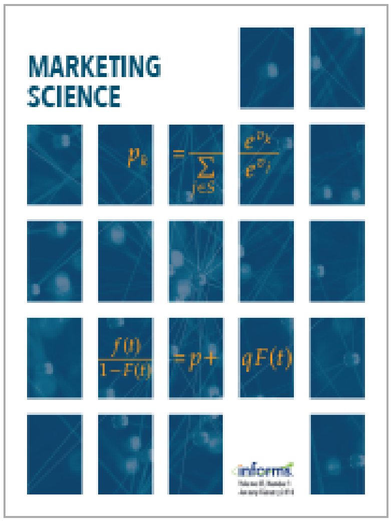


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Market Entry and Consumer Behavior: An Investigation of a Wal-Mart Supercenter

Vishal P. Singh, Karsten T. Hansen, Robert C. Blattberg,

MARKETING SCIENCE
Vol. 25, No. 5, September–October 2006, pp. 457–476
ISSN 0732-2399 | EISSN 1526-548X | 06 | 2505 | 0457

DOI 10.1287/mksc.1050.0176

Market Entry and Consumer Behavior: An Investigation of a Wal-Mart Supercenter

Vishal P. Singh
Tepper School of Business, Carnegie Mellon University, 244 Posner Hall, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213, vsingh@andrew.cmu.edu

Karsten T. Hansen, Robert C. Blattberg
Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208, {karsten-hansen@northwestern.edu, r-blattberg@northwestern.edu}

This paper provides an empirical study of entry by a Wal-Mart supercenter into a local market. Using a unique frequent-shopper database that records transactions for over 10,000 customers, we study the impact of Wal-Mart's entry on consumer purchase behavior. We develop a joint model of interpurchase time and basket size to study the impact of competitor entry on two key household decisions: store visits and in-store expenditures. The model also allows for consumer heterogeneity due to observed and unobserved factors. Results show that the incumbent supermarket lost 17% volume—amounting to a quarter million dollars in monthly revenue—following Wal-Mart's entry. Decomposing the lost sales into components attributed to store visits and in-store expenditures, we find that the majority of these losses were due to fewer store visits with a much smaller impact attributed to basket size. We also find that Wal-Mart lures some of the incumbent's best customers, and that retention of a small number of households can significantly reduce losses at the focal store. Finally, certain observed household characteristics such as distance to store, shopping behavior, and product purchase behavior are found to be useful in profiling the defectors to Wal-Mart. Implications and strategies for supermarket managers to compete with Wal-Mart are discussed.

Key words: entry; retail competition; Wal-Mart supercenter; frequent-shopper data
History: This paper was received March 18, 2004, and was with the authors 9 months for 2 revisions; processed by Michel Wedel.

> The time has come, as everyone knew it would. Wal-Mart, which through its four formats had already been selling more groceries than anyone in America, is now the country's biggest supermarket operator . . . racking up $4 billion more in annual sales than former top dog Kroger. (*Progressive Grocer*, May 2003)

1. Introduction

The role of supermarkets in the grocery retailing industry has undergone dramatic changes over the last decade. Rapid growth of alternative retail formats, in the form of mass discounters, wholesale clubs, and supercenters, has transformed not only the competitive structure of the industry, but also the way in which consumers shop. The biggest threat to the supermarket industry comes from none other than the world's largest retailer: Wal-Mart. In spite of being a relatively new player, Wal-Mart, through its supercenter format, has already become number one in the grocery industry. Patterned after the European hypermarket, a supercenter combines a full-line discount store with a full-line supermarket under one roof. These stores carry both general merchandise and food, including groceries and perishables. They also offer a variety of ancillary services such as pharmacy, dry cleaning, hair salon, and photo development services; and gas stations, providing consumers with a true one-stop shopping experience. For an industry already crowded with many players, there are various reasons why Wal-Mart's supercenter format poses an extraordinary challenge. As discussed in §2, Wal-Mart has been able to keep its costs below the industry level, which in turn translates into lower prices for the consumers. Given the razor-thin margins in the grocery industry, Wal-Mart's everyday low prices are difficult, if not impossible, to match. Indeed, as quoted in the *Wall Street Journal* (2003), items at Wal-Mart cost 8%–27% less than at Kroger, Albertsons, or Safeway, even after taking into account discounts from these competitors' loyalty cards and specials. Besides costs, another factor driving the grocery prices down at the supercenter has to do with the main motivation for why Wal-Mart and other discount stores entered the grocery business in the first place: store traffic. A typical supercenter has only 30% of the area devoted to grocery. According to industry analysts, Wal-Mart offers lower prices on food in order to bring traffic into the supercenters with the hope of selling higher-margin general merchandise, and even has the potential of treating the entire food business as a loss leader. National chains and independents alike are feeling the pressure from Wal-Mart. In the past decade, 29 chains have sought bankruptcy-court protection, with Wal-Mart as a catalyst in 25 of those cases (*Wall Street Journal* 2003). Not surprisingly, Wal-Mart supercenter is seen as a serious menace to the traditional grocery industry, with 80% of supermarket managers citing competition from supercenters as their biggest concern in the coming year (National Grocers Association 2003).

Despite their unprecedented growth and the threat they pose to the traditional grocery industry, relatively little is known about how entry of a supercenter in a market changes consumer purchase behavior or what it does to the bottom line of an incumbent supermarket. Although there have been a number of business press articles covering this new retail format, they provide little information on the issue. Instead, the commentary has ranged from predictions on extinction of traditional grocery to general guidelines on how to compete with this new format. Academic research, on the other hand, has primarily focused on stores that are similar in terms of their product offerings and cost structures (Lal and Matutes 1989, Pesendorfer 2002), or supermarkets that differ only in terms of their pricing formats, that is, every day low pricing (EDLP) versus Hi-Lo (Bell and Lattin 1998, Lal and Rao 1997, Messinger and Narasimhan 1997). With minor exceptions (Fox et al. 2004, Singh 2002), there is limited attention given to alternative retail formats, such as mass merchandisers or supercenters.

This paper provides an empirical study of the impact of a Wal-Mart supercenter entry on sales of a traditional supermarket. We utilize a unique frequent-shopper database that records purchases for over 10,000 households before and after Wal-Mart's entry. The data are drawn from a store located in a small town on the East Coast. The store in question has a well-developed frequent-shopper program, with over 85% of the sales captured on shopper cards. The database records all transactions made in the store, and captures such information as time and date of the transaction, price, promotion, and quantity for every UPC sold. This information was recorded at the individual level for all the customers in the store for a period of 20 months, from November 1999 to June 2001. In August of 2000, a Wal-Mart supercenter entered 2.1 miles from the store. Thus, we observe a reasonably long purchase history both before and after Wal-Mart's entry.

Our primary focus in this paper is on analyzing changes in consumer purchase behavior following the competitor's entry. Entry of a discount store in the market can influence a household's buying behavior in several ways. At the two extremes, some consumers may not change their purchase behavior at all, while others may completely abandon the incumbent and defect to Wal-Mart. Other consumers may shift part of their purchases to Wal-Mart while continuing to patronize the incumbent store. For this group, the lost volume can come from three sources: fewer store visits, smaller baskets, or a combination of the two. Furthermore, these changes in household behavior could be related to factors such as distance to the incumbent and Wal-Mart, household demographics, and other characteristics related to shopping behavior. The primary questions addressed in this paper are

- What is the impact of Wal-Mart's entry on the incumbent supermarket's total sales?
- To what extent are the total observed losses attributed to customer attrition, reduction in store visit frequency, and smaller basket size?
- What are the observed demographic and purchase behavior characteristics of the households that defect to Wal-Mart?

Answers to these questions can be quite important from a managerial perspective. For instance, decomposition of total sales into components attributed to store visits and basket size can be useful in understanding the source of lost volume and in developing store-level marketing policies. Suppose, for example, we find that the lost sales are primarily due to households not visiting the store as frequently as they did prior to Wal-Mart's entry but that, once the shopper is in the store, basket size remains constant. This in turn suggests the need for developing strategies that are primarily geared toward generating store traffic, such as use of deep promotions and feature advertisements. On the other hand, suppose we find that the frequency of store visits remains constant but that the basket size is smaller. In this case, the focus should be on in-store merchandising to increase expenditure once the customers are at the store. Similarly, identifying households based on their observed characteristics can also be quite important because it can allow the retailer not only to target customers with similar characteristics at this store, but also to transfer the findings to other store locations where the retailer comes in competition with Wal-Mart (or other such formats). Store opening information is generally available well in advance, and so preemptive actions can be taken for the households who are at high risk of defection.

To evaluate Wal-Mart's impact on consumer purchase behavior, we develop a joint model of interpurchase time and basket size. Although a popular approach to model interpurchase time used in the marketing literature is the proportional hazard model in continuous time (e.g., Jain and Vilcassim 1991), it has the limitation of only accounting for marketing mix and other covariates when an event occurs (e.g., when a purchase is made). On the other hand, a discrete-time approach (Gupta 1991, Wedel et al. 1995) can explicitly account for the covariates in periods during which households do not make a purchase. We take a discrete-time approach and model the household store visit decision using a discrete-choice framework with time-varying coefficients. These time-varying coefficients capture the duration dependence embodied in consumers' choice process. The model, based on an underlying utility-maximizing framework, can be interpreted as a hazard model (Seetharaman and Chintagunta 2003). Besides accounting for the full time path of the covariates, this modeling approach has the advantage of allowing for nonproportional hazards—a feature that is empirically relevant for our data.

Consumers' in-store expenditures are modeled using a semilog specification that has been used extensively in marketing (e.g., Blattberg and Neslin 1990). Both of these household decisions (store visit and in-store expenditure) are modelled jointly, and Wal-Mart's impact on these decisions are captured by allowing for a structural break at the time of competitor entry. The model also allows for consumer heterogeneity, modelled using a hierarchical structure. In particular, the full set of model parameters is allowed to vary across consumers due to both observed (e.g., demographic) and unobserved factors. For inference we use a hierarchical Bayesian approach that, as discussed in Allenby and Rossi (1999), is well suited to making inference at the individual level.

Our results show that the incumbent store lost 17% of its volume—amounting to a quarter million dollars in monthly revenue—following Wal-Mart's entry. The magnitude of the lost sales is quite alarming considering that supermarkets generally operate on a principle of low margins and high volume, with profit margins of only about 1% to 2%. Decomposing the lost volume into store visits and in-store expenditures, we find that the majority of the losses are due to fewer store visits, with little change seen in the basket size once consumers are in the store. This is an important finding, because it suggests that strategies designed to drive store traffic could be an effective way to recover some of the lost volume. We also find that the incumbent loses some of its best customers to Wal-Mart and that a small increase in retention of these customers can significantly mitigate the losses at the incumbent store. For instance, we find that if the retailer is able to retain 5% (10%) of its best customers, it can reduce its total losses by 41% (64%). Finally, in terms of consumer characteristics, we find that Wal-Mart's impact is most pronounced for households living in close proximity to it, which is consistent with the general finding in the retail site-selection literature (e.g., Huff 1964, Brown 1989, Craig et al. 1984). Certain shopping characteristics (e.g., 9–5 weekday shoppers) and purchase behavior characteristics (e.g., store-brand buyers) are found to be more useful than household demographics in profiling defectors to Wal-Mart.

This research makes several contributions to marketing theory and practice. Foremost among these is that we provide an empirical analysis of the impact of Wal-Mart supercenter entry on a traditional retailer. As discussed above, academic research has primarily focused on competition between supermarkets with little attention given to this new retail format. Similarly, past research has studied competition and store choice issues in a static environment, whereas this study considers both the short- and long-run impact of entry in a changing competitive environment, namely, entry by a new competitor. Given the dramatic changes taking place in the retail industry, results from the study should be of interest to both academics and practitioners.

The research is also salient to the growing body of literature focusing on customer management. For instance, our findings that a small proportion of customers account for a large proportion of store losses give credence to the general recommendation in the customer relationship management (CRM) literature on the importance of customer retention. Similarly, our analysis demonstrates how a retailer can exploit the information contained in its frequent-shopper database to understand and respond to its most valuable customers. This is a vital topic because, although the information contained in frequent-shopper databases is commonly assumed to be valuable, many retailers are struggling to leverage this information. The potential difficulty of converting data into valuable marketing strategies is illustrated by the case of Safeway PLC (U.K.), which abandoned its customer card program, citing a potential savings of $80 million per year in administrative costs (*BBC News* 2000). Thus, a secondary objective of this study is to shed some light on the potential uses of the purchase history information, especially in the face of competition.

