Impact of Wal-Mart Supercenter on a Traditional Supermarket: An Empirical Investigation

Vishal P. Singh[1] , Karsten T. Hansen[2], Robert C. Blattberg[3]

February 11, 2004

[1]vsingh@andrew.cmu.edu, GSIA, Carnegie Mellon University.
[2]karsten-hansen@northwestern.edu, Kellogg School of Management, Northwestern University.
[3]r-blattberg@kellogg.northwestern.edu, Kellogg School of Management, Northwestern University.

Abstract

Supermarkets operate in an increasingly competitive environment. The rapid growth of alternative retail formats has transformed not only the competitive structure of the industry, but also the way in which consumers shop. The biggest challenge to the industry is coming from none other than the world's largest retailer: Wal-Mart. Although a relatively new player, Wal-Mart through its supercenter format has become the nation's largest grocer and is cited by supermarket managers as their biggest concern in the coming years. Despite the dramatic proliferation of supercenters, relatively little is known about the impact it has on the performance of a traditional grocery store or how it changes consumer buying behavior. This paper provides an empirical study of entry by a Wal-Mart supercenter into a local market. Using a unique frequent shopper database from a supermarket, we study the impact of Wal-Mart's entry on household purchase behavior. The database records purchases for over 10,000 households before and after Wal-Mart's entry. We develop a joint model of inter-purchase time and basket size and allow for a structural break at the time of competitive entry. The model allows us to evaluate the impact of Wal-Mart on household store visit frequency and basket size, while allowing for consumer heterogeneity. We investigate the shopping and demographic characteristics of the consumers that are most likely to shift purchases to Wal-Mart.

Our 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 a much smaller impact on basket sizes. This in turn suggests that strategies designed to increase store traffic could be quite effective in mitigating losses to Wal-Mart. Interestingly, we find that a small proportion of customers account for a large proportion of the losses. For example, 10% of the households account for 45% of the store's lost revenue, while 20% of the customers account for almost 70% of the lost revenue. In terms of consumer characteristics, we find that distance to store while useful, explains little of the variation in household heterogeneity. Households that respond to Wal-Mart are likely to have an infant and pet in the family, and are more likely to be weekend shoppers. Furthermore, we find that these households are large basket consumers, confirming the findings in Bell and Lattin (1998) that large basket buyers are more likely to choose an EDLP operator. On the other hand, households that spend a large proportion of their grocery expenditures on fresh produce, seafood, and home meal replacement items are less likely to defect to Wal-Mart. Implications and strategies for supermarket managers to compete with Wal-Mart are discussed.

Keywords: Entry, Retail Competition, Wal-Mart Supercenter, Frequent Shopper Data

"*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, price 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 the number one player in the grocery industry. Patterned after the European hypermarket, 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, photo development, and gas stations providing consumers with a true "one-stop shopping" experience. While Meijer and Fred Meyer initiated the supercenter concept as early as 1960, it was only with the arrival of Wal-Mart in the 1990's that this format experienced a systematic growth. For example, in 1993 Wal-Mart operated only 10 supercenters, while in the current counting it has close to 1,400 supercenters. Wal-Mart has also chosen the supercenter format as the vehicle for growth with plans to add up to 200 new supercenters every year for the next five years (Company Website).

For an industry already crowded with many players, there are various reasons why Wal-Mart supercenter poses an extraordinary challenge. As discussed in Section 2 below, Wal-Mart has been able to keep its cost below the industry level, which in turn translates into lower prices for the consumers. Given the razor thin margins in the grocery industry, these everyday low prices are difficult, if not impossible, to match. Indeed, as quoted in a recent Wall Street Journal (27 May, 2003), items at Wal-Mart cost 8% to 27% less than Kroger, Albertsons or Safeway, including 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 got into 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 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. Not surprisingly, Wal-Mart supercenter is seen as a serious menace to the traditional grocery industry with 80% of supermarket managers citing competition from supercenter as their biggest concern in the coming year (National Grocers Association, 2003 Point of Impact Survey).

Despite their unprecedented growth and the threat they pose to traditional grocery industry, relatively little is known how entry of a supercenter in a market changes consumer purchase behavior or what it does to the bottom-line of an incumbent supermarket. While there have been a number 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 1994, Pesendorfer 2000), or on stores that differ only in terms of their pricing formats, i.e., EDLP vs. Hi-Lo (Bell and Lattin 1998, Lal and Rao 1997, Messinger and Narasimhan 1997). With minor exceptions (Fox et al. 2002, Singh 2003), there is limited attention given to alternate retail formats like mass merchandisers or supercenters.

This paper provides an empirical study on 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. Our primary focus in this paper is on changes in consumer purchase behavior following the competitor entry. At the household level, entry of a discount store in the market can influence their 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. Yet another group of customers 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. In this paper, we decompose the total observed losses in sales into components attributed to customer attrition, reduction in store visit frequency, and reduction in basket size. From a managerial perspective, such a decomposition can be quite useful in understanding the source of lost volume, and in developing store level marketing policies. For instance, suppose we find that the lost sales are primarily because households do not visit the store as frequently as they did prior to Wal-Mart's entry but once in the store, their basket size remains constant. In this case, households split their shopping trips between the incumbent and Wal-Mart. This in turn suggests developing strategies primarily geared towards generating store traffic, for example via feature advertisements. On the other hand, suppose we find that the frequency of store visits remains constant but the basket size is smaller. In this case, the focus should be on in-store merchandising to increase expenditure once the customers are in the store.

A second objective of this paper to explore the heterogeneity in consumer response to Wal-Mart and determine the characteristics of the households that are most likely to respond to Wal-Mart. This heterogeneity in response could be driven by household specific unobserved factors such as a like or dislike of shopping at a discount store. At the same time, it is conceivable that household behavior is also driven by their observed characteristics. For instance, a general finding in the retail site selection literature (Huff 1964, Brown 1989, Craig et. al. 1989, etc.) as well as business press (Progressive Grocer 69th Annual Report, 2002) is that location is one of the most important factor in determining store choice. Similarly, Fox et al. (2003) find that larger households are more likely to frequent mass merchandisers. In this paper, we explore a large set of household specific variables such as distance to the store and demographics, and other variables related to their shopping behavior (for example weekend shopper) and product purchase behavior (for example, proportion of expenditures on fresh seafood, proportion of expenditures on store brands etc.). Note that from a managerial perspective, identifying households based on their observed characteristics 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 it comes in competition from Wal-Mart (or other stores). Since store opening information is generally available well in advance, preemptive actions can be taken for the households who are at high risk of defection.

The data for this study comes from a store located in a small town in the East Coast region of the US. The sub-urban nature of the store location provides us with an opportunity to measure the impact of Wal-Mart's entry in a relatively controlled environment[1]. The store in question has a well developed frequent shopper program with over 85% of the sales captured on shopper card. The database records all transactions made in the store, and captures information such as time and date of the transaction,

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

price, promotion, and quantity for every UPC sold. This information is recorded at the individual level for all the customers in the store, provided they use their shopper card. 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 store. Thus, we observe reasonably long purchase history both before and after Wal-Mart's entry. Another useful information in the database is the mailing addresses for the card holders. These addresses were geo-coded to compute travel distance for each household to the incumbent store, and to Wal-Mart. In addition, we also created a number shopping related variables such as weekday versus weekend shopper, store-brand buyer, etc. that could potentially be useful in characterizing the households. Finally, we supplement the data with block level demographic variables.

To evaluate Wal-Mart's impact on consumer purchase behavior, we develop a joint model of inter-purchase time and basket size. While a popular approach to model inter-purchase time used in the marketing literature is the proportional hazard model in continuous time (for example, Jain and Vilcassim 1991), it has the limitation of only accounting for marketing mix and other covariates when an event occurs (for example, 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 where households do not make a purchase. We take a discrete-time approach and model 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 et al. 2003). Besides accounting for the full time path of the covariates, this modeling approach has the advantage of allowing for non-proportional hazards – a feature which, as shown in section 5, is empirically relevant for our data.

