij} = \sum \ln(1 + \exp(V_{htij}))

on a category by category basis, we observe systematic differences in "focal categories" for the two types of shoppers. For small basket shoppers a large proportion of EBA is driven by products like ice cream and bacon; for large basket shoppers, bathroom tissue and paper towels are more prominent. This suggests that retailers who seek to appeal to large basket shoppers should pay special attention (through pricing and advertising) to staples.

Finally, many of the households in our sample shop predominantly at one store. If households are effectively constrained to shop at a single store, then our theory predicts that at the margin they should tailor their behavior to the pricing format of this store. Specifically, EDLP loyals are better off buying large baskets and shopping less frequently, while HILO loyals are better off buying smaller baskets and shopping more frequently. To investigate this issue, we take those households who exhibit a high degree of loyalty to one store (178 households in A, 349 households in B) and estimate a regression model using expected basket size as the dependent variable. Our regression controls for the influence of other variables:

\[ EBSIZE_{h} = \alpha + \beta_{1} SIZE_{h} + \beta_{2} INCOME_{h} + \beta_{3} DIST_{h} + \beta_{4} EDLPLOY_{h}, \]

where SIZE\(_{h}\), INCOME\(_{h}\), and DIST\(_{h}\) are the family size, income, and distance to the store, respectively. EDLPLOY\(_{h}\) is a dummy variable (EDLPLOY\(_{h}\) = 1 for EDLP loyals, EDLPLOY\(_{h}\) = 0 for HILO loyals). Our theory suggests we should find \(\beta_{4} > 0\) (Table 11).

Note that in both markets EDLP-loyal shoppers have a larger expected basket size. Thus, there is some indirect support of our theory even among those shoppers who are effectively constrained to shop at one store.

6.2. Summary

This paper had one key objective: to explain and demonstrate the relationship between preference for store price format under price uncertainty and consumer shopping behavior. This objective led to two testable hypotheses. (1) When heterogeneity in consumer shopping behavior and retail price format exists, expected basket attractiveness will influence store choice. (2) The relationship between expected basket attractiveness and store choice is such that large basket shoppers prefer EDLP stores and small basket shoppers prefer HILO stores. We presented a simple theory of shopping behavior and developed three propositions to show why the two hypotheses should hold. The hypotheses were tested using multicategory scanner panel data from purchase records for 1,042 panelists in two metro markets. Both were strongly supported.

Limitations and Extensions. The intuition offered in support of our theory is based on a stylized model of shopping behavior that departs from the real world in some respects. For example, we make an assumption that shoppers are aware only of the price distributions before entering the store, but not exactly when certain product categories are available on special. In fact, stores use feature advertising to make customers aware of when discount prices are available. While it would be possible to allow for price signaling through feature advertising, we did not do so in order to preserve the relative parsimony of the theory. In fact, research by Bodapati (1996) suggests that a fairly small proportion of shoppers (on the order of 20%) is influenced by feature advertising in their choice of stores; furthermore, only a fraction of the categories carried by the store are featured in any given week. Note that our empirical test controls for feature advertising (via \(ADV_{htij}^{st}\)) and the results are consistent with Bodapati (1996).

We also recognize that the empirical test is based upon an incomplete dataset and our estimation methods are sequential. However, we are encouraged that

<table>
<thead>
<tr>
<th>Variable</th>
<th>Market A</th>
<th>Parameter</th>
<th>t Statistic</th>
<th>Market B</th>
<th>Parameter</th>
<th>t Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>(INTERCEPT)</td>
<td>14.604</td>
<td>(3.044)</td>
<td>9.638</td>
<td>(2.719)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_{1}(SIZE))</td>
<td>1.860</td>
<td>(1.470)</td>
<td>5.653</td>
<td>(4.807)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_{2}(INCOME))</td>
<td>2.184</td>
<td>(3.271)</td>
<td>2.271</td>
<td>(5.135)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_{3}(DIST))</td>
<td>-0.339</td>
<td>(-0.289)</td>
<td>-0.598</td>
<td>(-0.641)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_{4}(EDLPLOY))</td>
<td>14.831</td>
<td>(3.305)</td>
<td>11.532</td>
<td>(3.227)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted (R^{2})</td>
<td>0.197</td>
<td></td>
<td>0.230</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
we are able to substantiate our consumer behavior theory using only 12 product categories.