The rest of this paper is organized as follows. The next section provides a brief overview of the supercenter format, including suggestions made in the business press to counter Wal-Mart. Section 3 presents the data used in the study. Section 4 develops a joint model of interpurchase time and basket size, and §5 presents the empirical results from the model. In §6, we explore various household characteristics that can be useful in identifying potential defectors to Wal-Mart. We conclude in §7 with a discussion on limitations of the current study and directions for future research in this area.

2. The Supercenter Format

In this section we provide a brief overview of the supercenter format.[1] We discuss the motivation of discount stores to get into the grocery business, the challenges this format presents to supermarkets, and solutions suggested by some industry analysts. This format has received limited attention in the academic literature, so our discussion is primarily drawn from the business press.

Supercenters, which average 180,000 square feet, are retail stores that combine a discount department store with a full-service supermarket. They offer a wide variety of general merchandise and food items, including meat, produce, deli, and other perishables. In addition, many include ancillary services such as pharmacy, dry cleaning, vision center, Tire and Lube Express, hair salon, income tax preparation (in season), and so forth, providing consumers with a true "one-stop shopping" experience. Meijer and Fred Meyer started this format as early as the 1960s, but it is only with the arrival of Wal-Mart that this format has shown dramatic growth. The first Wal-Mart supercenter was opened in 1988, and in 1993 the company operated only 10 such stores. With 192 supercenters added in 2002, the company currently has over 2,121 supercenters. This unprecedented march by Wal-Mart into the grocery business is taking its toll along the way. According to the 2002 "Channel Blurring" study by ACNielsen, since 1999 consumer visits per year to supermarkets were down 12% while visits to supercenters were up 40%. National chains and independents alike are feeling the pressure from Wal-Mart. In the past decade, 29 chains have sought bankruptcy-court protection, with Wal-Mart as a catalyst in 25 of those cases (*Wall Street Journal* 2003).

Transition to Grocery

What motivated Wal-Mart to enter the grocery business? There are a number of reasons cited for the move, including change in the top management and the arrival of David Glass as the CEO (who had a background in the grocery business). Furthermore, by the late 1980s the discount retail industry was close to saturation, and was highly concentrated with three major players: Wal-Mart, K-Mart, and Target. The supermarket industry, on the other hand, was highly fragmented with small- to medium-size regional chains. Although this industry structure facilitated the transition to grocery, the main motivation for Wal-Mart's venture into the industry was store traffic. Indeed, industry experts believe that Wal-Mart is using food mainly as a traffic driver, with the hope of spillover to higher-margin general merchandise items, which account for 65%–70% of supercenter sales. The strategy seems to be working, with some reports suggesting that the general merchandise sales are 25%–50% higher at a supercenter than they are at discount stores in the same area (or after conversion of a discount store to supercenter).[2] The supercenter format has been so successful that Wal-Mart has chosen this path for expansion, with plans to add 200 supercenters every year for the next five years (company website). According to Trade Dimensions, with the current growth rate, over three-fourths of Kroger and Albertsons stores will be within 10 miles of a Wal-Mart supercenter within this decade.

Pricing at Wal-Mart

A general consensus in industry reports is that the prices at Wal-Mart supercenter are about 15% lower than traditional supermarkets.[3] Besides the store traffic considerations discussed above, there are several other cost-related factors driving the prices down at Wal-Mart. Foremost, Wal-Mart's size gives the company several advantages over smaller competitors, including bargaining power with the manufacturer and economies of scale in distribution systems. Furthermore, Wal-Mart's large size allows the company to bypass wholesalers with the majority of the merchandise at the supercenters, including perishables, supplied through its distribution centers. This, coupled with an EDLP strategy (which not only helps create a low-price image in a consumer's mind but also offers many operational advantages in demand forecasting) and Wal-Mart's proprietary Retail Link software, gives Wal-Mart a tremendous advantage in logistics and inventory control. According to an independent study by McKinsey & Co., Wal-Mart's efficiency gains were the source of 25% of the entire U.S. economy's productivity improvement from 1995 to 1999. Last but not least, another factor keeping the costs low at Wal-Mart is its nonunionized labor. For the majority of supermarkets, labor, which constitutes approximately 70% of the overhead, is unionized. None of Wal-Mart's employees belong to a union, and industry analysts believe that they get paid significantly less than the industry average.[4]

[1] We shall limit our discussion primarily to Wal-Mart. Target and K-Mart each has its own version of a supercenter, but Wal-Mart is by far the biggest player in the industry.

[2] Packaged Facts (1997).

[3] The *Wall Street Journal* (2003) article cited above quotes 8%–27% lower prices at supercenters. A *Time Magazine* (2003) article, "Can Wal-Mart Get Any Bigger?" reports 15% lower prices.

[4] There are a number of lawsuits pending against Wal-Mart due to its labor practices. In one such case in Jacksonville, Texas, the meat-cutters at Wal-Mart voted to form a union in February 2000. Within months, Wal-Mart had decided to close down the meat plant and replace it with case-ready meat packaged by suppliers (*KFPT News* 2002).

How to Compete with Wal-Mart?

> Competition from Wal-Mart supercenters may be inevitable, but it is not a death sentence.... (Thomas Zaucha, president and CEO of the National Grocers Association)

Given such cost asymmetries, how can supermarkets compete with Wal-Mart? Although there is no one answer, industry experts have given many suggestions. These range from shutting down the store to improving efficiency and cutting costs. In general, the recommendations fall into two broad groups: Become more like Wal-Mart, or differentiate (Rogers 2001). Indeed, there has been a move in the supermarket industry toward consolidation through mergers and acquisitions (e.g., Kroger and Fred Myers, American and Albertson) with the hope of leveraging similar bargaining power and economies of scale as Wal-Mart. Similarly, there has been a drive in the industry toward cost cutting (*Wall Street Journal* 2003). Many stores have also expanded their general merchandise items and other services such as pharmacy and banking in order to provide their own version of one-stop shopping.

Others argue it is not possible to beat Wal-Mart at its own game and recommend differentiation with a focus on the two main weaknesses of supercenters: perishables and convenience. These recommendations include providing a clean friendly store; improving fresh produce and custom-cut meat departments; emphasizing deli, ready-to-eat foods, and salad bars; broadening product assortment; and increasing the focus on understanding customer needs. Another weakness of the supercenters is that they are generally located outside the city limits. However, Wal-Mart has recognized this limitation, and is testing with scaled-down versions of supercenters ranging in size from 40,000 to 50,000 square feet. According to Wal-Mart its Neighborhood Market format will charge the same low price as its supercenters, while providing the same location convenience as supermarkets and convenience stores.

Given the discussion above, it is not surprising that supermarket managers consider Wal-Mart one of their most formidable competitors. Many of the suggestions that industry experts provide on how to tackle Wal-Mart go well beyond the scope of this paper. However, before developing any general principles on the issue, a necessary first step is an understanding of what a Wal-Mart does to a retailer's bottom line, and how it changes consumer purchase behavior. To this extent we present an empirical study that attempts to address some of these issues.

3. Frequent-Shopper Database

The data used in the study come from a single store of a large supermarket chain on the East Coast. The store in question is located in a small suburban town, which provides us with an opportunity to analyze the impact of a Wal-Mart supercenter's entry in a relatively controlled environment.[5] Based on our discussion with the store managers, the store can be classified as typical Hi-Lo format. Besides standard grocery products, the store offers a variety of services such as 24-hour shopping, in-store postal and banking services, video rental, photo developing, pharmacy, and speciality departments such as bakery, deli, salad bar, seafood, and custom-cut meat.

Table 1 Shopper Card Penetration

	No card	Employee card	Card holder
Sales ($) (%)	12.4	2.0	85.6
Number of transactions (%)	36.0	3.0	61.0
Average basket size ($)	8.00	16.60	33.50

The store also has a well-developed frequent-shopper program. Although the original purpose of such frequent-shopper programs was to create store loyalty by rewarding the best customers, over time the role of these programs has changed to being just another promotional tool. However, they do provide retailers with a wealth of information about their customers. A secondary objective of this study is to demonstrate how retailers can utilize the information contained in their database, especially when faced with competition. The frequent-shopper data that we use are unique, in that they record all transactions made in the store and capture information such as time and date of the transaction, card holder information (if a shopper card is used), and the dollar volume, unit price, quantity, and promotion for *every* UPC sold. At the same time, these data have the drawback (unlike typical scanner panel data) of only making the purchase information available for the store in question. Thus, if a card holder in our sample shops at other stores (including Wal-Mart), those purchases are not recorded.[6] While this may seem like a major shortcoming, one must realize that this is the information typically available to the retailer (unless it purchases data from outside vendors such as Information Resource Inc. (IRI) or ACNielsen).

The data are available for a period of 20 months, from November 1999 to June 2001. In August of 2000 a Wal-Mart supercenter entered 2.1 miles from the focal store. Thus, we observe reasonably long time series both before and after Wal-Mart's entry. Note that our database contains information on all transactions made at the store, but that we can only track purchases for households that are members of the frequent-shopper program. As seen in Table 1, the

[5] Over 70% of the households in our database own a house and, on average, have lived at their current residence for 14 years.

[6] Besides Wal-Mart, there is one other major grocery store in the area, as well as several smaller food shops.

Figure 1 Household Locations

Note. Distance between the focal store and Wal-Mart is 2.1 miles.

usage of the shopper card program for this retailer is quite high, with card holders accounting for over 85% of total store sales. About 2% of the sales are on the employee card and thus cannot be traced back to any individual card holder. Although noncard purchases account for 32% of all transactions, the average order size is significantly lower ($8 compared with $34 for transactions using the shopper card). The noncard purchases often tend to come from the coffee or snack shop, and the pharmacy. For most grocery categories the shopper card penetration rate is well over 90%.

Purchase history information is available for over 22,000 card holders. However, many of these card holders are casual buyers who make few purchases at this store, or who had highly irregular purchase patterns before Wal-Mart entered. The estimation results presented below use data from the top 10,000 customers, who account for 77% of card holder pre-entry sales.

3.1. Customer Location

A useful piece of information in the database is the mailing addresses for all the card holders. These addresses were geo-coded to compute each household's travel distance from the focal store and from Wal-Mart. Based on the findings in the previous literature (see, for example, research on retail site selection by Huff 1964, Brown 1989, Craig et al. 1989, etc.) as well as business press (*Progressive Grocer* 2002), we would expect location to play an important role in determining the likelihood of shifting purchases to Wal-Mart. In Figure 1 we plot the locations of the households. The location of the focal store is shown by the large star, and the location of Wal-Mart is shown by the pin. On average, consumers live about 3.5 miles from the focal store and 4.8 miles from Wal-Mart.[7] As is evident from Figure 1, many customers are clustered around the focal store. However, despite the apparent proximity, over one-third of the customers live outside the three-mile radius (considered the trading area of a typical grocery store).[8] In our empirical application, we tried several specifications for incorporating distance, including defining census tract neighborhoods, and linear and quadratic distance terms as suggested in the Hotelling-type theoretical models.

[7] These numbers are based on a straight-line distance. While one would prefer to use travel times rather than distance, there is some evidence that straight-line distance is a good proxy for actual travel time. For example, Phibbs and Luft (1995) find a correlation of 0.987 between straight-line distance and travel time, although this correlation drops to 0.826 for distances below 15 miles. Note also that our distance variable is more accurate than that used in previous research that have used the centroid of the zip code in which the household is located to compute distances (e.g., Bell et al. 1998).

[8] Sixty-six percent of the households live within the three-mile radius of the focal store and 78% live within a five-mile radius.

3.2. Observed Household Characteristics

Besides location, we use a large set of household-specific variables that could be useful in understanding the type of customers that defect to Wal-Mart. These variables are constructed using the census data as well as the transaction history of the households *prior* to Wal-Mart's entry. Table 2 provides the summary statistics on these variables, which fall into three broad categories:

• **Demographics:** The first two demographic variables in Table 2 (INC and HHSIZE) refer to the median income and median household size for the census block group in which the household resides. In general, we find significant variation in household demographics. For instance, the median income level in the block-group ranges from a low of $8,700 to a high of over $105,000. The next two demographic variables (BABY and PET) are dummies, indicating the presence of a baby or a pet, respectively. These were computed from the household purchase history data. For example, if a household is observed to purchase baby products such as diapers or baby food, it indicates the presence of an infant in the family. Similarly, purchase of dog food or cat litter indicates the presence of a pet.