Consumer's in-store expenditures are modeled using a semi-log specification which has been used extensively in marketing (See for example, Blattberg and Neslin 1989). Both 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, which is captured by using a hierarchical structure. In particular, the full set of model parameters are allowed to vary across consumers due to both observed (e.g. demographics) and unobserved factors.

For inference, we use a hierarchical Bayesian approach. In particular, we use a Markov Chain Monte Carlo procedure to simulate the posterior distribution of the model parameters and to compute household level estimates of preferences. As discussed in Allenby in Rossi (1999), Bayesian procedures are well suited when one is interested in making inference at the individual level.

Our results show that the incumbent store lost 17% 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%. Decomposing the lost volume into store visits and in-store expenditures, we find that the majority of the losses came due to fewer store visits with little impact on the basket size once consumers are in the store. This is an important finding as it suggests that strategies designed to drive store traffic could be an effective way for recovering some of the lost volume. Interestingly, we find that a small proportion of customers account for a large proportion of the losses. For example, 10% of the households account for 45% of the store's lost revenue, while 20% of the customers account for almost 70% of the lost revenue. In terms of consumer characteristics, we find that distance to store while useful, explains little variation in household heterogeneity. Households that respond to Wal-Mart are likely to have an infant and pet in the family, and are more likely to be weekend shoppers. Similarly, store-brand buyers have a higher likelihood of moving purchases to Wal-Mart, while households that spend a large proportion of their grocery expenditures on fresh produce, seafood, and home meal replacement items are less likely to so. Finally, we find that the households that respond to Wal-Mart are large basket consumers, confirming the findings in Bell and Lattin (1998) that large basket buyers are more likely to choose an EDLP operator. Using a holdout sample, we find that these shopper characteristics can be quite useful in identifying potential defectors to Wal-Mart.

This research makes several contributions to marketing theory and practice. Foremost among these is that the we provide an empirical analysis of the impact of Wal-Mart supercenter entry on a traditional retailer. 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 small proportion of customers account for a large proportion of store losses give credence to the contention in the CRM literature on the importance of customer retention. Similarly, our analysis demonstrates how a retailer can exploit the information contained in their frequent shopper database to understand and respond to their 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 (UK) which abandoned its customer card program citing a potential saving of $80 million per year in administrative costs. 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 the 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. The data used in the study is presented in Sections 3. Section 4 develops a joint model of interpurchase time and basket size and the empirical results from the model are presented in Section 5. In section 6 we explore various household characteristics that can be useful in identifying potential defectors to Wal-Mart. We conclude in section 7 with a discussion on limitations of the current study and directions for future research in this area.

2 The Supercenter Format[2]

In this section we provide a brief overview of the supercenter format. 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. Since this format has received limited attention in the academic literature, 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. While Meijer and Fred Meyer started this format as early as 1960's, it is only with the arrival of Wal-Mart that this format has shown a 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 has close to 1,400 supercenters in the current counting. This unprecedented march by Wal-Mart into the grocery business is taking

[2] We shall limit our discussion primarily to Wal-Mart. Both Target and K-Mart have their own version of supercenter, but Wal-Mart is by far the biggest player in the industry.

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 percent while visits to supercenters went up 40 percent. The pressure from Wal-Mart is being felt by national chains and independents alike. In the past decade, 29 chains have sought bankruptcy-court protection, with Wal-Mart as a catalyst in 25 of those cases (The Wall Street Journal, May 27, 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 arrival of David Glass as the CEO (who had a background in grocery business). Furthermore, by the late 80's, the discount retail industry was close to saturation, and was highly concentrated with three major players: Wal-Mart, K-Mart, and Target. Supermarket industry on the other hand, was highly fragmented with small to medium sized regional chains. While 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 as mainly a traffic driver, with the hope of spillover to higher margin general merchandise items (that 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 that of discount stores in the same area (or after conversion of a discount store to supercenter)[3]. The supercenter format has been so successful that Wal-Mart has chosen this path for expansion with plans of adding 200 supercenters every year for the next five years (Company Website). According to Trade Dimensions, with the current growth rate, over three-fourth of Kroger and Albertsons stores would 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[4]. 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 the wholesalers with majority of the merchandise at the supercenters, including perishables, supplied through

[3]"The U.S. Market for Supermerchants: Warehouse Clubs and Supercenters", 1997, report by Packaged Facts

[4]The WSJ article cited above quotes 8% to 27% lower prices at supercenters. A Time Magazine article "Can Wal-Mart get any bigger" (Vol. 161, issue 2, 2003) reports 15% lower prices.

its distribution centers. This coupled with an EDLP strategy (which not only helps create a low-price image in 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[5]. Last, but not the least, another factor keeping the costs low at Wal-Mart is its nonunionized labor. For majority of supermarkets, labor, that constitutes approximately 70% of the overhead, is unionized. On the other hand, none of the Wal-Mart's employees belong to a union and industry analysts believe that they get paid significantly lower than the industry average[6].

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? While there is no one answer, there have been a number of suggestions given by industry experts. 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[7]. Indeed, there has been a move in the supermarket industry towards consolidation through mergers and acquisitions (e.g., Kroger and Fred Myers, American and Albertson stores, etc.) with the hope leveraging similar bargaining power and economies of scale as Wal-Mart. Similarly, there has been a drive in the industry towards cost cutting (WSJ, 27 May, 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 that 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 department, emphasis on deli, ready-to-eat foods,

[5] "Can Wal-Mart get any bigger" Time Magazine (Vol. 161, issue 2, 2003)

[6] 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 decided to close down the meat plant and replace it with case-ready meat packaged by suppliers.

[7] David Rogers "With Wal-Mart, Look, Don't Listen", Supermarket Business, Jan 15, 2001

and salad bars, broader product assortment, and an increased focus on understanding customer needs. Another weakness of the supercenter 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 location convenience of supermarkets and convenience stores.

Given the discussion above, it is not surprising that supermarket managers find Wal-Mart to be one of their most formidable competitors. Many of the suggestions provided by industry experts 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 retailers 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 comes from a single store of a large supermarket chain in the East coast. The store in question is located in small suburban town which provides us with an opportunity to analyze the impact of entry by a Wal-Mart supercenter in a relatively controlled environment. 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.

The store also has a well developed frequent shopper program. While the original purpose of such frequent shopper programs was to create store loyalty by rewarding the best customers, over time, their role 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. Indeed, the frequent shopper data that we use are quite 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 in that (unlike typical scanner panel data) the purchase information is only 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[8]. While this may seem like a major shortcoming, one must realize that this is the information that is typically available to the retailer (unless they purchase data from outside vendors like IRI or ACNielsen).

As seen in Table 1, the usage of the shopper card program 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 non-card purchases account for 32% of all transactions, the average order size is significantly lower (average of $8 compared to $34 for transactions using the shopper card). The non-card 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%.

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. Purchase history information is available for over 22,000 card members. However, several of these cardholders are casual buyers with very few purchases at this store. Our approach in this paper is to use data for the top 10,000 customers based on their total expenditure prior to Wal-Mart's entry. The selected customers account for 77% of card member sales. Our focus on top 10,000 customers is in line with the literature on customer relationship management that advocates understanding the behavior of the best customers (e.g. Reichheld and Teal, 1996)[9].