**Future Research.** Our findings suggest many avenues for future work on store choice and shopping behavior. For example, do consumers “self-select” certain store price formats, modify their shopping behavior to take best advantage of the store’s pricing policy, or both? The spirit of our results (that EDLP and HILO formats allow for segmentation in a market where consumers self-select) is consistent with the equilibrium results in Lal and Rao (1997). In particular, we focus on the behavior of consumers and show that segmentation by expected basket size allows one to predict preference for store price formats. Further examination of this issue is a promising area for future work. As to the second issue, Ho, Tang, and Bell (1997) show that rational cost-minimizing consumers will tailor their shopping behavior to the pricing policy of the store. Rational EDLP shoppers visit the store less frequently and buy larger quantities per visit; rational HILO shoppers visit the store more frequently and buy smaller quantities per visit. Our analysis in §6.1 is consistent with their findings. Our analysis of the constituent parts of expected basket attractiveness suggests that some categories are more salient than others in determining the price image of the store. The relationship between store price image and category-level pricing is an important area for future research.

Finally, our results seem to suggest a higher level of in-store responsiveness among the small basket shoppers. A parallel finding in cross-category brand choice appears in the work of Ainslie and Rossi (1997). They find that more frequent shoppers are more price sensitive in brand choice. How do we reconcile these findings with the notion of “fill-in” trips (i.e., a special trip to the store to pick up one or two items that have run out at home), a small-basket trip where we expect the shopper to be less responsive to marketing activity? One possibility is that all shoppers engage in some need-driven shopping behavior, but that the proportion of need-driven purchases is no greater for small-basket shoppers than large-basket shoppers. This would imply that not all small-basket trips are fill-in trips; our evidence suggests that there must be a fair amount of opportunistic purchase behavior going on among small-basket shoppers. Attempting to capture these trip-specific differences (something our model does not do) appears to be a potentially fruitful avenue for future research.17

17 We thank Lee Cooper, Xavier Dreze, Dave Montgomery, Don Morrison, Catarina Sismeiro, and especially Randy Bucklin for their comments and the Marketing Science Institute for generous funding through the Alden G. Clayton Doctoral Dissertation Award. The comments and advice of Rajiv Lal and V. Padmanabhan are greatly appreciated. We also appreciate the help of Douglas Honnold, IRI, for authorizing the creation of the market basket database and that of Tara Merrill from IRI for her assistance in getting access to the data. Finally, we are extremely grateful to the editor, area editor, and two anonymous reviewers for their valuable comments on earlier versions of this paper.

**Appendix A. Proofs**

**Proof of Proposition 2.** We derive the store choice probability elasticity with respect to changes in expected basket attractiveness (EBA). First, some notation. Let $P(s)$ be the probability that store $s$ is chosen, and $\psi$ be the response parameter in store utility for EBA. $P_p(inc)$ is the purchase incidence probability for category $c$ and $\gamma$ is the price response parameter in the incidence utility. Given the expression for EBA in Equations (6) and (7) it is easy to see (Ben-Akiva and Lerman 1985) that the elasticity with respect a price change in a single product category is $\left[1 - P(s)\right] \cdot \psi \cdot P_p(inc) \cdot \gamma \cdot E$ for the EDLP store and $\left[1 - P(s)\right] \cdot \psi \cdot P_p(inc) \cdot \gamma \cdot \left[nL + (1 - n)H\right]$ for the HILO store. Given an assumption that category purchase decisions are independent, the elasticity with respect to price changes in the entire basket is simply the summation of this category elasticity, over all categories. Since $V_{iBS} > V_{ABS}$ implies $\rho(inc|LBS) > \rho(inc|SBS)$, it must follow that large basket shoppers have greater elasticities with respect to basket prices. Q.E.D.

**Proof of Proposition 3.** Let $V$ denote the deterministic utility the shopper obtains from category $c$, net of the price. We begin with a set of prices $L$, $H$, and $E$ and a deal probability $\pi$. By setting Equations (6) and (7) equal to each other, we can find the shopper type with $V = V^*$ who is indifferent between the EDLP and HILO formats. That is, the shopper type for whom the stores offer the same expected basket attractiveness. Specifically, we have $V^*$ such that:

$$
\pi \cdot \ln[1 + \exp(V^* - \gamma L)] + (1 - \pi) \cdot \ln[1 + \exp(V^* - \gamma H)] = \ln[1 + \exp(V^* - \gamma E)].
$$

(A.1)

Taking the exponent of both sides:

$$
[1 + \exp(V^* - \gamma L)]^{[1 + \exp(V^* - \gamma H)]^{1 - \pi}} = [1 + \exp(V^* - \gamma E)].
$$

(A.2)