• **Shopping Variables:** We use the time and day of trip information to construct a variable that relates to household shopping behavior: MSHOP (the percent of total visits that were made between 9 A.M. and 5 P.M. on weekdays, excluding holidays). This variable can be treated as proxy for a shopper's search cost. For instance, if a household is observed to make the majority of its purchases between 9 A.M. and 5 P.M. on weekdays, it suggests the presence of a retired or otherwise unemployed member in the household.

Table 2 Demographic, Shopping, and Product Purchase Variables

Variable	Description	Mean	Std	Min	Max
STDIST	Distance—focal store	3.49	4.70	0.02	49.06
PROXWM	Proximity to Wal-Mart	4.78	4.42	0.18	48.79
INC	Median income in census tract	44,352	16,860	8,713	105,218
HHSIZE	Household size in census tract	2.36	0.31	1.21	3.94
BABY	Indicator for presence of baby	10%			
PET	Indicator for presence of pet	29%			
MSHOP	Fraction of trips between 9 A.M. and 5 P.M. weekdays	0.39	0.24	0.00	1.00
E-SHOP	Fraction expenditure on produce	0.10	0.06	0.00	0.74
E-MEAT	Fraction expenditure on meat	0.14	0.08	0.00	0.77
E-HMR	Fraction expenditure on HMR	0.02	0.05	0.00	0.92
E-PL	Fraction expenditure on store brand	0.14	0.07	0.00	0.59

Note. HMR = home meal replacement.

• **Product Purchase Behavior:** The last set of variables was created using household purchase behavior in different product types (again created using transactions prior to Wal-Mart's entry). The primary motivation for these variables comes from business press reports that argue that one of the major weaknesses of the Wal-Mart supercenter is the fresh food area. As discussed in §2, Wal-Mart supercenters primarily rely on prepackaged produce and meat from suppliers. Thus households that allocate a large proportion of their expenditures to fresh produce (E-SHOP) and speciality meat and seafood (E-MEAT) are less likely to abandon this store in favor of Wal-Mart. Similarly, large purchases in prepared food departments such as salad bar and deli (E-HMR: proportion of expenditure in home meal replacement) could result in a higher affinity for the store compared with Wal-Mart. Finally, E-PL refers to the proportion of total expenditures on the store brands.[9] The likely impact of this variable in determining whether a household defects to Wal-Mart is not entirely clear. Previous research (Corstjens and Lal 2000) suggests that store brands can create store loyalty. At the same time, researchers have also found store brand buyers to be more value driven and price sensitive (Hoch 1996, Hansen et al. 2006, Pauwels and Srinivasan 2004), and if more price-sensitive households frequent discount stores, we can find the opposite effect.

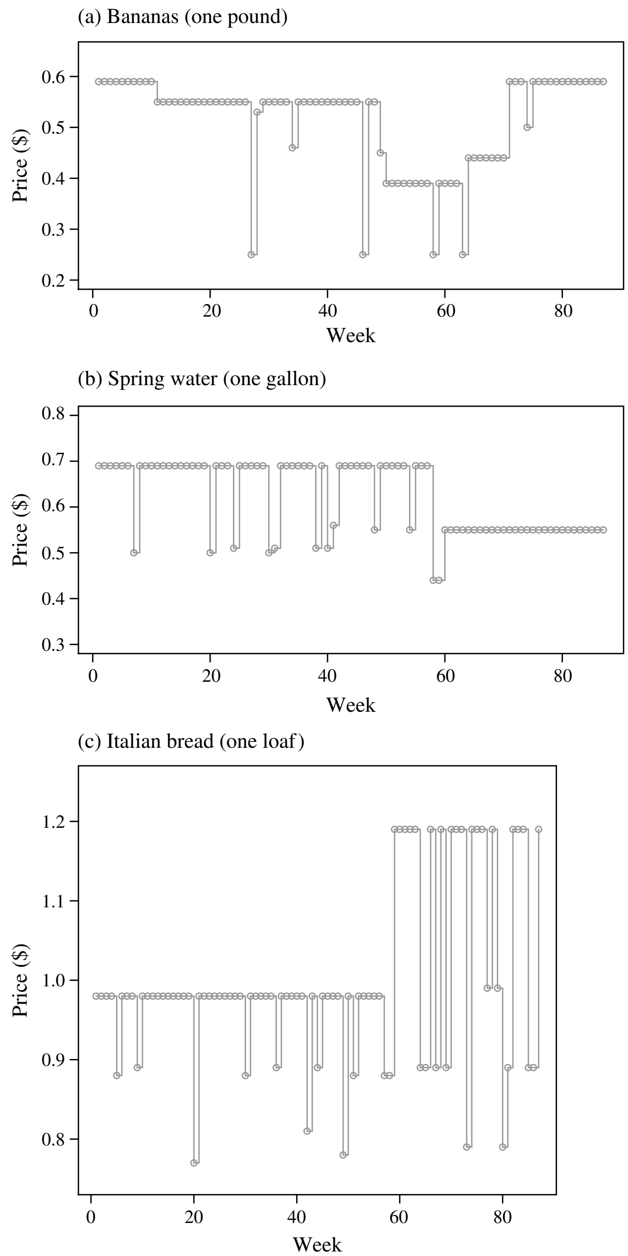
3.3. Pricing Environment

As discussed above, the database records the price and promotion information for every UPC sold in the store. We expect these marketing mix variables to influence various household decisions such as the decision to visit the store and basket size once in the store. However, creating variables to capture the overall store-pricing environment is a nontrivial task. The store carries over 50,000 unique UPCs that are classified into several hundred categories. Furthermore, several of these products (e.g., in produce and meat departments) do not carry a fixed UPC bar code that remains constant over time. Instead, these products are assigned a temporary code that changes from week to week. This makes the task of creating a price series for these products difficult if not impossible. The matter is further complicated by different price reactions by the incumbent to Wal-Mart's entry in various product categories. Figure 2 shows prices for three products: bananas, spring water, and Italian bread. It is evident that the initial reaction for bananas is to match the $0.39 per pound price of Wal-Mart with occasional $0.25 per pound promotions.[10] Over time

[9] Excludes nonpackaged goods such as items in meat and produce departments.

[10] Price of $0.39 per pound of banana is a heavily promoted item at Wal-Mart supercenters in their advertisements as well as in-store special displays.

Figure 2 Store Pricing Series for Three UPCs



prices went back to $0.59 per pound. The strategy followed in the other two products is quite different. In spring water the store seems to have moved from a Hi-Lo pricing to an EDLP pricing strategy, while the situation is reversed in Italian bread.[11]

Given the above-mentioned complexities, any measure to capture the overall store price environment would be an approximation at best. Our strategy is therefore to rely on aggregate measures designed to proxy the weekly price environment at the store. In the empirical application, we experimented with several measures including a basket-price index, a household-specific basket-price index, and overall store- and department-level promotion indices. The basket-price index was created using the share weighted price of the top 100 (250) selling items in the store. For household-specific basket prices, we used the top 100 (250) items for that household prior to Wal-Mart's entry. Finally, promotional indices were created by aggregating the total promotional discount offered on all items falling in that department. For instance, suppose the produce department consists of the following three products with regular and promoted price as: bananas ($0.55 per pound, $0.45 per pound), broccoli ($1.39 per pound, $1.39 per pound), and red grapes ($1.00 per pound, $0.90 per pound). The promotion measure for produce in that week would be $0.20.

3.4. Store Sales

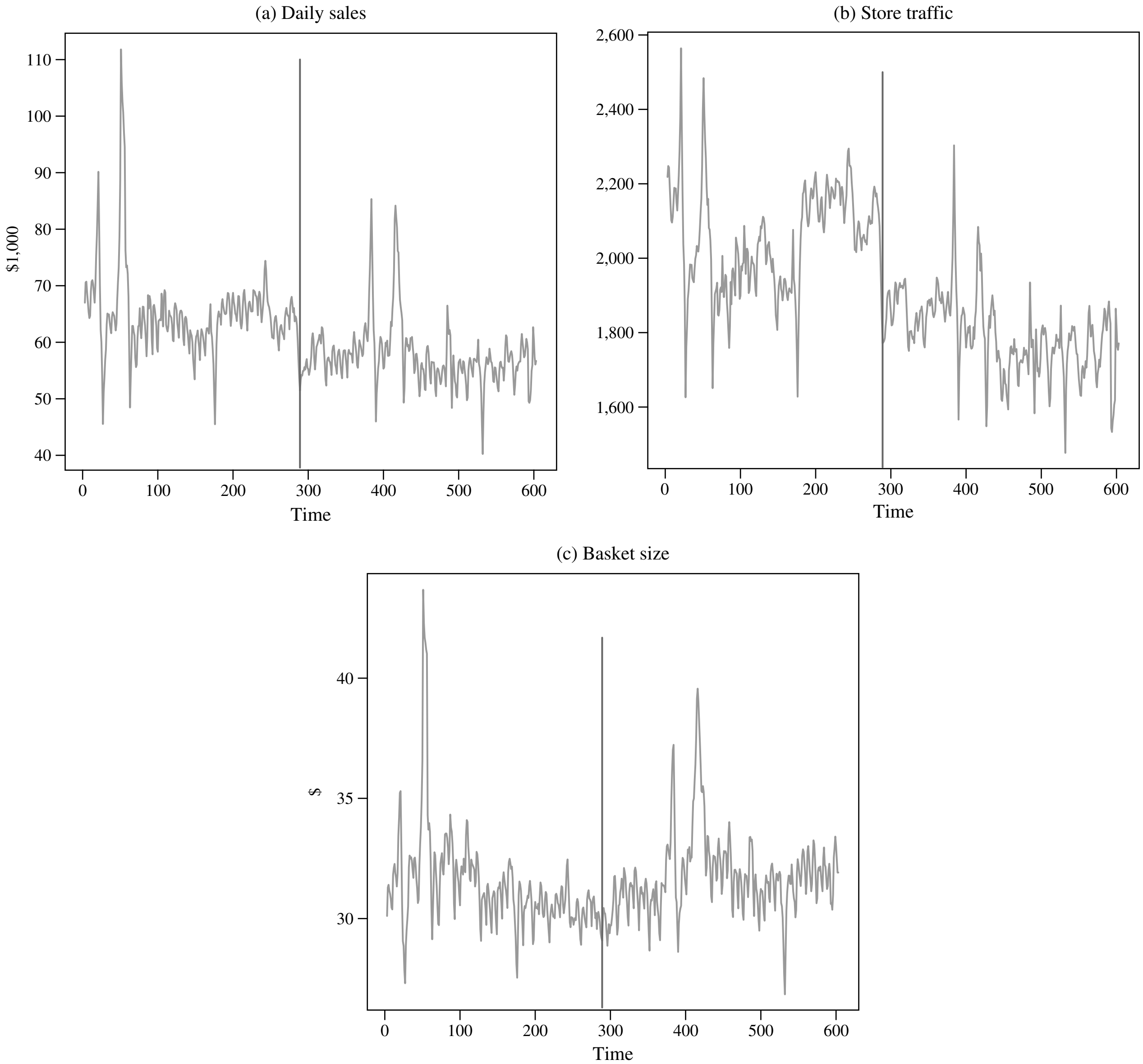
Figure 3 shows the daily store sales, store traffic (i.e., number of transactions per day), and the average basket size over the sample period. The entry of Wal-Mart is indicated by the vertical line. The spikes early in the data and around Day 400 are the Thanksgiving and Christmas weeks. Two key observations from the figures must be highlighted. First, there appears to be a significant fall in the baseline volume for the incumbent store. Second, looking at the graphs for the store visits and basket size, it seems that a large proportion of the lost revenues at the store level is due to fewer store visits, with little change in the average basket size. To formalize and test this at the household level, we next describe a model that captures these two fundamental household decisions: whether to visit the store, and basket size once at the store.