3.1 Customer Location

Another 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 Wal-Mart. Based on the findings in the previous literature (see for example research on retail site selection Huff 1964, Brown 1989, Craig et. al. 1989, etc.) as well as business press (Progressive Grocer 69th Annual Report, 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 around the

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

[9]While ideally one would want to sort customers in terms of their profitability rather than spending, we do not observe costs in our data and thus rely on aggregate expenditures.

store. The location of the focal store is shown by the large star while the location of Wal-Mart is shown by the large square. On average, consumers live about 3.5 miles from the focal store and 4.8 miles from Wal-Mart[10]. As is evident from Figure 1, a large number of the customers are clustered around the focal store. However, despite the apparent proximity, over one-third of the 10,000 best customers considered in our analysis live outside the 3 mile radius (considered as the trading area of a typical grocery store).[11] 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.

The addresses were also used to collect a variety of census block-group demographic data. 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. Similarly, certain block-groups are characterized by very young customers (median age of 19) while other have a median age of over 66. Details on the demographic data as well as their usefulness in determining the likely reaction to Wal-Mart entry are discussed in section 6 below.

3.2 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 decision to visit the store and basket size once in the store. However, creating variables to capture the overall store-pricing environment is a non-trivial task. The store carries over 50,000 unique UPCs that are classified into several hundred categories. Furthermore, unlike the packaged goods typically used in marketing, several of these products (for example 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 very difficult, if not impossible. The matter is further complicated by different price reactions by the incumbent to Wal-Mart's entry in

[10]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 .987 between straight line distance and travel time, although this correlation drops to .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, Ho and Tang 1998).

[11]66% of the households live within the 3 mile radius of the focal store and 78% live within a 5 mile radius.

various product categories. In Figure 2 we plot the pricing series three products: Banana, Gallon of store brand spring water, and loaf of Italian bread. It is evident that the initial reaction in the Banana is to match the 39 cents/lb price of Wal-Mart with occasional 25 cent/lb promotions[12]. Over time the prices went back to 59 cents/lb. 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[13].

Given the above mentioned complexities, any measure to capture the overall store price environment in any given week would be an approximation at best. Our strategy is therefore to rely on aggregate measures that are designed to proxy the overall price environment of the store on a given week. In particular, we constructed promotion indexes for the 5 major departments– Grocery, Health and Beauty (HBC), General Merchandize (GM), Produce, and Meat. These measures were constructed as follows: for each department and each price cycle (prices are set every Sunday on a weekly basis) we aggregated the dollars off the regular price for every product falling in that department. For instance, suppose the produce department consists of the following 3 products with regular and promoted price as: Banana ($0.55/lb, $0.45/lb), Broccoli ($1.39/lb, $1.39/lb) and Red grapes ($1.00/lb, $0.90/lb). The promotion index for produce in that week would be $0.20. We take this as a proxy for the department level promotions. In addition, we also experimented with various measures of basket price (for instance, a sum of the price for the top 100 products). However, these measures revealed little variation over time. Table 2 shows that the average promotions in every department increased after Wal-Mart's entry. Not surprisingly the department with the largest promotional volume is Grocery. Note also that the promotional indices exhibit a good amount of variation over the sample period, ranging from 20 percent of the mean value for groceries to about 75 percent of the mean value for meat.

3.3 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

[12]Price of .39 cents/lb of banana is a heavily advertised item at Wal-Mart supercenter in their advertisements as well as in-store special displays.

[13]Our discussions with the store manager did not reveal any particular insight in 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 on emphasizing the fresh produce, seafood, meat, and deli items–a point that we return to in section 6 below.

by the vertical line. The two spikes early on in the data and around day number 400 are the Thanksgiving and Christmas weeks. Two key observations from the figures must be highlighted. First, there seems a significant fall in the baseline volume for the incumbent store. This is particularly troubling considering that supermarkets generally operate on a principle of low margins and high volume, with profit margins of only about 1%. 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 are 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 the two fundamental household decisions: decision to visit the store and basket size once at the store. The model also allows for heterogeneity across households due to both observed and unobserved factors.

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, entry by a competing store 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, e.g., 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 then to estimate the expected value of quantities like $\Delta_1 \equiv V^W - V$, $\Delta_2 \equiv (V^W - V)/V$ 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 of 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 be different across consumers due to both observed (such as demographics) as well as unobserved factors. It is crucial for the store to understand what "types" of consumers/households display the biggest change in store expenditure and what the causes are. 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) or smaller basket size per trip, or a combination of the two? Finally, if the store environment changes after Wal-Mart's entry, e.g., 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. Table 2 shows that this is indeed the case in the sample.

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 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 post Wal-Mart entry, and letting x_t denote variables describing store environment we can now define "pure" Wal-Mart effects at the individual and aggregate level by holding x_t fixed. For example,

$$\mathrm{E}[nt_h^W|x_t] - \mathrm{E}[nt_h|x_t], \quad \text{and} \quad \mathrm{E}[\log b_{ht}^W|x_t] - \mathrm{E}[\log b_{ht}|x_t], \tag{4}$$

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

4.1 Interpurchase Time and Basket Size

We model the two consumer decisions jointly using a flexible model of inter-purchase time (to capture when to visit the store) and semi-log regression (to capture basket size once at the store) and allow for a structural break at the time of competitor entry. Heterogeneity across households is captured by using a hierarchical structure where the parameters are allowed to vary across consumers due to both observed and unobserved factors.

Over the past two decades, a number of models have been proposed to capture the purchase-timing decisions of households (see Seetharaman and Chintagunta 2002 for a recent 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 (Jain and Vilcassim 1991, Chintagunta and Halder 1998 etc.) 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 where household 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 where 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 process. An advantage of using this approach is that we can use a very flexible specification for duration dependence that allows us to approximate any shape of the household specific hazard function. In proportional hazard models, like 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 non-proportional hazard funtions.

We take a discrete-time approach and use "days" as the basic time unit[14]. In each time period the individual decides on whether or not to visit the store and make a purchase.[15] Let $U_i(t)$ be (indirect) net utility for individual i of making a purchase from the store in period t. Assume that

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

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 which can be made as as flexible as desired. For instance, we could have $f(\tau_{it}) = \left(1,\ \tau_{it},\ \tau_{it}^2, \ln \tau_{it}, ...\right)'$. 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.[16]

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

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

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

[14] 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.

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

[16] Note that by interacting τ_{it} with P_t we can allow for more general duration dependence, e.g., capture the notion that promotions' effectiveness depends on time since last purchase.

right hand side parameters and variables as (X_{it}, β_i),these durations then have likelihood

$$\Pr(T_{i1} = t_1, T_{i2} = t_2 | \beta_i, X_i^{t_1+t_2}) = \left\{ \prod_{t=1}^{t_1-1} \Pr(D_{it} = 0 | X_{it}, \beta_i) \right\} \times \Pr(D_{it_1} = 1 | X_{it_1}, \beta_i) \times \left\{ \prod_{t=t_1+1}^{t_1+t_2-1} \Pr(D_{it} = 0 | X_{it}, \beta_i) \right\} \times \Pr(D_{i,t_1+t_2} = 1 | X_{i,t_1+t_2}, \beta_i). \tag{7}$$

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

Post Wal-Mart entry, the utility is assumed to be:

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

where $\beta_{i0,W}$ captures the impact of competitor entry.[17]

To model the basket size once the household is in the store, we use a semi-log specification which has been used extensively in marketing for modeling sales and expenditures (see for example, Blattberg and Neslin 1989). 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

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

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 $\varepsilon_{e,it} | v_i \sim \mathrm{N}(0, v_i^{-1})$. Post Wal-Mart entry, the log basket size is modelled as:

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

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

4.2 Heterogeneity

Since we expect different households to react differently to Wal-Mart's entry, it is important to account for consumer heterogeneity in the model parameters. Furthermore, we

[17]More advanced specifications can easily be entertained. For example, it is possible to allow for change in sensitivity to marketing mix variables by letting $\beta_{ip} + \beta_{ip,W}$ be the new sensitivity vector after entry. Another generalization is to allow for both a short run and long run effect of Wal-Mart. This can be accomplished by adding a short-run dummy to the specification, e.g., a dummy equal to one in the first 60 days after entry. This could capture "curiosity effects" which differ from the long run effect.

also expect the responses by households to be related to their observed characteristics such as demographics. In this paper we use a parametric approach to model household heterogeneity. The model parameters are allowed to 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 that θ_i follows a multivariate normal distribution with a mean vector ΠZ_i and covariance matrix Ω :

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

where Z_i is a vector containing household characteristics. For ν_i we assume a Gamma distribution $G(\nu_i|a, b)$ with parameters (a, b).