We now investigate what happens for a household with $V > V^*$. We will show that the slope of $EBA(EDLP)$ is greater than the slope of $EBA(HILO)$ by taking derivatives and evaluating at the point $V^*$. For the HILO store we have:
Shopping Behavior and Consumer Preference for Store Price Format

\[ \frac{\partial \text{EBA(HILO)}}{\partial V} = \pi \frac{\exp(V - \gamma L)}{1 + \exp(V - \gamma L)} + (1 - \pi) \frac{\exp(V - \gamma H)}{1 + \exp(V - \gamma H)}, \]  
\[ \text{and for the EDLP store we have:} \]
\[ \frac{\partial \text{EBA(EDLP)}}{\partial V} = \frac{\exp(V - \gamma E)}{1 + \exp(V - \gamma E)} = 1 - \frac{1}{1 + \exp(V - \gamma E)}. \]  
\[ \text{Evaluating at } V^*, \text{ we know that (A.2) must hold. Thus, we can substitute} \]
\[ \frac{1}{1 + \exp(V - \gamma E)} \]
\[ \text{in (A.4) to give:} \]
\[ \frac{\partial \text{EBA(EDLP)}}{V} \bigg|_{V^*} = 1 - \left[ \frac{1}{1 + \exp(V^* - \gamma L)} \right]^\pi \left[ \frac{1}{1 + \exp(V - \gamma H)} \right]^{1 - \pi}, \]  
where the term being subtracted is the geometric mean. Similarly, we can rewrite (A.3) as:
\[ \frac{\partial \text{EBA(HILO)}}{V} \bigg|_{V^*} = 1 - \left[ \pi \frac{1}{1 + \exp(V^* - \gamma L)} + (1 - \pi) \frac{1}{1 + \exp(V^* - \gamma H)} \right], \]  
where the term being subtracted is the arithmetic mean (both the geometric and arithmetic means are positive and in the interval (0, 1)). Since the geometric mean is less than the arithmetic mean, we have the result that for any shopper with \( V > V^* \), \( \text{EBA(EDLP)} > \text{EBA(HILO)}. \)

Now, it remains to be shown that given \( \pi, L, \) and \( H \) there is a range of values of \( E \) for which \( V^* \) in Equation (A.1) always exists. We prove this by taking limits in Equation (A.1). Specifically:
\[ \lim_{V^* \to \infty} \pi \cdot \ln[1 + \exp(V^* - \gamma L)] + (1 - \pi) \cdot \ln[1 + \exp(V^* - \gamma H)] = \ln[1 + \exp(V^* - \gamma E)] \]  
\[ \text{becomes} \]
\[ \pi(V^* - \gamma L) + (1 - \pi)(V^* - \gamma H) = (V^* - \gamma E). \]  
Solving for \( E \) yields \( E = \pi L + (1 - \pi)H \). This result is intuitive—since the shopper buys with certainty in either store \( (V^* \to \infty) \), the EDLP store can charge a constant price equal to the average price in the HILO store and keep the shopper indifferent. Note also that the proportion of times the shopper buys at \( L \) is equal to \( \pi \). For the case of \( \lim_{V^* \to \infty} \) both sides of (A.1) go to 0. However, using Bayes’ Theorem we can find the relative proportion of times on which the shopper buys at price \( L \):
\[ \pi' = \frac{p(L \mid \text{inc})}{p(L \mid \text{inc}) + (1 - \pi) \cdot p(L \mid \text{dec})}, \]  
which simplifies to
\[ \pi' = \frac{\pi}{\pi + (1 - \pi) \cdot \exp(L - H) > \pi}, \]  
because \( \exp(L - H) < 1 \). This result is also very intuitive: the proportion of times, \( \pi' \), on which a (small basket) shopper buys at \( L \) is greater than for the large basket shopper. Thus, for any \( L, H, \) and \( \pi \) there is always an \( E \in (\pi L + (1 - \pi)H, \pi L + (1 - \pi)H) \) such that Proposition 3 holds. Q.E.D.

Appendix B. Summary of Product Categories

Table 12 provides the complete breakdown of sales by store for brand-size indices in 12 categories in both Market A and Market B. In each category, we construct brand-size indices so as to capture a minimum of 70% of the category volume (this number is a conservative estimate of the cutoff employed in published studies that utilize scanner panel data).

For each category we list the number of brands and sizes and total sales for the brand-size index. The indices were constructed by aggregation of brands and sizes across attributes such as flavor (where marketing mix activity does not vary substantially from UPC to UPC). To follow are the categories and subcategories: Subcategories (e.g., ground coffee) are created with reference to the original stub files and the IRI Marketing Factbook. The full list of brand-size indices for each category is available from the authors.