4. Model

In this section we develop a model to evaluate the impact of Wal-Mart's entry on household purchase behavior. As discussed above, a competing store's entry is likely to result in lost volume for the incumbent store. Suppose we define volume as total expenditure for all households shopping at the store over a period of a certain length, for example, T days. Let V be store volume for this period before Wal-Mart's entry and V^W store volume after entry. A casual method to evaluate the overall impact of Wal-Mart's entry is to estimate the expected value of quantities such as $\Delta_1 \equiv V^W - V$, $\Delta_2 \equiv (V^W - V)/V$, or $\Delta_3 = \log V^W - \log V$. These quantities can be estimated by simply comparing before–after averages of observed store volume. However, there are (at least) three shortcomings to this approach. First, it is important for the incumbent to understand the source of these lost sales in terms of longer interpurchase times and smaller baskets. Second, the impact of Wal-Mart's entry is likely to

[11] Our discussions with the store manager did not reveal any particular insight into the matter. It seemed that the store was experimenting with different pricing schemes. However, the store manager did indicate that the focus post–Wal-Mart entry had shifted to emphasizing the fresh produce, seafood, meat, and deli items—a point we return to in §6.

Figure 3 Daily Sales, Store Traffic, and Basket Size, Before and After Entry



Note. Entry date shown by vertical line.

be different across consumers due to observed (such as demographics) as well as unobserved factors. It is crucial for the store to understand what types of consumers or households display the biggest change in store expenditure and what the causes are of those changes. In other words, if a given household reduces its store expenditure over a certain period after Wal-Mart's entry, is this due to longer interpurchase times (i.e., fewer trips per period), smaller basket size per trip, or a combination of the two? Finally, if the store environment changes after Wal-Mart's entry, for example, if pricing and promotion strategies change, then it is important to control for these changes. For example, if the store promotes more aggressively after Wal-Mart's entry, the promotion effect will be confounded with the pure Wal-Mart effect.

To overcome the shortcomings described above, we start by decomposing overall store volume as

$$V = \sum_{h=1}^{H} e_h, \tag{1}$$

where e_h is household h's store expenditure over a period of length T. This can in turn be decomposed as

$$e_h = \sum_{t=1}^{T} d_{ht} b_{ht}, \tag{2}$$

where d_{ht} is equal to one if the store is visited on day t of the period and b_{ht} is the basket size (in dollars) of the trip. The total number of trips over the period for household h is

$$nt_h = \sum_{t=1}^{T} d_{ht}. \tag{3}$$

Letting a superscript W denote quantities after Wal-Mart's entry, and letting x_t denote variables describing the store environment (e.g., promotional activity), we can now define "pure" Wal-Mart effects at the individual and aggregate level by holding x_t fixed. For example,

$$\begin{aligned} &\mathrm{E}[nt_h^W \mid \{x_t\}_{t=1}^T] - \mathrm{E}[nt_h \mid \{x_t\}_{t=1}^T] \quad \text{and} \\ &\mathrm{E}[\log b_{ht}^W \mid x_t] - \mathrm{E}[\log b_{ht} \mid x_t] \end{aligned} \tag{4}$$

is the expected change in number of trips per period and expected change in log basket size per trip for household h.

4.1. Interpurchase Time and Basket Size

We model the two consumer decisions using a flexible model of interpurchase time (to capture when to visit the store) and semilog regression (to capture basket size once at the store). Both these household decisions are modeled jointly, and heterogeneity across households is captured by using a hierarchical structure where the full vector of model parameters (from both equations) is allowed to vary across consumers due to both observed and unobserved factors. To study the impact of Wal-Mart's entry on household purchase behavior, we allow for a structural break at the time of competitor entry.

Over the past two decades, a number of models have been proposed to capture the purchase-timing decisions of households (see Seetharaman and Chintagunta 2003 for a review). A majority of the empirical studies in marketing have used the proportional hazard model (proposed by Cox 1972) to characterize the purchase-timing behavior of households either in continuous time (e.g., Jain and Vilcassim 1991, Chintagunta and Haldar 1998) or discrete time (Gupta 1991, Helsen and Schmittlein 1993, Wedel et al. 1995). An advantage of the discrete-time approach is that it explicitly accounts for marketing mix and other covariates in periods during which households do not make a visit. For instance, in the current application it may be important to take into account the marketing mix variables on not only the purchase occasions, but also the periods in which households decide not to visit the store.

Our approach in this paper is to employ a discrete-choice framework with time-varying coefficients to capture the duration dependence embodied in consumers' choice processes. An advantage of using this approach is that we can use a flexible specification for duration dependence that allows us to approximate any shape of the household-specific hazard function. In proportional hazard models, such as those typically used in the literature, the impact of any covariate is to shift the baseline hazard up or down proportionately. Our specification is more flexible and allows for nonproportional hazard functions.

Using *days* as the basic time unit,[12] assume that an individual in each time period decides whether to visit the store and make a purchase.[13] Let U_{it} be net benefits for household i of making a purchase from the store in period t. The household will visit the store at t if $U_{it} > 0$. Assume

$$U_{it} = \beta_{i0}' f(\tau_{it}) + \beta_{ip}' p_t + \varepsilon_{it}, \quad t < T_W, \tag{5}$$

where T_W refers to the time periods before Wal-Mart's entry, τ_{it} is time since last purchase, ε_{it} is iid standard normal, and $f(\cdot)$ is some known vector function that can be made as flexible as desired. For instance, we could have $f(\tau_{it}) = (1, \tau_{it}, \tau_{it}^2, \ln \tau_{it}, \ldots)'$. p_t is a vector of time-varying covariates affecting utilities and includes time-varying marketing mix variables such as price and promotion for the incumbent store. The specification in (5) could be extended to include other factors such as expenditures on the previous purchase occasion, weekend, holiday, seasonality, and so forth (see the empirical application below).

Define D_{it} as one when $U_{it} > 0$ and zero otherwise. The probability of purchase at time t conditional on last purchase τ_t days ago is

$$\Pr(D_{it} = 1 \mid \beta_i, \tau_{it}, p_t) = \Phi(\beta_{i0}' f(\tau_{it}) + \beta_{ip}' p_t). \tag{6}$$

This is the hazard rate induced by (5) and captures the notion of individual specific hazard. The model in (5) implies a model for purchase times. Suppose we observe a purchase duration of length t_1, followed by a purchase duration of t_2. Stacking all the right-hand-side parameters and variables as (X_{it}, β_i), these durations then have likelihood

$$\begin{aligned} &\Pr\big(T_{i1} = t_1, T_{i2} = t_2 \mid \beta_i, X_i^{t_1+t_2}\big) \\ &\quad = \left\{ \prod_{t=1}^{t_1-1} \Pr(D_{it} = 0 \mid X_{it}, \beta_i) \right\} \times \Pr\big(D_{it_1} = 1 \mid X_{it_1}, \beta_i\big) \\ &\quad\quad \times \left\{ \prod_{t=t_1+1}^{t_1+t_2-1} \Pr(D_{it} = 0 \mid X_{it}, \beta_i) \right\} \\ &\quad\quad \times \Pr\big(D_{i,\,t_1+t_2} = 1 \mid X_{i,\,t_1+t_2}, \beta_i\big), \end{aligned} \tag{7}$$

where $X_i^{t_1+t_2}$ is the entire path for the covariates: $X_i^{t_1+t_2} = \{X_{it}\}_{t=1}^{t_1+t_2}$.

After Wal-Mart's entry, the utility is assumed to be

$$U_{it} = (\beta_{i0} + \beta_{i0,\,W})' f(\tau_{it}) + \beta_{ip}' p_{it} + \varepsilon_{it}, \quad t > T_W, \tag{8}$$

where $\beta_{i0,\,W}$ captures the impact of competitor entry. In the empirical application we separate out the

[12] Most marketing applications using discrete hazard models have assumed *week* as the unit of analysis. The primary motivation for the assumption is that the marketing mix variables change on a weekly basis. However, in our sample over one-third of the households visit the store more than once a week.

[13] Like most other marketing data sets, we observe a store visit only if a purchase is made.

short-run (to capture "curiosity effects") and long-run impact of Wal-Mart and also allow other model parameters (e.g., marketing mix sensitivities) to change after entry.

To model the basket size once the household is in the store, we use a semilog specification that has been used extensively in marketing for modeling sales and expenditures (e.g., Blattberg and Neslin 1990). In particular, let b_{it} be log expenditures for household i in time period t (which is zero unless $U_i(t) > 0$). If a store visit is made at time t, the pre-entry log basket size b_{it} is assumed to be

$$b_{it} = \lambda_{i0} + \lambda_{ip} p_t + \lambda_{i\tau}\tau_{it} + \varepsilon_{b,\,it}, \quad t < T_W, \tag{9}$$

where p_t is the marketing mix environment on store visit t. The parameter $\lambda_{i\tau}$ captures the impact on basket size due to the recency of the previous visit. In general, we would expect a smaller basket size if the customer had visited the store recently. Finally, we assume $\varepsilon_{e,\,it} \mid v_i \sim N(0, v_i^{-1})$. After Wal-Mart's entry, the log basket size is modeled as

$$b_{it} = (\lambda_{i0} + \lambda_{iW}) + \lambda_{ip} p_t + \lambda_{i\tau}\tau_{it} + \varepsilon_{b,\,it}, \quad t \geq T_W, \tag{10}$$

where λ_{iW} captures the impact of Wal-Mart on the basket size.

4.2. Heterogeneity

Because we expect different households to react differently to Wal-Mart's entry, it is important to account for consumer heterogeneity in the model parameters. We also expect household responses to be related to observed characteristics, such as demographics. In this paper we use a parametric approach to model household heterogeneity and let the model parameters vary across households due to both observed and unobserved factors. Let $\theta_i = (\beta_i, \lambda_i)$ be the full vector of coefficients from the purchase timing and expenditure equations discussed above. We assume θ_i follows a multivariate normal distribution with a mean vector ΠZ_i and covariance matrix Ω:

$$\theta_i \mid \Pi, Z_i \sim N(\Pi Z_i, \Omega), \tag{11}$$

where Z_i is a vector containing household characteristics.

For inference we use a hierarchical Bayesian approach. In particular, we use a Markov Chain Monte Carlo (MCMC) procedure to simulate the posterior distribution of the model parameters and to compute household-level estimates of preferences. As discussed in Allenby and Rossi (1999), Bayesian procedures are well suited for these models, especially when one is interested in making inference at the individual level. We use standard conjugate priors on the model parameters. For the parameters in the Π matrix, we use a joint normal prior with mean zero and precision 0.01. We use a Wishart prior for Ω^{-1} with degrees of freedom equal to $\dim(\theta_i) + 2 = 30$ and set the scale so the prior mean is equal to the identity matrix. Finally, we specify the prior for v_i to be gamma with shape $a_0 = 5$ and inverse scale equal to b. We use a gamma prior for the hyper parameter b with shape parameter equal to 2 and scale parameter equal to 1. The MCMC sampling algorithm was run for a total of 20,000 iterations, and we dropped the first 2,000 iterations to allow for burn-in, leaving us with 18,000 draws on which to base posterior calculations. These procedures have become quite standard in the literature, so we do not discuss the estimation algorithm in detail (a detailed discussion of this is available from the authors).

5. Estimation Results

Model Specification

Before we discuss the results, two important considerations with respect to the model specification are worth mentioning. The first relates to the distance variables discussed in §3.1. As seen from Figure 1, households live in a wide area around the two stores. This dispersion causes a problem if one wants to incorporate distance to the two stores directly as a household-specific covariate. This is because the distances will tend to be highly correlated. The overall correlation between distance to the incumbent and distance to Wal-Mart is 0.96 in our sample, which makes it impossible to estimate two separate distance effects directly. Although one could alleviate this problem to some extent by restricting the sample to households living in a small radius around the two stores, such a restriction will throw away a substantial part of the sample. Instead, we use two approaches that mitigate the problem to a certain extent. For the model presented below, we use the household distance to the incumbent and create an indicator variable for proximity to Wal-Mart (households living within one mile). We also present results for another specification (§6.2) where we use fairly fine grid of region fixed effects (census tracts) that provides us with a clean nonparametric estimate of the distance effects.

The second consideration is with regard to the function f, which, after some experimentation, was specified as

$$f(\tau_{it}) = (1, 0.1\tau_{it}, 0.01\tau_{it}^2, 1/\tau_{it}), \tag{12}$$

where the scaling in the second and third element is to stabilize estimated coefficients. This specification allows for a wide range of different hazard shapes. Recall from above that this function determines the shape of each household's hazard. For example, prior to Wal-Mart's entry, the hazard for household i will be

$$h_i(\tau; p_\tau) = \Phi(\beta_{0i,1} + \beta_{0i,2}(\tau/10) + \beta_{0i,3}(\tau^2/100) + \beta_{i0,4}/\tau + \beta_{ip}' p_\tau),$$

where p_τ is the vector of other covariates. We also separate out the impact of Wal-Mart entry in the short- (defined as the first three months after entry) and long-run effects.