For inference, we use a hierarchical Bayesian approach. In particular, we use a Markov Chain Monte Carlo procedure to simulate the posterior distribution of the model parameters and to compute household level estimates of preferences. As discussed in Allenby in Rossi (1999), Bayesian procedures are well suited for these models especially when one is interested in making inference at the individual level. Since these procedures have become quite standard in the literature, we outline the estimation algorithm in the appendix.

Using the estimated model we can simulate expenditures over a certain period (e.g., a month) and break expenditures down into number of trips and basket size, both at the household level and at the store level. We can do this holding promotions fixed to estimate causal Wal-Mart effects. We can also simulate the effects of various targeted promotion policies and gauge to what extent these can be used by the store to fight Wal-Mart.

5 Estimation Results

We estimated the model described above using the algorithm outlined in the appendix. The marketing mix vector p_t consisted of the five promotion variables discussed above.[18] The function f was, after some experimentation, 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.

[18]The promotion variables were each normalized to have mean 1 over the estimation period. The standard deviation for the normalized promotional indices were 0.57 (produce), 0.75 (meat), 0.36 (hbc), 0.47 (general merchandise) and 0.19 (grocery).

The variables included in the second stage of the hierarchical model (11) was $Z_i = \{D^r_{ji}\}^n_{i=1}$, where D^r_{ji} is equal to 1 if household i is located in census tract region j. The average composition of household characteristics as well as average distance to the focal store and Wal-Mart distance by region is shown in table 3. The regions show considerable variation in both distance and average income, ranging from regions very close to the store (region 15, 10 and 17) to regions close to Wal-Mart (region 18) to regions far away from both stores (e.g., region 13). Note that while demographics are only available to us at a block-group level, we do have the exact distances to the two stores at the household level. Standard specifications of models of site choice typically include distance (and sometimes distance squared) to each site as an approximation of travel costs. While we could follow this approach, we chose the dummy specification for several reasons. First, the correlation between distance to the focal store and distance to Wal-Mart is .96 in our sample.[19] This makes it impossible to estimate two separate distance effects. Second, we use a fairly fine grid of regions in our dummy specification and this provides us with a clean "non-parametric" estimate of the distance effects. Finally, the dummy specification allows us to capture other census tract-specific characteristics apart from distance (income, age, education etc.).

In Table 5 we report estimates of the hierarchical coefficients Π and the diagonal elements of Ω in (11) for the coefficients relating to the store trip model (5) and (8). Recall from above that the Z-variables in the model are dummies for 19 different residential regions. So the dummy coefficient estimates may be interpreted as the mean coefficient for the corresponding region. It is evident that there is variability in mean coefficients across regions, implying that the mean store visit frequency varies across region. Since it is hard to directly interpret the coefficients to the variables which are functions of τ ("time since last purchase"), we plot the implied distributions of store visit times. Figure 4 plots the distribution of inter-store-visit times prior to Wal-Mart's entry for the three regions 3, 9 and 15.[20] Note from table 3 that region 15 is located only 0.39 miles from the store, while region 3 is located 7.69 miles from the store. Region 9 is in between – located 2.16 miles from the store. It is evident from the figure that store distance is an important factor in determining the average inter-store-visit times in a region. As

[19]The reason for this is obvious: Since the households live in a wide area around the two stores, distances will tend to be correlated. While one could alleviate this problem to some extent by restricting the sample to households living in a small radius around the two stores, this will throw away a substantial part of the sample. We are specifically interested in solving the problem facing the retailer and so we will not artificially restrict the sample.

[20]These distributions were computed holding the promotional indices fixed at their in-sample mean value.

expected, longer distance to the store implies longer average inter-store-visit times. All promotion coefficients are either positive or (not significantly different from) zero. An increase in all promotion indices by one standard deviation increases the latent index in the duration model (i.e., mean utility value in (5)) from .027 for region 1 to .062 for region 4. By computing the expected number of visits per households under the two promotional levels, these coefficients translate into an overall effect on aggregate monthly store visits equal to 6.7%.

In Figure 5 pre- and post-Wal-Mart distributions of inter-store-visit times are displayed for regions 15 and 18. As noted above region 15 is located only 0.39 miles from the store while region 18 is the region closest to Wal-Mart (households in region 18 has an average distance to the store equal to 2.27 miles and an average distance to Wal-Mart equal to 0.82 miles). The average inter-store-visit time increases for both regions after Wal-Mart's entry (since probability mass shifts from smaller duration times to larger). However, the Wal-Mart effect is clearly much stronger for region 18 – the region closest to Wal-Mart.

As noted above, our estimation procedure also gives us a set of household level coefficient estimates (for example, the posterior mean of θ_i for each household). Figure 6 shows the effects of a promotion on the estimated hazard for household nr. 3572 in the sample. For the first hazard in Figure 6(a) (labelled "Low Promotion") the promotional indexes is equal to 1. For the second hazard (labelled "High Promotion") the promotional indexes are all increased by one standard deviation. As expected the "High promotion" hazard shifts up – implying a reduced expected inter-store-visit time (Figure 6(c)). Figure 6(b) shows the ratio of the hazards in Figure 6(a). If the proportionality assumption assumed for standard proportional hazard models were satisfied this ratio should be independent of time, i.e., a straight line. Clearly, this is not the case. This demonstrates the value of the non-proportionality assumption. In addition, we can easily simulate the effects of different time-paths for the promotional variables.

Table 6 shows estimates for the basket size equation (9) and (10). There is substantial variation in the average basket size across regions (as reflected in the intercept). The coefficient to the Wal-Mart dummy is negative for all regions and may be interpreted as the average percentage drop in basket size after Wal-Mart's entry (holding promotions constant). The effect varies from 3 to 10 percent. The smallest change in found in region 15 (the region closest to the store) while the biggest effect is for region 18 (the region closest to Wal-Mart). The coefficient to τ (time since last purchase) ranges from 1 to 2 percent. So the basket size on average increases by about 1-2 percent per day since last purchase. An increase in all promotion indices by one standard deviation increases the basket size from 1.3% for region 1 to 5% for region 2 and 18.

The above results indicate that both store visit duration times and basket sizes are affected by Wal-Mart's entry and that average effects vary by region. To quantify the combined effect of entry on duration times and basket size by region, we simulated expected monthly expenditure for each household pre- and post-Wal-Mart entry using the household level estimates, again holding the promotion level fixed. This amounts to computing the expected value of (2) for $T = 30$ for each household, before and after Wal-Mart's entry. In Figure 7 and 8 the resulting distribution of relative change (i.e., percentage change) (Figure 7) and absolute change (Figure 8) in mean expenditure is shown. The average effect for the whole sample was minus 19 percent for the relative change distribution (shown by the plus sign in the figure) and minus \$25 for the absolute change distribution. The overall effect (the sum of all effects in Figure 8) is -\$244,405. This amounts to 17% of the pre-Wal-Mart monthly store volume. However, both figures show a wide distribution around this mean (ranging from about minus to plus 100 percent change for the relative change distribution and from about minus \$350 to \$150 for the absolute change distribution). How much of this variation can be explained by regional characteristics? In table 7 we show the overall effect on average monthly expenditure by region as well as the effect on monthly number of trips and basket size. For example, the average percentage change of households' monthly expenditure in region 1 was minus 24 percent. This drop came about in part from an average reduction in number of trips per month equal to 20 percent. The average drop in basket size for region 1 was 6 percent[21]. Table 7 shows that the average effect of entry on monthly expenditure varies from minus 15 percent (for region 6 and 17) to minus 37 percent (for region 18). The average effect for all regions is, as noted above, minus 19 percent. The main cause of the change in expenditure is a reduction in the number of trips per month. Compared to trip frequency the drop in basket size is small. Table 7 shows another important feature of the Wal-Mart-effect distribution: The regional mean is associated with a substantial amount of within regional variation. This should come as no surprise in light of the results shown in Table 5 and 6: the amount of total variation in heterogeneity explained by the regional dummies (shown in the column labeled "Z-fraction") ranges from 1 to 6