Appendix C. Likelihood Functions

Shopper Types

The formation of the likelihood is at the household level. The log likelihood function for each household is:
\[ \sum_{t=1}^{T} \log \left[ F(\mu_t, \sigma_t^2) \right], \]  
where \( T \) indexes the total initialization trips made by each household, and \( F(\mu_t, \sigma_t^2) \) is:
\[ \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp \left( -\frac{1}{2\sigma_t^2} \left( \log Y_t^s - \mu_t \right)^2 \right). \]  
\( Y_t^s \) denotes the spending of the household at shopping occasion \( t \).
Table 12 Summary Table for Market A and Market B

<table>
<thead>
<tr>
<th>Category</th>
<th>Market A</th>
<th></th>
<th></th>
<th>Market B</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Number</td>
<td>Number</td>
<td>Total</td>
<td>% Cat.</td>
<td>Total</td>
</tr>
<tr>
<td>Bacon</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>5,270</td>
<td>81.93</td>
<td>4,730</td>
</tr>
<tr>
<td>Butter</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>3,962</td>
<td>94.67</td>
<td>3,994</td>
</tr>
<tr>
<td>Margarine</td>
<td>11</td>
<td>1</td>
<td>11</td>
<td>11,396</td>
<td>92.33</td>
<td>11,381</td>
</tr>
<tr>
<td>Ice Cream</td>
<td>11</td>
<td>1</td>
<td>11</td>
<td>8,667</td>
<td>76.88</td>
<td>8,509</td>
</tr>
<tr>
<td>(Soda) Crackers</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>1,950</td>
<td>84.72</td>
<td>2,460</td>
</tr>
<tr>
<td>(Liquid) Detergent</td>
<td>10</td>
<td>4</td>
<td>29</td>
<td>9,087</td>
<td>74.88</td>
<td>5,087</td>
</tr>
<tr>
<td>(Ground) Coffee</td>
<td>7</td>
<td>4</td>
<td>20</td>
<td>4,605</td>
<td>80.43</td>
<td>4,987</td>
</tr>
<tr>
<td>Hod Dogs (Franks)</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>5,188</td>
<td>82.89</td>
<td>5,495</td>
</tr>
<tr>
<td>Soft Drinks (Cola)</td>
<td>4</td>
<td>5</td>
<td>14</td>
<td>9,377</td>
<td>77.46</td>
<td>16,526</td>
</tr>
<tr>
<td>Granulated Sugar</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>3,150</td>
<td>90.73</td>
<td>3,681</td>
</tr>
<tr>
<td>Tissue</td>
<td>10</td>
<td>4</td>
<td>25</td>
<td>16,899</td>
<td>95.66</td>
<td>25,693</td>
</tr>
<tr>
<td>Paper towels</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>12,854</td>
<td>82.57</td>
<td>19,744</td>
</tr>
</tbody>
</table>

\[\text{Brand Choice}\]

The log-likelihood function for the brand choice models in each product category is simply:

\[ LL = \sum_{h} \sum_{c} \sum_{i} \sum_{k} \delta_{ik} \cdot \ln [p_{ik}(i, k)], \quad (C.3) \]

where \( \delta_{ik} \) is 1 if brand-size \( i, k \) in category \( c \) is chosen by household \( h \) at visit \( t \) and 0 otherwise.

\[\text{Purchase Incidence}\]

For purchase incidence we have:

\[ LL = \sum_{h} \sum_{c} \sum_{i} \sum_{t} \delta_{ht} \cdot \ln[p_{ht}(inc)] + (1 - \delta_{ht}) \cdot \ln(1 - p_{ht}(inc)), \quad (C.4) \]

where \( \delta_{ht} \) equals 1 if household \( h \) purchased category \( c \) at shopping trip \( t \) and 0 otherwise. Note that even though the category utility functions are independent, the joint occurrence of products in the basket is reflected in the formulation of the likelihood function. The parameter estimates from this purchase incidence model are then employed to generate \( \text{EBA}_{hs} \) and \( \text{EBA}_{xij} \) values for each household.

\[\text{Store Choice}\]

The log-likelihood function is:

\[ LL = \sum_{h} \sum_{s} \delta_{hs} \cdot \ln[p_{hs}(s)], \quad (C.5) \]

where \( \delta_{hs} \) is 1 if store \( s \) is chosen by household \( h \) at time \( t \) and 0 otherwise.

References


This paper was received May 23, 1996, and has been with the authors 12 months for 2 revisions; processed by Russell S. Winer.