Results

In Table 3 we report estimates of the hierarchical coefficients Π and the diagonal elements of Ω in (11) for the coefficients relating to the store trip models (5) and (8). The first column in Table 3 represents the mean numbers followed by the impact of observed characteristics on the parameters. The column labelled "SD" is the standard deviation of the parameter estimates representing the unobserved heterogeneity. Finally, the last column, "DEMO," shows the fraction of heterogeneity across households that is explained by the observed characteristics. Looking across the rows the first four rows capture the duration dependence prior to Wal-Mart's entry, followed by the short- and long-run effects. The variables *holiday* and *weekend* are indicator variables, while *prom* is the store-level promotion index. Note that all demographic variables except the indicator variables have been mean centered.

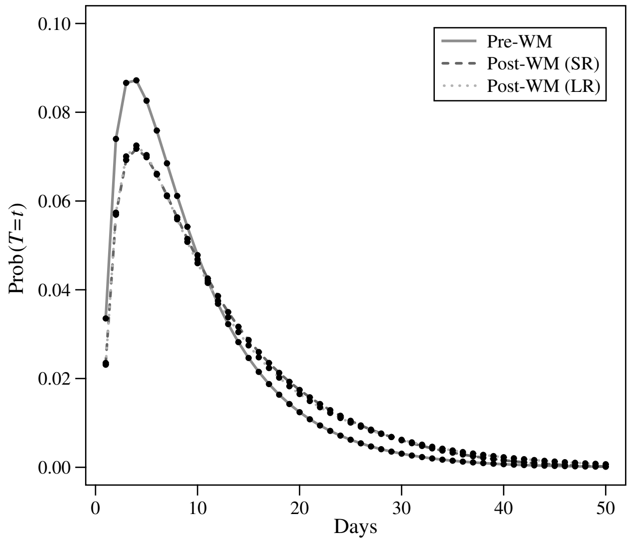
The demographic interactions show that several of the household characteristics are significant. However, the parameter variation these observed characteristics explain is not large (ranges from 2%–24%). On the other hand, the large standard deviation terms indicate that most of the heterogeneity across

Table 3 **Hierarchical Coefficient Estimates, Duration Model**

Attribute	Constant	STDIST	PROXWM	MSHOP	E-PL	E-PROD	E-MEAT	E-HMR	HHSIZE	BABY	PET	INC	SD	DEMO
Constant	**−0.96**	−0.10	−0.23	**−1.17**	**1.17**	**−2.97**	**−2.32**	−0.58	**−0.30**	0.07	0.04	0.22	1.79	0.06
	(0.05)	(0.09)	(0.14)	(0.16)	(0.50)	(0.61)	(0.47)	(0.66)	(0.14)	(0.12)	(0.08)	(0.14)	(0.05)	(0.01)
$10^{-1} \times \tau$	**−0.12**	**−0.02**	**−0.04**	0.02	**0.23**	0.14	−0.07	0.11	**0.07**	0.00	**0.17**	0.03	0.44	0.04
	(0.01)	(0.01)	(0.02)	(0.03)	(0.09)	(0.11)	(0.08)	(0.13)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
$10^{-2} \times \tau^2$	**0.03**	0.00	0.01	0.00	**−0.06**	−0.03	**0.04**	−0.01	−0.01	0.00	**−0.03**	0.00	0.09	0.03
	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.03)	(0.02)	(0.03)	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
τ^{-1}	**−0.82**	**−0.15**	**−0.08**	**−0.60**	**0.63**	**−0.35**	**0.61**	**−0.47**	**−0.20**	0.03	0.02	0.01	0.56	0.12
	(0.01)	(0.02)	(0.03)	(0.04)	(0.12)	(0.15)	(0.11)	(0.17)	(0.03)	(0.03)	(0.02)	(0.03)	(0.01)	(0.01)
D_{SR}	**−1.13**	−0.09	0.07	**0.68**	**−2.45**	1.23	0.98	0.80	−0.19	−0.12	**−0.23**	**−0.47**	1.85	0.04
	(0.07)	(0.12)	(0.19)	(0.24)	(0.70)	(0.80)	(0.63)	(0.89)	(0.21)	(0.17)	(0.12)	(0.18)	(0.08)	(0.01)
$10^{-1} \times D_{SR} \times \tau$	**−0.05**	−0.02	−0.02	**0.10**	**−0.27**	0.00	**0.27**	0.39	−0.06	−0.01	−0.03	0.02	0.39	0.03
	(0.02)	(0.02)	(0.04)	(0.05)	(0.14)	(0.17)	(0.13)	(0.22)	(0.04)	(0.03)	(0.02)	(0.04)	(0.02)	(0.01)
$10^{-2} \times D_{SR} \times \tau^2$	**0.02**	−0.01	−0.01	0.00	0.01	0.03	−0.02	−0.01	0.01	0.00	**0.02**	0.01	0.10	0.02
	(0.00)	(0.01)	(0.01)	(0.01)	(0.03)	(0.04)	(0.03)	(0.06)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)
$D_{SR} \times \tau^{-1}$	−0.03	−0.04	−0.08	−0.04	0.08	−0.13	0.07	0.32	−0.07	0.03	−0.01	0.02	0.41	0.02
	(0.02)	(0.04)	(0.05)	(0.06)	(0.18)	(0.23)	(0.17)	(0.25)	(0.05)	(0.04)	(0.03)	(0.05)	(0.02)	(0.01)
D_{LR}	**−1.11**	−0.10	0.03	**0.60**	**−2.76**	1.35	**1.32**	1.19	−0.22	−0.06	**−0.22**	**−0.43**	1.85	0.05
	(0.07)	(0.12)	(0.19)	(0.23)	(0.69)	(0.79)	(0.63)	(0.88)	(0.20)	(0.17)	(0.11)	(0.18)	(0.08)	(0.01)
$10^{-1} \times D_{LR} \times \tau$	**−0.05**	−0.01	0.00	**0.06**	**−0.38**	0.00	0.12	0.13	−0.01	**−0.08**	**−0.07**	0.03	0.29	0.04
	(0.01)	(0.01)	(0.03)	(0.03)	(0.09)	(0.11)	(0.08)	(0.15)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
$10^{-2} \times D_{LR} \times \tau^2$	**0.01**	**−0.01**	−0.01	−0.01	**0.06**	0.04	−0.02	0.01	0.00	0.01	**0.02**	0.01	0.08	0.03
	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.03)	(0.02)	(0.04)	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
$D_{LR} \times \tau^{-1}$	**−0.05**	−0.04	0.02	**0.12**	0.02	−0.14	**−0.32**	−0.28	0.02	**−0.07**	**−0.04**	−0.02	0.38	0.02
	(0.02)	(0.03)	(0.04)	(0.05)	(0.14)	(0.18)	(0.13)	(0.19)	(0.04)	(0.03)	(0.02)	(0.04)	(0.01)	(0.01)
holiday	**0.12**	**−0.02**	0.00	**−0.34**	0.00	**0.26**	0.08	**−0.18**	**0.05**	−0.01	**0.03**	**0.05**	0.19	0.16
	(0.01)	(0.01)	(0.02)	(0.02)	(0.06)	(0.08)	(0.06)	(0.09)	(0.02)	(0.02)	(0.01)	(0.02)	(0.00)	(0.02)
holiday $* D$	**−0.06**	0.00	0.00	**0.13**	0.06	−0.16	−0.02	0.16	−0.04	0.03	−0.01	−0.03	0.22	0.04
	(0.01)	(0.01)	(0.02)	(0.03)	(0.09)	(0.10)	(0.08)	(0.12)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
weekend	**0.03**	**0.04**	0.02	**−0.12**	**0.15**	**0.23**	**0.22**	0.00	**0.04**	**−0.02**	**0.03**	**−0.04**	0.37	0.02
	(0.01)	(0.01)	(0.02)	(0.02)	(0.06)	(0.08)	(0.06)	(0.09)	(0.02)	(0.01)	(0.01)	(0.02)	(0.00)	(0.00)
weekend $* D$	0.00	**−0.08**	−0.01	**−0.73**	**−0.14**	−0.08	0.11	0.07	−0.03	0.01	0.00	0.03	0.32	0.24
	(0.01)	(0.01)	(0.02)	(0.02)	(0.07)	(0.08)	(0.06)	(0.09)	(0.02)	(0.02)	(0.01)	(0.02)	(0.00)	(0.01)
prom	−0.01	−0.01	0.02	**0.22**	−0.06	**0.42**	**0.26**	0.11	0.01	0.00	0.00	**−0.04**	0.26	0.06
	(0.01)	(0.01)	(0.02)	(0.02)	(0.07)	(0.09)	(0.07)	(0.10)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
prom $* D$	**0.14**	0.02	−0.01	**−0.07**	**0.27**	−0.17	−0.12	−0.11	0.03	0.01	**0.04**	**0.07**	0.27	0.04
	(0.01)	(0.02)	(0.03)	(0.03)	(0.10)	(0.11)	(0.09)	(0.12)	(0.03)	(0.02)	(0.02)	(0.03)	(0.01)	(0.01)

Note. $D_{SR} = 1$ in the first three months following entry after which $D_{LR} = 1$, and $D = D_{SR} + D_{LR}$. Posterior means with posterior standard deviation are in parentheses.

Figure 4 **Distributions of Inter-Store-Visit Times**



Note. WM = Wal-Mart; SR = short run; LR = long run.

households is due to unobserved factors. Prior to Wal-Mart's entry, consumers derive higher utility shopping on week-ends and during holidays, although the effect of holiday shopping goes down after entry. Interestingly, promotional sensitivity is significantly higher following Wal-Mart's entry. This could be due to changes in consumer base (higher proportion of promotional-sensitive consumer shopping at the store), an inherent change within consumers, or some promotional tactics used by the retailer.

Because it is hard to directly interpret the coefficients of the variables that are functions of τ ("time since last purchase"), we plot the implied distributions of store-visit times. Figure 4 shows the distribution of store-visit times prior to Wal-Mart entry, and the distribution in the short and long run after entry. It is evident that the average inter-store-visit time increases after Wal-Mart's entry (since probability mass shifts from smaller duration times to larger). There does not appear to be a significant difference in the short- and long-run effects.