[21]For this exercise, the basket size at the household level was calculated as $bsize = \mathrm{E}[V|\hat{\theta}_h]/\mathrm{E}[nt|\hat{\theta}_h]$, where V is monthly expenditure, nt is number of trips in a month and $\hat{\theta}_h$ is the household level estimate for household h. Since this is a non-linear function of expected expenditure and expected number of trips, the percentage change in number of trips and basket size will not necessarily add up to the percentage change in expenditure.

percent for the duration model and from 1 to 10 percent for the basket model. So we should not expect to be able to explain the variation in the Wal-Mart effect very well by the regional dummies as is evident by the large amount of with-in cell variation in Table 7. We can conclude that while mean effects do vary across regions, there is still a considerable amount of unexplained variation left.

The above results indicate that it is hard to predict which households are affected the most by Wal-Mart's entry using households' neighborhood characteristics only (such as store distance and Wal-Mart distance). This raises the obvious question of how one might then explain the variation in Wal-Mart effects. As a start, we can dig deeper into the distribution of entry effects. Figure 7 and 8 show a considerable amount of heterogeneity in entry effects. To probe deeper, we define the total entry effect as the change in monthly expenditure for all households and decompose it as follows:

$$\text{entry effect} = TL_{-} + TL_{+} \equiv \sum_{h:\Delta_h<0} \Delta_h + \sum_{h:\Delta_h>0} \Delta_h, \qquad (13)$$

where Δ_h is the change in expected monthly expenditure for household h. The total loss is the sum of those households' expenditure change which is negative plus the sum of those which is positive. The total entry effect is ,as mentioned above, $-\$244,405$, while TL_{-} is $-\$289,429$. A natural question is to what extent a small fraction of households are responsible for the majority of TL_{-}. Figure 9 shows the "Lorenz curve" for the effects distribution for those households who contribute to TL_{-}. This figure confirms that the lower tail of the effects distribution contributes a substantial amount to the overall effect. In particular, the bottom 10 percent of households in the effects distribution is responsible for about 30 percent of TL_{-} (and 45 percent of the total store losses). The bottom 20 percent is responsible for about half of TL_{-} (69% of total store losses). This suggests that – as a beginning – it might be of interest to try to profile the households in the very left tail of the effects distribution.

6 Profiling the Defectors to Wal-Mart

In the previous section we found that distance to the store (based on the regions that household live in) while useful, explain little variation in heterogeneity of response. We also found that a small proportion of households seem to account for a large proportion of the observed losses at the store level. In this section we explore the extent to which the households that respond most to Wal-Mart can be profiled based on their parameter estimates (such as promotion sensitivity) and other observed household 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 it comes in competition from Wal-Mart (or other stores). 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 to not only identify the households based on their individual card numbers or the region they live in, but to also map these households on some actionable demographic or other characteristics, which then becomes the basis of 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 high risk of defection.

6.1 Household Level Parameter Estimates

We test the extent to which household level coefficients of the model described above is informative about the Wal-Mart effect on store expenditure using the following:

- Pre-Wal-Mart basket size (as reflected in the intercept of the log-basket equation λ_{i0} in (9))
- Pre-Wal-Mart trip frequency (a function of the parameters in the duration model),
- Pre-Wal-Mart monthly expenditure (a function of both the basket and duration model) and
- Promotional sensitivity.

In Figure 10(a) we show the average value of each of these four measures by decile of the effects distribution. For example, Figure 10(a) shows the average value of the log basket size intercept for each decile of the effects distribution. A strong relationship between location in the effects distribution and basket size is evident: The average log basket size for households who are associated with the largest loss to the store (the 0-10% group) is the biggest of all. The average log basket size falls roughly linearly with decile (though the average intercept increases slightly for the 90%-100% group). The trip frequency plot (fig. 10(b)) is U-shaped, implying that both the far left tail and right tail of the effects distribution have higher trip frequencies than the middle of the distribution. The 0-10% group has roughly the same trip frequency as the 90% - 100% group. The 0-10% group has the largest pre-Wal-Mart expenditure (fig. 10(c)) and average expenditure is falling up to the 50%-60% group after which it increases. Sensitivity to grocery promotions (fig.10(d)) displays a monotonic relationship with the percentiles of the effects distribution with the 0-10% group being the least sensitive to promotions and the 90%-100% group the most sensitive.

These findings indicate that households who are large basket shoppers and are non-sensitive to grocery promotions are more likely to undergo a large negative change in their monthly expenditure (at the focal store) than the remaining population. In addition, households who are located in the left tail of the effects distribution have higher trip frequencies and monthly expenditure than households located in the middle of the distribution. These are plausible findings: Large basket shoppers who are non-sensitive to promotions in their trip-frequency, will in general gain by moving from a Hi-Lo pricing environment (the focal store) to an EDLP store (Wal-Mart). On the other hand, small basket shoppers who are sensitive to promotions will benefit from staying with the focal store and take advantage of promotions when they occur.

To demonstrate the value of these calculations in order to pin-point household segments with large negative changes in expenditure, Figure 11 shows the effects distributions for two segments of roughly 500 households each. These segments were determined by first sorting households into a 5×5 matrix with cells corresponding to quintiles of the distribution of log basket size intercepts and grocery promotion sensitivities. We then plotted the distribution of effects for household with highest promotion sensitivities and smallest basket sizes, and the distribution of effects for household with smallest promotion sensitivities and biggest basket sizes. This simple segmentation strategy – using parameters which can be estimated on pre-Wal-Mart data – is able to separate out a shopper-segment who contributes significantly to the overall loss in store volume.

6.2 Observed Household Characteristics

We now turn our attention to observed household variables that can help identify the potential defectors to Wal-Mart. In section 5, we found that distance to the store can be an important factor. However, given the large standard deviations in table 7, there is significant heterogeneity still left to be explained. Consider for example, the hazard functions for three selected households in Figure 12: hh-A lives 0.09 miles from the focal store while hh-B lives 0.21 miles from Wal-Mart. It is evident from the plots that entry of Wal-Mart had very different impact on these two households, confirming the findings above that distance to store can be quite important. Next consider the hazards for hh-B and hh-C which also lives next to Wal-Mart (the households are in fact neighbors living .04 miles apart). Surprisingly, we find that entry of Wal-Mart had little impact on the hazard of household C. However, a deeper probe into the purchase behavior can explain why we observe these different reactions. Prior to Wal-Mart entry, household C is found to visit the store very frequently with average basket size of only \$9 (as opposed to basket size of \$51 for hh-B). More importantly, 77% of the total trips of hh-C occur between 9 a.m and 5 p.m on (non-holiday) weekdays (compared with only 4% for hh-B). This in turn implies presence of a retired or otherwise unemployed person in hh-C that continues to frequent the incumbent store for bargains.