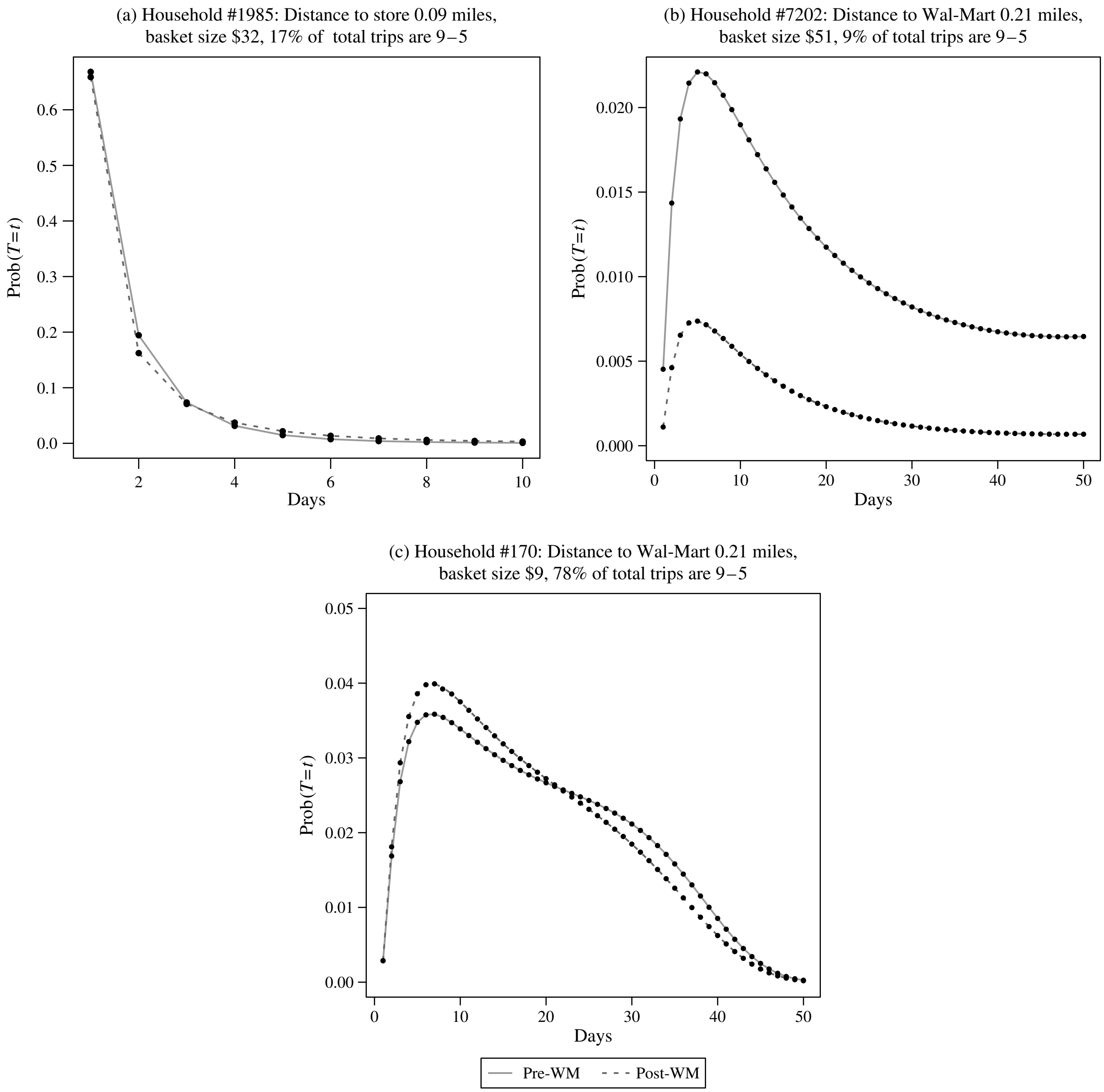
To indicate the potential impact of customer characteristics on store visits, we plot in Figure 5 the implied distributions of store-visit timing before and after Wal-Mart's entry for three select households. The top graph shows the distribution for household number 1985 in the sample. This household is located very close to the incumbent store. The middle graph shows the visit-time distribution for household number 7202, which is located right next to Wal-Mart. It is apparent from the graphs that the entry of Wal-Mart had very different impacts on these two households, suggesting that household location can be potentially important in explaining impacts across households. Next consider the visit-time distribution for household number 170 (the lower panel in Figure 5). This household is also located close to Wal-Mart and is in fact a neighbor of household number 7202. Surprisingly, we find that entry of Wal-Mart had little impact on household 170. However, a deeper probe into the purchase behavior can explain why we observe these different reactions by these households. Prior to Wal-Mart' entry, household 170 visits the incumbent store twice as frequently and with a significantly smaller basket size compared with household 7202. More importantly, 78% of the total trips for household 170 occur between 9 A.M. and 5 P.M. on (nonholiday) weekdays (compared with only 9% for household 7202). This in turn implies the presence of a retired or otherwise unemployed person in household 170. Thus, besides distance, a number of other household characteristics are important in determining the likely impact of Wal-Mart's entry. We explore these issues further in §6.

Table 4 shows the estimates for the basket-size Equations (9) and (10). The Wal-Mart dummy for the short and long run are negative and significant, indicating that the average basket size goes down after Wal-Mart's entry. The coefficient to τ shows that basket size tends to increase with time since the previous purchase. Basket sizes also tend to be higher during weekends and holidays, although the latter effect is mitigated by Wal-Mart's entry. The promotion parameter has an incorrect sign pre-entry, perhaps due to the aggregate store-level measure that we use. However, similar to the store-visit equation, the promotional sensitivity increases significantly after Wal-Mart. Although not reported here, correlation between the basket size intercept and promotion sensitivity indicates that small basket households are more promotion sensitive, which is consistent with the findings in the previous literature (Bell and Lattin 1998). Looking across the demographic coefficients, we find several of the household characteristics to be significant and the total proportion of the heterogeneity variation explained is a bit higher compared to the store-visit equation. However, there is evidence of substantial unobserved heterogeneity in the coefficients as indicated by the large standard deviation parameters.

Monthly Expenditures

One of the advantages of the Bayesian procedure used for inference is that it provides us with household-level estimates (for example, the posterior mean of θ_i for each household) as a byproduct of the MCMC procedure. These household-level parameters can be used to quantify the combined effect of entry on duration times and basket size by simulating expected monthly expenditure for each household (holding other variables fixed at their mean levels). This amounts to computing the expected value of (2) for $T = 30$ for each household before and after Wal-

Figure 5 Distributions of Inter-Store-Visit Times for Three Households



Mart's entry. In Table 5 we report the average short- and long-run effects of Wal-Mart's entry on monthly expenditures, store visits, and basket size across the population. The average effect for the whole sample is −$24 in monthly expenditures and the overall effect across the population is −$241,319. This translates to approximately 17% of the monthly store volume before Wal-Mart's entry. Such a large drop in volume is alarming for the retailer, considering that supermarkets operate on the principle of high volume with profit margins only in the range of 1%–2%.

Three observations are notable from the numbers presented in Table 5. First, the effect of Wal-Mart seems a little larger in the short run compared with the long run. This could indicate the presence of some curiosity effect on the part of some households. Second, the majority of the losses appear to come from a drop in store visits (15% drop in the long run) as opposed to a change in basket size (2% drop). This finding could be important for the retailer: It suggests that strategies geared toward driving store traffic, for example feature advertising, could be useful in mitigating losses to Wal-Mart. Finally, the large standard deviations in Table 5 suggest that the impact of Wal-Mart varies dramatically across households. The large standard deviations in turn suggest that characterizing the households that are affected by Wal-Mart's entry the most may be of interest.

Table 4 Hierarchical Coefficient Estimates, Basket Size Model

Attribute	Constant	STDIST	PROXWM	MSHOP	E-PL	E-PROD	E-MEAT	E-HMR	HHSIZE	BABY	PET	INC	SD	DEMO
Constant	**3.37**	**0.23**	0.01	**−1.08**	**−1.26**	**−1.30**	**−1.08**	−0.11	−0.05	0.21	0.04	−0.19	0.95	0.12
	(0.05)	(0.10)	(0.14)	(0.18)	(0.53)	(0.63)	(0.49)	(0.63)	(0.15)	(0.14)	(0.09)	(0.14)	(0.07)	(0.03)
D_{SR}	**−0.98**	0.08	0.29	**1.11**	**−1.11**	**1.86**	0.43	1.12	0.23	0.05	−0.16	0.28	0.71	0.28
	(0.08)	(0.14)	(0.21)	(0.24)	(0.78)	(0.85)	(0.76)	(0.81)	(0.22)	(0.19)	(0.13)	(0.21)	(0.13)	(0.09)
D_{LR}	**−0.97**	0.08	0.28	**1.13**	−1.14	**2.04**	0.45	1.29	0.18	0.08	−0.14	0.29	0.71	0.28
	(0.08)	(0.14)	(0.21)	(0.24)	(0.78)	(0.85)	(0.77)	(0.81)	(0.22)	(0.19)	(0.13)	(0.21)	(0.13)	(0.09)
τ	**0.01**	**−0.01**	0.00	**0.01**	0.00	0.01	0.00	0.01	0.00	0.00	**0.01**	0.00	0.04	0.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
prom	**−0.04**	−0.01	0.02	**0.12**	0.08	**0.21**	0.12	−0.01	**0.06**	0.00	**0.02**	**0.04**	0.15	0.09
	(0.01)	(0.01)	(0.02)	(0.03)	(0.08)	(0.09)	(0.07)	(0.09)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
prom $*D$	**0.13**	−0.01	−0.04	**−0.14**	0.11	−0.28	−0.03	−0.13	−0.03	−0.01	0.02	−0.04	0.11	0.21
	(0.01)	(0.02)	(0.03)	(0.03)	(0.11)	(0.12)	(0.11)	(0.12)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.07)
weekend	**0.02**	0.01	0.00	**−0.11**	0.11	0.01	0.09	−0.03	−0.02	0.01	**0.04**	−0.03	0.28	0.02
	(0.01)	(0.01)	(0.02)	(0.02)	(0.06)	(0.07)	(0.05)	(0.08)	(0.02)	(0.01)	(0.01)	(0.02)	(0.00)	(0.00)
weekend $*D$	**0.03**	0.01	0.01	**−0.24**	−0.12	0.03	−0.09	−0.02	0.00	0.00	**−0.02**	0.03	0.26	0.05
	(0.01)	(0.01)	(0.02)	(0.02)	(0.07)	(0.09)	(0.07)	(0.09)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
holiday	**0.12**	−0.01	−0.03	**−0.07**	−0.02	0.16	**0.18**	**−0.22**	0.01	−0.03	0.00	**0.06**	0.21	0.04
	(0.01)	(0.01)	(0.02)	(0.02)	(0.08)	(0.09)	(0.07)	(0.10)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
holiday $*D$	**−0.07**	0.01	0.04	0.01	0.10	0.07	−0.07	0.17	0.00	0.01	0.03	−0.04	0.23	0.02
	(0.01)	(0.02)	(0.03)	(0.03)	(0.11)	(0.13)	(0.10)	(0.12)	(0.03)	(0.03)	(0.02)	(0.03)	(0.01)	(0.01)

Note. $D_{SR} = 1$ in the first three months following entry after which $D_{LR} = 1$, and $D = D_{SR} + D_{LR}$. Posterior means with posterior standard deviation are in parentheses.

To dig deeper into the issue of distribution of entry effects, we assigned households into deciles based on the long-run Wal-Mart effects. In Figure 6 we plot the pre–Wal-Mart monthly expenditures, monthly store visits, and basket size for the households with the largest impact (decile 1) along with the distributions for households in deciles 2 to 10. It is apparent that Wal-Mart has the highest impact on households with large pre-entry expenditures at the incumbent. In particular, the households in the top decile have significantly higher monthly store visits (4.81 versus 3.51) and basket sizes ($63.60 versus $39.71), translating into substantially higher monthly expenditures ($255 versus $113). Aggregating across the households in the top decile, we find that this 10% of the households alone accounts for 64% of the observed losses for the incumbent.[14]

This last finding is troubling for the retailer, because Wal-Mart is stealing some of its best customers. At the same time, it also presents interesting targeting opportunities. As discussed above, traffic generation using feature advertisements could be a useful strategy to pursue. However, feature advertisement represents a communication tool that can rarely be customized to the individual level. On the other hand, customized coupons can readily be used to target the high-value customers that have defected to Wal-Mart. The frequent-shopper database already contains useful purchase history and mailing address information, so pursuing such a strategy could be fruitful. By retaining the top 5% of these customers, the retailer can reduce its losses to Wal-Mart by 41%. Thus, understanding the needs and preferences of this handful of customers and the use of targeted mechanisms could be rewarding in the long run.