The example above suggests that there are a number of other characteristics that can help identify the households most likely to respond to Wal-Mart. For the purpose, we use a large set of household specific variables, several of which are computed using the transaction history of the households *prior* to Wal-Mart's entry. Table 4 provides the summary statistics on these variables which fall into 3 broad categories:

- **Demographics**: The first four demographic variables in Table 4 (*% Rent*, *Income, Age and HH Size*) refers to the percent of population living in rented dwellings, the median income, median age, and median household size for the census block group that the household belongs to. The next two demographic variables (*Baby* and *Pet*) are dummies indicating the presence of a baby and pet. These were computed from the household purchase history data. For example, if a household is observed to purchase baby products such as diaper or baby food it indicates presence of an infant in the family. Similarly purchase of dog food or cat litter indicates presence of a pet.

- **Shopping Variables**: We use the time and day of trip information to construct two variables that relate to household shopping behavior: *% of trips on Weekend* (the percent of trips that a household makes on a weekend) and *% of trips between 9-5* (the percent of total visits that were made between 9 and 5 on weekdays excluding holidays). These two variables can be treated as proxies for shopper's search cost. For instance, if a household is observed to make majority of their purchases between 9 and 5 on the weekdays, it suggests presence of retired or otherwise unemployed member in the household.

- **Product Purchase Behavior**: The last set of variables were created using household purchase behavior in different product types (again created using transactions prior to Wal-Mart entry). The primary motivation for these variables comes from business press reports that argue that one of the major weakness of Wal-Mart supercenter is in the fresh food area. As discussed in section two, Wal-Mart supercenters primarily rely on pre-packaged produce and meat from supplier. Thus households that allocate a large proportion of their expenditures to fresh produce (*% Expenditure on Produce*), fresh meat (*% Expenditure on Meat*), and fresh seafood (*% Expenditure in Seafood*) 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 (*% Expenditure on HMR*: proportion of expenditure in home meal replacement) could result in higher affinity for the store compared to Wal-Mart. Finally, *% Private* refers to proportion of total expenditures on the store brands[22]. 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) suggest that store brands can create store loyalty. At the same time, researchers have also found store brand buyers to be more price sensitive (Hoch 1996, Hansen et al 2003), and if more price sensitive households frequent discount stores then we can find the opposite effect.

To test the usefulness of these variables in profiling households that are most likely to respond to Wal-Mart, we construct the following indicator variable:

$$DEFECT = 1 \text{ for 20\% of households with largest } \Delta_{Expenditure} \quad (14)$$
$$= 0 \; otherwise \quad (15)$$

where $\Delta_{Expenditure}$ refers to the change in household's monthly expenditure following Wal-Mart's entry[23]. We then run a simple probit regression as a function of household characteristics described above. Note that instead of taking this 2-step approach, we could have directly incorporated these demographic variables in the model in section 5 (as additional "Z" variables). However, several of these constructed variables are potentially endogenous, and could have biased the model estimates.

The results are presented in Table 8. These results are for a sample of 8,000 cardholders, with another 2,000 kept as a holdout sample. All of the demographic variables from the census data are found to be insignificant. On the other hand, the two demographics

[22]Excludes non-packaged goods like items in meat and produce departments.

[23]We repeated the exercise using the bottom 10, 15, and 25% of the households and found similar results.

constructed using the transaction data are highly significant and show that households with an infant or pet are more likely to respond to Wal-Mart. Similarly, we find that household specific shopping or product purchase variables are quite useful in identifying households that move to Wal-Mart. For instance, Table 8 shows that households that shop mostly on weekends are more likely to respond to Wal-Mart, while households that shop during 9 am and 5 p.m. on weekdays are less likely to do so. Similarly, households that spend a large proportion of their in-store expenditures on fresh produce, seafood, and home meal replacement items are less likely to move to Wal-Mart. In terms of store-brand purchases, we find that store-brand buyers have a higher likelihood of moving purchases to Wal-Mart. This 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). Finally, confirming the findings above and those found in the previous literature (Bell and Lattin, 1998), we find that large basket consumers are more likely to choose an EDLP operator.

To assess the usefulness of the households variables for targeting the likely defectors to Wal-Mart, we compute the predicted probabilities (based on estimates above) for a hold-out sample of 2,000 cardholders. Note that 400 of these households are in the 20% group that reduced the expenditure the most (i.e., for who the indicator variable DEFECT takes value 1). The objective of the retailer is to identify these 400 households based on their observed characteristics. The gains chart reflecting the information contained in the household variables is presented in Figure 13. The 45-degree line represents the performance if targeting is done at random. Gains due to information in the household variables is reflected in the extent to which the line lies above the 45-degree line. For example, the chart shows that by targeting the top 20% based on predicted probabilities using the (above mentioned) household variables, the retailer would be able to capture 46.4% of the targets. By contrast, 20% picked up at random would include only 20% of the target cardholders. Thus, the chart shows that the household specific variables can be quite useful for such targeting purposes.

7 Discussion and Future Research

One of the biggest challenge 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 is cited by supermarket managers as their biggest concern in the coming years. Using a unique frequent shopper database, we provide an empirical study of the impact of entry by a Wal-Mart supercenter on the sales of a traditional grocery store. We model the two key households decision of whether to visit the store and in-store expenditure using a flexible model of inter-purchase 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 their distance to the stores, demographics, and other shopping characteristics.

Our results show that the incumbent store lost significant volume–amounting to a quarter million dollars in monthly revenue–following Wal-Mart's entry. Decomposing the lost volume into store visits and in-store expenditures, we find that the majority of the losses came due to fewer store visits with little impact on the basket size once consumers are in the store. This is an important finding as it suggests that strategies designed to drive store traffic could be an effective way for recovering some of the lost volume. Interestingly, a small proportion of customers are found to account for a large proportion of the losses. For example, 10% of the households account for 45% of the store's lost revenue, while 20% of the customers account for almost 70% of the lost revenue. In terms of consumer characteristics, we find that distance to store while useful, explains little variation in household heterogeneity. Households that respond to Wal-Mart are likely to have an infant and pet in the family, and are more likely to be weekend shoppers. Similarly, store-brand buyers have a higher likelihood of moving purchases to Wal-Mart, while households that spend a large proportion of their grocery expenditures on fresh produce, seafood, and home meal replacement items are less likely to so. Finally, we find that the households that respond to Wal-Mart are large basket consumers, confirming the findings in Bell and Lattin (1998) that large basket buyers are more likely to choose an EDLP operator. Using a holdout sample, we find that these shopper characteristics can be quite useful in identifying potential 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. Even though a Wal-Mart supercenter carries all products typically found in a supermarket, variation in the quality of products (for example 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 has changed. 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), it is important to do a category-by-category analysis to analyze the differential impact across product groups. In doing so, one can draw upon the extensive literature on developing defensive marketing strategies (for example, the various strategies suggested in the DEFENDER type models, Hauser and Shugan 1983) to enhance category level retention.

There are also several shortcomings related to the data used in the analysis. First, we do not observe consumer purchases outside of 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 as they are found to be large basket (lower store visit frequency) and hence have fewer opportunities to take advantage of the promotions. Finally, another avenue for future research is based on our finding that the majority of the losses at the store were 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 non-trivial task. However, with better data and advancements in computing power, we hope that some of these issues can be addressed in the future.

References

[1] Allenby and Rossi [1999] "Marketing Models of Consumer Heterogeneity, *Journal of Econometrics*, 89,1-2, 57-78

[2] Bell,D. and J. Lattin[1998] "Shopping Behavor and Cosumer Preference for Store Price Format: Why Large Basket Shoppers Prefer EDLP" *Marketing Science*, (1) 66-88

[3] Bell, D., Teck-Hua Ho and Christopher S. Tang (1998), "Determining Where to Shop: Fixed and Variable Costs of Shopping," *Journal of Marketing Research*, 35, 352-69

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

[5] Blattberg, Robert C. and Edward J. Fox (1995), "Category Management," *Center of Retail Management, Northwestern University*

[6] Brown, Stephen [1989], "Retail Location Theory: The Legacy of Herald Hotelling," Journal of Retailing, 65, 450-70

[7] Chintagunta, P. [1993] "Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households" *Marketing Science,* 12,2,184-208

[8] Chintagunta, P and S. Haldar [1998] "Investigating Purchase Timing Behavior in Two Related Product Categories" *Journal of Marketing Research*, 35,1, 43-53

[9] Cox, D. R. [1972], "Regression models and life tables (with discussion)," Journal of the Royal Statistical Society, Ser. B, 34, 187–220.