6. Profiling the Defectors to Wal-Mart

Results in the previous section show that the impact of Wal-Mart varies across households and that a small proportion of customers account for a large proportion of losses at the incumbent store. We now explore the extent to which the households that respond most

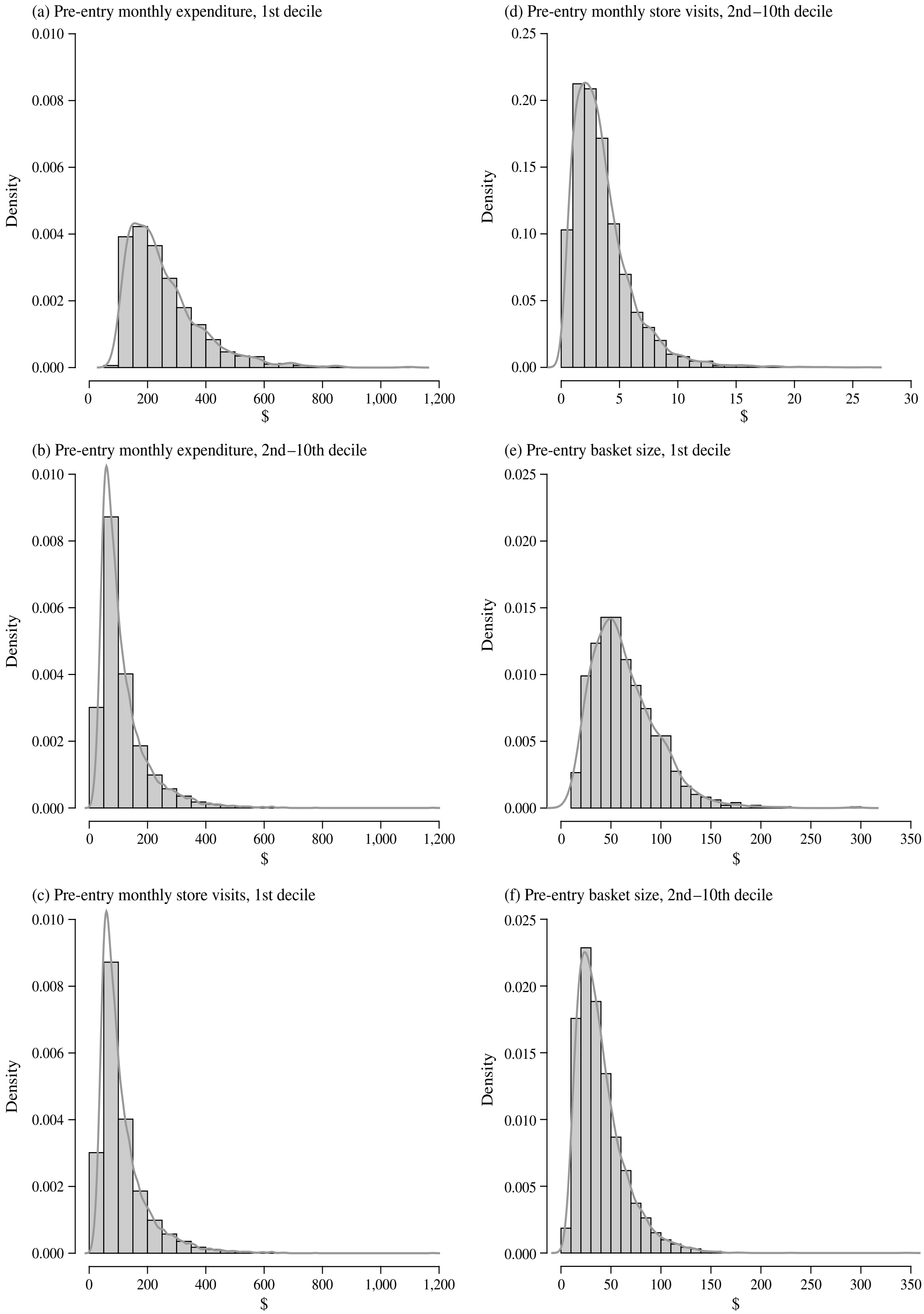
Table 5 Before and After Wal-Mart's Entry: Monthly Expenditures, Store Visits, and Basket Size

	Before–Wal-Mart	After–Wal-Mart (LR)	After–Wal-Mart (LR)
Expenditure ($)	127.23	103.64	104.61
	(98.64)	(96.43)	(97.03)
Store visits	3.64	3.14	3.07
	(2.62)	(2.73)	(2.61)
Basket size ($)	42.06	39.79	40.67
	(26.02)	(25.03)	(24.91)

Note. Standard deviation is in parentheses. SR = short run; LR = long run.

[14] Another approach to looking at the distribution of entry effects would be to assign households to deciles based on their pre-entry expenditures (most valuable to least) and then analyze Wal-Mart's impact on each decile. Looking at the joint distribution of customer value and Wal-Mart's impact (not reported in the paper), we find that almost 40% of the best customers fall in the top decile of the impact distribution.

Figure 6 Pre-entry Monthly Expenditure, Store Visits, and Basket Size by Decile of Impact Distribution



to Wal-Mart can be profiled based on their locations and other observed characteristics. Note that from a managerial perspective, such profiling of households that respond to Wal-Mart can be quite important. For instance, it can allow the retailer to not only target customers with similar characteristics at this store, but also transfer the findings to other store locations where the retailer competes with Wal-Mart. To be more specific, consider the problem of identifying the consumers that are at high risk of defection at a different location where a Wal-Mart is scheduled to open in the next few months. To take findings from the experience of this store to the new location, it is important not only to identify the households based on their individual card numbers or the region in which they live, but also to map these households on some actionable demographic or other characteristics. Such mapping then becomes the basis for identifying potential defectors at the new location. Since store opening information is generally available well in advance, it can allow the retailer to take preemptive actions for the households who are at a high risk of defection.

6.1. Marginal Effects for Observed Household Characteristics

In the results presented in Tables 3 and 4, several of the household characteristics were found to be significant. We now compute the marginal effects for the observed household characteristics entering the second stage of the hierarchical model. Note that evaluating the direct impact of the elements of the covariate vector is nontrivial because the covariates affect both store-visit frequencies and basket sizes. We evaluate the overall effect of each covariate on the size of the Wal-Mart effect as follows: Let e_i denote monthly expenditure for a household. The mean pre-entry expenditure at the incumbent for a population with covariate vector $Z = z$ is

$$\begin{aligned}&\mathrm{E}[e_i \mid D_{\mathrm{SR}}=0, D_{\mathrm{LR}}=0, Z=z, \theta]\\ &\quad=\int \mathrm{E}[e_i \mid D_{\mathrm{SR}}=0, D_{\mathrm{LR}}=0, Z=z, \theta, \beta_i]\phi(\beta_i \mid z\pi, \Omega)\, d\beta_i,\end{aligned}$$

while the mean short- and long-run expenditures after entry are computed analogously with $D_{\mathrm{SR}} = 1$ and $D_{\mathrm{LR}} = 1$, respectively.

We consider the effect of the covariates on a specific Wal-Mart effect: The change in the log of mean expenditure in the short run and long run, i.e.,

$$\begin{aligned}\Delta_{\mathrm{SR}}(\theta; z) \equiv{}& \log(\mathrm{E}[e_i \mid D_{\mathrm{SR}}=1, D_{\mathrm{LR}}=0, Z=z, \theta])\\ &-\log(\mathrm{E}[e_i \mid D_{\mathrm{SR}}=0, D_{\mathrm{LR}}=0, Z=z, \theta]),\\ \Delta_{\mathrm{LR}}(\theta; z) \equiv{}& \log(\mathrm{E}[e_i \mid D_{\mathrm{SR}}=0, D_{\mathrm{LR}}=1, Z=z, \theta])\\ &-\log(\mathrm{E}[e_i \mid D_{\mathrm{SR}}=0, D_{\mathrm{LR}}=0, Z=z, \theta]).\end{aligned}$$

$\Delta_{\mathrm{SR}}(\theta; z)$ is the change in the log of mean expenditure from pre-entry to post-entry in the short run, while $\Delta_{\mathrm{LR}}(\theta; z)$ is the long-run change for a population characterized by covariate vector z.[15] We compute the effects of changing z on $\Delta_{\mathrm{SR}}(\theta, z)$ and $\Delta_{\mathrm{LR}}(\theta, z)$. In particular, for each covariate z_j we compute

$$\begin{aligned}\delta_{\mathrm{SR}}(\theta, z_{j,\%5}) &\equiv \Delta_{\mathrm{SR}}(\theta, z_{j,\%5}) - \Delta_{\mathrm{SR}}(\theta, \bar{z}),\\ \delta_{\mathrm{SR}}(\theta, z_{j,\%95}) &\equiv \Delta_{\mathrm{SR}}(\theta, z_{j,\%95}) - \Delta_{\mathrm{SR}}(\theta, \bar{z}),\\ \delta_{\mathrm{SR}}(\theta, \bar{z}_j + \mathrm{sd}_j) &\equiv \Delta_{\mathrm{SR}}(\theta, \bar{z} + \mathrm{sd}_j) - \Delta_{\mathrm{SR}}(\theta, \bar{z})\end{aligned}$$

(and similarly for δ_{LR}), where $\bar{z}$ is the sample average of the covariate vector, $z_{j,5\%}$ ($z_{j,95\%}$) is equal to $\bar{z}$ except for covariate j, which is equal to its 5th percentile (95th percentile), while $\bar{z} + z_{j,\mathrm{sd}}$ is equal to $\bar{z}$ except for covariate j, which is equal to $\bar{z}_j$ plus a one standard deviation increase. These effects capture the marginal effect of changing covariate j on the short- and long-run Wal-Mart effect. In Table 6 we report the posterior mean and standard deviation of $\delta_{\mathrm{SR}}(\theta; z)$ and $\delta_{\mathrm{LR}}(\theta; z)$ for different values of z. For example, for store distance (STDIST) we see that changing the store distance variable from its mean to its 5th percentile (i.e., a population that is very close to the local store) decreases the Wal-Mart effect by 1%, while changing it to the 95th percentile increases the mean Wal-Mart effect by 4%. A one standard deviation increase in distance increases the mean Wal-Mart effect by 2%. In general, the marginal effects of both distance variables are rather small, a point we return to below.

A few covariates do have a substantial impact on the size of the Wal-Mart effect. For example, the measure for shopping costs seems quite important. Recall that the variable MSHOP (percent of trips between 9 a.m.–5 p.m. weekdays) was created as proxy for the presence of a retired or otherwise unemployed member in the household. From the marginal effects we see that the households that do the majority of their shopping between 9 a.m. and 5 p.m. on weekdays are significantly less likely to abandon the incumbent. This finding is not surprising, because these households would tend to have better opportunities to take advantage of the promotions offered at all of the stores in the neighborhood. Similarly, the proportion of expenditures on private label is quite important. In particular, a one standard deviation increase in the private label expenditure ratio (E-PL) increases the size of the Wal-Mart effect by 13% in the short run (16% in long run) compared with the mean effect, thus almost doubling the overall effect for this subpopulation. On the other hand, households with the lowest E-PL have a Wal-Mart effect 14% smaller than the

[15] Note that it is not possible to define the effect as the change in the mean of log expenditure, because expenditure has nonzero probability mass at zero (corresponding to moving all expenditures to Wal-Mart).

Table 6 Effects of Covariates on Mean Wal-Mart Effect

	Short-run effect: $\Delta_{SR}(\theta; z)$				Long-run effect: $\Delta_{LR}(\theta; z)$			
Attribute	$Z=\bar{z}$	$Z=z_{5\%}$	$Z=z_{95\%}$	$Z=\bar{z}+\text{sd}_z$	$Z=\bar{z}$	$Z=z_{5\%}$	$Z=z_{95\%}$	$Z=\bar{z}+\text{sd}_z$
STDIST	−0.22 (0.02)	−0.21 (0.02)	−0.28 0.03	−0.25 (0.02)	−0.22 (0.02)	−0.20 (0.02)	−0.28 0.03	−0.25 (0.02)
PROXWM	−0.22 (0.02)	−0.22 (0.02)	−0.30 0.03	−0.24 (0.02)	−0.22 (0.02)	−0.21 (0.02)	−0.31 0.03	−0.24 (0.02)
MSHOP	−0.22 (0.02)	−0.40 (0.02)	−0.05 (0.02)	−0.12 (0.02)	−0.22 (0.02)	−0.37 (0.02)	−0.06 (0.02)	−0.12 (0.02)
E-PL	−0.22 (0.02)	−0.12 (0.02)	−0.35 (0.02)	−0.30 (0.02)	−0.22 (0.02)	−0.06 (0.02)	−0.39 (0.02)	−0.32 (0.02)
E-PROD	−0.22 (0.02)	−0.22 (0.02)	−0.23 (0.02)	−0.23 (0.02)	−0.22 (0.02)	−0.24 (0.02)	−0.20 (0.02)	−0.21 (0.02)
E-MEAT	−0.22 (0.02)	−0.28 (0.02)	−0.14 (0.02)	−0.18 (0.02)	−0.22 (0.02)	−0.29 (0.02)	−0.11 (0.02)	−0.16 (0.02)
E-HMR	−0.22 (0.02)	−0.24 (0.02)	−0.18 (0.02)	−0.19 (0.02)	−0.22 (0.02)	−0.25 (0.02)	−0.16 (0.02)	−0.17 (0.02)
HHSIZE	−0.22 (0.02)	−0.19 (0.02)	−0.26 (0.02)	−0.25 (0.02)	−0.22 (0.02)	−0.18 (0.02)	−0.27 (0.02)	−0.25 (0.02)
BABY	−0.22 (0.02)	−0.22 (0.02)	−0.26 0.03	−0.23 (0.02)	−0.22 (0.02)	−0.22 (0.02)	−0.24 (0.03)	−0.22 (0.02)
PET	−0.22 (0.02)	−0.25 (0.02)	−0.18 (0.02)	−0.21 (0.02)	−0.22 (0.02)	−0.24 (0.02)	−0.17 (0.02)	−0.21 (0.02)
INC	−0.22 (0.02)	−0.26 (0.02)	−0.18 (0.02)	−0.20 (0.02)	−0.22 (0.02)	−0.27 (0.02)	−0.14 (0.02)	−0.18 (0.02)

Note. Posterior means with posterior standard deviation are in parentheses.

average in the short run (19% in the long run), making the effect virtually zero for this subpopulation. These findings indicate that households who spend a large fraction of their total grocery expenditure on the private label brand are—on average—much more likely to switch to Wal-Mart. Thus, we find that store-brand buyers have a higher likelihood of moving purchases to Wal-Mart, which is in contrast to the findings in the previous literature that suggest that store-brand buyers are also more store loyal (Corstjens and Lal 2000).