[10] Corstjens, Marcel and Rajiv Lal [2000], "Building Store Loyalty through Store Brands", Journal of Marketing Research, 37, 281-291

[11] Craig, Samuel C., Avijit Ghosh, Sara McLafferty [1984] "Models of Retail Location Process: A Review," Journal of Retailing, 60, 5-36

[12] Fox, E, A. Montgomery and L. Lodish [2003] "Consumer Shopping and Spending Across Retail Formats: A Multivariate Tobit Model" *Journal of Business, forthcoming*

[13] Gupta, S. [1991]“Stochastic Models of Interpurchase Time with Time Dependent Covariates” *Journal of Marketing Research*, 28, 1-15

[14] Hansen, Karsten, Vishal Singh, and Pradeep Chintagunta [2003], “Understanding Store Brand Purchase Behavior Across Categories”, Working Paper, Carnegie Mellon University

[15] Hauser, J. and S. Shugan [1983] “Defensive Marketing Strategies” *Marketing Science*, 2 (4) 319-360

[16] Helsen, K., David Schmittlein [1993]. “Analyzing duration times in marketing: evidence for the effectiveness of hazard rate models”. Marketing Science. 12 (4) 395–414

[17] Hoch, Stephen [1996], “How Should National Brands Think About Private Labels?”, Sloan Management Review, 37, 2, 89-102

[18] Huff, David (1964), “Defining and Estimating a Trading Area,” Journal of Marketing, 28, 34-38

[19] Jain, Dipak and Naufel Vilcassim, “Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach,” Marketing Science, 1-23

[20] Lal, R. and C. Matutes [1989] “Price Competition in Multi-Market Duopolies” *Rand Journal of Economics,*30, 516-537

[21] Lal, R. and R. Rao [1997] “Supermarket Competition: The Case of Everyday Low-Pricing”*Marketing Science*, 1, 60-80

[22] Messinger, P. and C. Narasimhan [1997] “A Model of Retail Formats Based on Consumers Economizing on Shopping Time” *Marketing Science* 16,1-23

[23] Pesendorfer M. (2002); Retail Sales: A Study of Pricing Behaviour in Supermarkets.Journal of Business 75: 33-66.

[24] Phibbs CS and Luft HS [1995]. “Correlation of Travel Time on Roads Versus Straight Line Distance”. Medical Care Research and Review, 52:532-542

[25] *Progressive Grocer,* various issues.

[26] *Supermarket Business News*, various issues

[27] Reichheld, Frederick F. and Thomas Teal [1996], "The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value", Harvard Business School Press

[28] Rossi, P., R. McCulloch and G. Allenby [1996] "On the Value of Household Information in Target Marketing" *Marketing Science,* 15.321-340

[29] Seetharaman, P.B. and P. Chintagunta [2003] "The Proportional Hazars Model for Purchase Timing: A Comparison of Alternative Specifications" *Journal of Business and Economic Statistics*

[30] Singh, V. [2003], "Competing With Mass Discounters: Implications for Retailer Pricing," *Working Paper*, Carnegie Mellon University.

[31] Wedel,M., W.A. Kamakura, W.S. Desarbo, and F.T. Hofstede [1995] "Implications for Asymmetry, Nonproportionality and Heterogeneity in Brand Switching from Peicewise Exponential Mixture Hazard Models," *Journal of Marketing Research,* 32,4,457-462

MCMC algorithm

We use a Gibbs sampling algorithm to simulate the posterior distribution of the model parameters. In the following we describe the conditional distributions making up the Gibbs sampler for our model.

The store visit model is

$$U_{it} = X_{it}'\beta_i + \varepsilon_{it},, \qquad t = 1, \ldots, T_i;\ i = 1, \ldots, N, \tag{16}$$

where $\varepsilon_{it} \sim \mathrm{N}(0, 1)$. The expenditure model is

$$e_{it} = X_{e,it}'\lambda_i + \varepsilon_{e,it}, \tag{17}$$

where e_{it} is observed only when consumer i makes a store visit on day t, and $\varepsilon_{e,it} \sim \mathrm{N}(0, v_i^{-1})$.

The joint posterior distribution of latent utilities for consumer i, U_i, and $\theta_i = (\beta_i, \lambda_i)$ and v_i is then

$$\begin{aligned} p(U_i, \theta_i, v_i | e_i, D_i, \Pi, \Omega, b) \propto \prod_{t:D_{it}=0} \mathrm{TN}_{(-\infty,0)}(U_{it}|X_{it}'\beta_i, 1) \times \prod_{t:D_{it}=1} \mathrm{TN}_{(0,\infty)}(U_{it}|X_{it}'\beta_i, 1) \times \\ \prod_{t:D_{it}=1} \mathrm{N}(e_{it}|X_{e,it}'\lambda_i, v_i^{-1}) \times N(\theta_i|\Pi, \Omega) p(v_i|b), \end{aligned} \tag{18}$$

where $\mathrm{TN}_{(r_1,r_2)}$ is the truncated normal distribution with truncation region (r_1, r_2), and $p(v_i|a, b)$ is the gamma prior on v_i. The full joint posterior of $(U, \theta, v, \Pi, \Omega, b)$ is the product of (18) for all households multiplied by the prior for (Π, Ω, b).

This leads to the following conditionals for U_{it}:

$$U_{it}|\theta_i, \Pi, \Omega \sim \begin{cases} \mathrm{TN}_{(-\infty,0)}(U_{it}|X_{it}'\beta_i, 1) & \text{if } D_{it} = 0, \\ \mathrm{TN}_{(0,\infty)}(U_{it}|X_{it}'\beta_i, 1) & \text{if } D_{it} = 1. \end{cases} \tag{19}$$

The conditional for θ_i is found by first stacking the observations for U_i and e_i:

$$\mathbf{Y_i} \equiv \begin{bmatrix} U_i \\ e_i \end{bmatrix} = \begin{bmatrix} X_i & 0 \\ 0 & X_{e,i} \end{bmatrix} \theta_i + \begin{bmatrix} \varepsilon_i \\ \varepsilon_{e,i} \end{bmatrix} = \mathbf{X_i}\theta_i + \boldsymbol{\varepsilon}_\mathbf{i}. \tag{20}$$

In the following we let K_θ denote the dimension of θ, and K_Z denote the dimension of Z_i.

Conditional on U_i, this has the form of a normal linear regression model with covariance matrix

$$\mathrm{V}(\boldsymbol{\varepsilon}_\mathbf{i}) = \Sigma \equiv \begin{bmatrix} I_T & 0 \\ 0 & \nu_i^{-1} I_{np_i} \end{bmatrix}, \tag{21}$$

where np_i denotes the number of store visits by consumer i. By standard results the conditional for θ_i is then

$$\theta_i|\Pi, \Omega, U_i \sim \mathrm{N}\big(\theta_i|\hat{\theta}_i, \hat{\Omega}_i\big), \tag{22}$$

where

$$\hat{\Omega}_i^{-1} = \mathbf{X_i}'\Sigma^{-1}\mathbf{X_i} + \Omega^{-1}, \tag{23}$$

$$\hat{\theta}_i = \hat{\Omega}_i\big[\mathbf{X_i}'\Sigma^{-1}\mathbf{Y_i} + \Omega^{-1}\Pi Z_i\big]. \tag{24}$$

Letting $ssq_i \equiv \sum(e_{it} - X'_{eit}\lambda_i)^2$, the conditional for v_i is

$$\begin{aligned} \nu_i &\propto v_i^{np_i/2} \exp\big\{-\frac{v_i}{2}ssq_i\big\}v_i^{a-1}\exp\big\{-bv_i\big\}, \\ &\propto \mathrm{G}\big(v_i|np_i/2 + a, ssq_i/2 + b\big), \end{aligned} \tag{25}$$

a gamma distribution.