Table 6 also shows that households that spend a large proportion on speciality meat and home meal replacement (HMR) items are less likely to defect to Wal-Mart. This seems consistent with the industry reports that these items are better catered at supermarkets rather than at Wal-Mart. Overall, the shopping- and purchase-related variables appear more important than demographic variables. Although the low explanatory power of demographics is a general phenomenon reported in the literature (e.g., see Rossi et al. 1996), we should acknowledge that our demographic variables were computed using census data and may not be as accurate in standard scanner panel data.

6.2. Customer Locations

The discussion above shows that store distances play a relatively minor role in determining the propensity to shift purchases to Wal-Mart. This seems quite in contrast to the findings in the retail site selection literature (Huff 1964, Brown 1989, Craig et al. 1984, and so on) as well as business press (*Progressive Grocer*, 69th annual report, 2002), namely that location is one of the most important factors in determining store choice. Given the importance of location in the retailing world, we re-estimate the model with a different specification that alleviates the high-correlation problem discussed above, to a certain extent. In particular, in the new specification we replace the household characteristics in the second stage of the hierarchical model (11) with $Z_i = \{D_{ji}^r\}_{i=1}^n$, where D_{ji}^r is an indicator variable equal to one if household i is located in census tract region j. This approach has several advantages over the specification that includes distances directly. First, as pointed out, the correlation between distance to the focal store and distance to Wal-Mart is 0.96 in our sample, which makes estimating two separate distance effects difficult. Second, we use a fairly fine grid of regions in our dummy specification that provides us with a clean nonparametric estimate of the distance effects. Finally, the dummy specification allows us to capture other census tract–specific characteristics apart from distance, such as demographics and actions of other retailers in the region. The market has another (independent) supermarket (located in Region 6) as well as several other small convenience or food stores. Thus each dummy coefficient

Table 7 **Average Characteristics for the 19 Census Tract Regions**

Group ID	Frequency	Average distance in miles to: Focal store	Wal-Mart	Percentage change in expenditure after Wal-Mart (std. dev.)
1	392	6.10	3.70	−25 (48)
2	242	6.88	7.69	−18 (36)
3	1,026	14.11	14.32	−27 (38)
4	241	2.38	1.85	−19 (36)
5	435	1.24	1.57	−16 (34)
6	292	1.33	1.86	−16 (34)
7	345	1.64	2.31	−15 (35)
8	309	2.16	4.51	−17 (34)
9	747	1.65	3.10	−17 (32)
10	1,221	0.64	2.72	−13 (37)
11	249	2.61	4.16	−18 (37)
12	315	5.00	4.63	−18 (40)
13	674	5.10	7.48	−11 (33)
14	626	5.98	7.98	−23 (37)
15	884	0.39	2.35	−15 (38)
16	820	0.71	3.16	−13 (38)
17	963	1.41	3.91	−13 (36)
18	240	2.27	0.82	−38 (34)
19	3,317	1.81	2.42	−18 (40)

will represent the coefficient for a region with a certain composition of distance to Wal-Mart and the local store and other tract-specific information.

The results are presented in Table 7. Note that the regions show considerable variation in distance, ranging from regions very close to the focal store (Regions 15, 10, and 16) to regions close to Wal-Mart (Region 18) to regions far away from both stores (e.g., Region 3). The last column of Table 7 shows the simulated effects and standard deviations by region in the long run (short-run effects are similar). With this specification, we do observe an overall effect of store distance. In particular, the impact is higher in areas close to Wal-Mart (−38% in Region 18) and areas far from both stores (Region 3). The impact of Wal-Mart's entry is lowest in regions close to the incumbent store. However, large standard deviation numbers indicate that even with this specification there is considerable variation across households within a region.

7. Discussion and Future Research

One of the biggest challenges facing the supermarket industry is competition from Wal-Mart. Although a relatively new player, Wal-Mart through its supercenter format has become the nation's largest grocer, and supermarket managers consider it their biggest concern in the coming years. Using a unique frequent-shopper database, we provide an empirical study of the impact of a Wal-Mart supercenter's entry on the sales of a traditional grocery store. We model the two key household decisions of whether to visit the store and in-store expenditure using a flexible model of interpurchase time and basket size. Heterogeneity across households is modeled using a hierarchical structure that allows the response parameters to vary due to observed and unobserved factors. In order to characterize the potential defectors to Wal-Mart, we use a large set of household-specific variables such as distance to the stores, demographics, and other shopping characteristics.

Results show that the incumbent store lost 17% volume—amounting to a quarter million dollars in monthly revenue—following Wal-Mart's entry. Decomposing the lost sales into components attributed to store visits and in-store expenditures, we find that the majority of these losses were due to fewer store visits with little change seen in basket size. This finding suggests that strategies designed to increase store traffic could be effective in mitigating losses to Wal-Mart. We also find that the retailer loses some of its best customers to Wal-Mart, and that a small increase in retention of these customers can significantly reduce losses attributed to Wal-Mart. Thus the retailer should focus its attention on these select households by understanding their needs and preferences. Finally, we find that certain observed household characteristics—such as distance, shopping behavior, and product purchase behavior—can be useful in profiling the defectors to Wal-Mart.

There are, of course, several caveats to our analysis and potential directions for future research. Foremost, our focus in this paper has been on the two broad household decisions of store visit and basket size, while ignoring the basket-composition aspect. Although a Wal-Mart supercenter carries all products typically found in a supermarket, variation in the quality of products (e.g., in produce and meat), as well as the breadth and depth of assortment, can lead to differential impact across departments and categories. Thus, it is conceivable that while the basket size remains constant, the basket composition changes. Since retailers increasingly employ category management tools where each category is treated as a strategic business unit and pricing, merchandising, promotions, and product mix are determined at the category level (Blattberg and Fox 1995), a category-by-category analysis is important to analyze the differential impact across product groups. In doing so, one can draw upon the extensive literature on developing defensive marketing strategies (e.g., the various strategies suggested in the DEFENDER type models, Hauser and Shugan 1983) to enhance category-level retention. Finally, given the asymmetries across grocery retailers and supercenters in costs and product assortments, it may be useful to study their competition from a game theoretic perspective to analyze the optimal response by incumbents to entry by a dominant retailer (Ailawadi et al. 2005).

There are also several shortcomings related to the data used in the analysis. First, we do not observe consumer purchases outside the store in question. Similarly, our analysis is based on expenditures rather than profitability. It is possible that the defectors to Wal-Mart are not only high-revenue customers, but are also more profitable. Finally, another avenue for future research is based on our finding that the majority of the losses at the store are due to fewer store visits. This suggests that it is important for the retailers to figure out the products that are best suited to drive store traffic. Given that the retailer has to choose a subset of 50 to 100 products from a total of over 50,000 unique UPCs, this can be a nontrivial task. However, with better data and advancements in computing power, we hope some of these issues can be addressed in the future.

References

Ailawadi, Kusum, Praveen Kopalle, Scott Neslin. 2005. Predicting competitive response to a major policy change: Combining game-theoretic and empirical analysis. *Marketing Sci.* **24**(1) 12–24.

Allenby, G., P. Rossi. 1999. Marketing models of consumer heterogeneity. *J. Econometrics* **89**(2) 57–78.

BBC News. 2000. Safeway scraps loyalty card. (May 4).

Bell, D., J. Lattin. 1998. Shopping behavior and consumer preference for store price format: Why large basket shoppers prefer EDLP. *Marketing Sci.* (1) 66–88.

Bell, D., Teck-Hua Ho, Christopher S. Tang. 1998. Determining where to shop: Fixed and variable costs of shopping. *J. Marketing Res.* **35** 352–369.

Blattberg, Robert C., Edward J. Fox. 1995. Category management. Center of Retail Management, Northwestern University, Evanston, IL.

Blattberg, Robert C., Scott A. Neslin. 1990. *Sales Promotions: Concepts, Methods, and Strategies*. Prentice-Hall, Englewood Cliffs, NJ.

Brown, Stephen. 1989. Retail location theory: The legacy of Herald hotelling. *J. Retailing* **65** 450–470.

Chintagunta, P. 1993. Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Sci.* **12**(2) 184–208.

Chintagunta, P., S. Haldar. 1998. Investigating purchase timing behavior in two related product categories. *J. Marketing Res.* **35**(1) 43–53.

Corstjens, Marcel, Rajiv Lal. 2000. Building store loyalty through store brands. *J. Marketing Res.* **37** 281–291.

Cox, D. R. 1972. Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

Craig, Samuel C., Avijit Ghosh, Sara McLafferty. 1984. Models of retail location process: A review. *J. Retailing* **60** 5–36.

Fox, E., A. Montgomery, L. Lodish. 2004. Consumer shopping and spending across retail formats. *J. Bus.* **77**(2) 525–560.

Gupta, S. 1991. Stochastic models of interpurchase time with time-dependent covariates. *J. Marketing Res.* **28** 1–15.

Hansen, Karsten, Vishal Singh, Pradeep Chintagunta. 2006. Understanding store brand purchase behavior across categories. *Marketing Sci.* **25**(1) 75–90.

Hauser, J., S. Shugan. 1983. Defensive marketing strategies. *Marketing Sci.* **2**(4) 319–360.

Helsen, K., David Schmittlein. 1993. Analyzing duration times in marketing: Evidence for the effectiveness of hazard rate models. *Marketing Sci.* **12**(4) 395–414.

Hoch, Stephen. 1996. How should national brands think about private labels? *Sloan Management Rev.* **37**(2) 89–102.

Huff, David. 1964. Defining and estimating a trading area. *J. Marketing* **28** 34–38.

Jain, Dipak, Naufel Vilcassim. 1991. Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Sci.* 1–23.

KPFT News. 2002. Wal-Mart's Texas union-busting tactics. (November 15).

Lal, R., C. Matutes. 1989. Price competition in multi-market duopolies. *Rand J. Econom.* **30** 516–537.

Lal, R., R. Rao. 1997. Supermarket competition: The case of everyday low-pricing. *Marketing Sci.* **1** 60–80.

Messinger, P., C. Narasimhan. 1997. A model of retail formats based on consumers economizing on shopping time. *Marketing Sci.* **16** 1–23.

National Grocers Association. 2003. NGA Marketing Survey.

Pauwels, Koen, Shuba Srinivasan. 2004. Who benefits from store brand entry? *Marketing Sci.* **23**(3) 364–390.

Pesendorfer, M. 2002. Retail sales: A study of pricing behaviour in supermarkets. *J. Bus.* **75** 33–66.

Phibbs, C. S., H. S. Luft. 1995. Correlation of travel time on roads versus straight line distance. *Medical Care Res. Rev.* **52** 532–542.

Progressive Grocer, various issues.

Supermarket Business News, various issues.

Reichheld, Frederick F., Thomas Teal. 1996. *The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value*. Harvard Business School Press, Cambridge, MA.

Rogers. 2001. With Wal-Mart, Look, Don't Listen. *Supermarket Business* (15 January).

Rossi, P., R. McCulloch, G. Allenby. 1996. On the value of household information in target marketing. *Marketing Sci.* **15** 321–340.

Seetharaman, P. B., P. Chintagunta. 2003. The proportional hazards model for purchase timing: A comparison of alternative specifications. *J. Bus. Econom. Statist.* **21**(3) 368–382.

Singh, V. 2002. Consumer behavior and firm strategies in a changing retail environment. Ph.D. thesis, Kellogg School of Management, Northwestern University, Evanston, IL.

Time Magazine. 2003. Can Wal-Mart Get Any Bigger? (January 13).

Wall Street Journal. 2003. "Price War in Aisle 3." (May 27).

Wedel, M., W. A. Kamakura, W. S. Desarbo, F. T. Hofstede. 1995. Implications for asymmetry, nonproportionality and heterogeneity in brand switching from piecewise exponential mixture hazard models. *J. Marketing Res.* **32**(4) 457–462.