To describe the conditionals for Π, Ω we first define

$$\begin{aligned} M_{zz} &= \sum_{i=1}^{n} Z_i Z_i', \\ M_{z\theta_j} &= \sum_{i=1}^{n} Z_i\theta_{ij}, \quad j = 1, \dots, K_\theta, \end{aligned} \tag{26}$$

and

$$M_{zy} = \begin{bmatrix} M_{z\theta_1} \\ M_{z\theta_2} \\ \vdots \\ M_{z\theta_{K_\theta}} \end{bmatrix}.$$

Finally, let $\pi = \mathrm{vec}(\Pi')$. Using a prior of the form,

$$\begin{aligned} \pi &\sim \mathrm{N}(\mu_\pi, \Lambda_\pi^{-1}), \\ \Omega^{-1} &\sim \mathrm{W}(\nu, S), \end{aligned}$$

where $\mathrm{W}(\nu, S)$ denotes the Wishart distribution with ν degrees of freedom and scale matrix S, standard results leads to the conditionals

$$p(\Omega^{-1}|\pi, \theta) = \mathrm{W}\Big(\Omega^{-1}|\nu + n, \big(S^{-1} + \sum_{i=1}^{N} u_i u_i'\big)^{-1}\Big),$$

where $u_i = \theta_i - \Pi Z_i$, and

$$p(\pi|\Omega, \theta) = \mathrm{N}\Big(\pi|\Omega_\pi\big[(\Omega^{-1} \otimes I_{K_Z})M_{zy} + \Lambda_\pi\mu_\pi\big], \Omega_\pi\Big),$$

where $\Omega_\pi = \left[(\Omega^{-1} \otimes M_{zz}) + \Lambda_\pi\right]^{-1}$. We set the prior parameters as follows: $\mu_\pi = 0.0$, $\Lambda_\pi = 0.1 I_{K_\theta * K_Z}$, $\nu = K_\theta + 2$, $S = I_{K_\theta}/(K_\theta + 2)$.

The conditional for b is seen to be

$$
\begin{aligned}
p(b|v) &\propto p(b) \prod_{i=1}^{n} p(v_i|a, b), \\
&\propto p(b) b^{na} \exp\big\{ - b \sum_{i=1}^{n} v_i \big\}.
\end{aligned} \tag{27}
$$

We use a gamma prior for b, $b \sim \mathrm{G}(a_0 = 5, a_1 = 2)$. This implies a Gamma distribution for the conditional of b:

$$
b \sim \mathrm{G}\big(b|na + a_0, \sum_{i=1}^{n} v_i + a_1\big). \tag{28}
$$

	NO CARD	EMPLOYEE CARD	CARD MEMBER
SALES	12.4%	2.0%	85.6%
TRANSACTIONS	36.0%	3.0%	61.0%
BASKET SIZE	$8.0	$16.6	$33.5

Table 1: Shopper Card Penetration.

	PRE WAL-MART		POST WAL-MART		WHOLE PERIOD	
DEPARTMENT	MEAN	STD	MEAN	STD	MEAN	STD
PRODUCE	12.66	7.64	13.25	7.33	12.97	7.48
MEAT	164.19	123.19	202.25	146.87	184.16	137.34
HBC	125.81	46.25	155.28	51.08	141.27	50.98
GM	74.35	25.97	97.78	48.21	86.57	40.86
GROCERY	519.79	101.93	614.69	97.04	569.26	110.08

Table 2: Pre- and Post-Wal-Mart Promotion Index.

		AVERAGE DISTANCE IN MILES TO		
GROUP ID	FREQUENCY	FOCAL STORE	WAL-MART	INCOME ($)
1	392	6.10	3.70	33,909
2	242	6.88	7.69	41,896
3	1026	14.11	14.32	42,306
4	241	2.38	1.85	58,297
5	435	1.24	1.57	30,938
6	292	1.33	1.86	43,280
7	345	1.64	2.31	50,626
8	309	2.16	4.51	56,620
9	747	1.65	3.10	60,165
10	1221	0.64	2.72	40,935
11	249	2.61	4.16	46,907
12	315	5.00	4.63	64,052
13	674	5.10	7.48	63,316
14	626	5.98	7.98	56,878
15	884	0.39	2.35	22,945
16	820	0.71	3.16	28,546
17	963	1.41	3.91	35,719
18	240	2.27	0.82	60,480
19	3317	1.81	2.42	58,978

Table 3: Average characteristics for the 19 census tract regions

VARIABLE	MEAN	STD	MIN	MAX
RENT	0.71	0.24	0.03	0.99
MEDIAN INCOME	44352	16860	8713	105218
MEDIAN AGE	40.52	4.79	19.30	66.00
HH SIZE	2.36	0.31	1.21	3.94
BABY	10%			
PET	29%			
% OF TRIPS ON WEEKEND	0.31	0.19	0.00	1.00
% OF TRIPS BETWEEN 9-5	0.39	0.24	0.00	1.00
% Expditure on Produce	0.10	0.06	0.00	0.74
% Expditure on Meat	0.14	0.08	0.00	0.77
% Expditure on Seafood	0.03	0.05	0.00	0.56
% Expditure on HMR	0.02	0.05	0.00	0.92
% Expditure on Store Brand	0.14	0.07	0.00	0.59

Table 4: Demographic, Shopping, and Product Purchase Variables

ATTRIBUTE	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	S.D.	Z FRACTION
CONSTANT	-1.19	-1.28	-1.37	-1.26	-1.09	-1.11	-1.15	-1.23	-1.12	-0.96	-1.30	-1.32	-1.16	-1.29	-0.89	-1.02	-1.12	-1.25	-1.21	0.58	0.06
	(0.04)	(0.05)	(0.03)	(0.06)	(0.04)	(0.05)	(0.05)	(0.05)	(0.03)	(0.02)	(0.05)	(0.05)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.06)	(0.05)	(0.01)	(0.01)
$10^{-1} \times \tau$	-0.11	0.00	-0.01	-0.01	0.10	0.00	-0.10	0.11	0.01	0.01	0.04	-0.06	0.04	0.08	-0.05	0.01	0.08	-0.05	0.03	0.57	0.01
	(0.04)	(0.05)	(0.02)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.03)	(0.02)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.04)	(0.01)	(0.00)
$10^{-2} \times \tau^2$	0.02	0.00	0.00	0.00	-0.02	0.00	0.01	-0.02	-0.01	-0.01	-0.01	0.01	-0.01	-0.02	0.00	0.00	-0.02	0.01	-0.01	0.12	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
$1/\tau$	-0.74	-0.93	-0.98	-0.91	-0.67	-0.82	-0.93	-0.79	-0.69	-0.60	-0.94	-0.97	-0.78	-1.01	-0.53	-0.61	-0.63	-0.94	-0.88	0.56	0.08
	(0.05)	(0.06)	(0.03)	(0.06)	(0.04)	(0.05)	(0.05)	(0.05)	(0.03)	(0.02)	(0.06)	(0.06)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.06)	(0.05)	(0.01)	(0.01)
dw	-0.07	-0.05	-0.09	-0.09	-0.02	-0.04	-0.12	-0.09	-0.08	-0.06	-0.07	-0.09	-0.02	-0.07	-0.09	-0.05	-0.02	-0.15	-0.06	0.30	0.02
	(0.03)	(0.04)	(0.02)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.01)	(0.00)
$10^{-1} \times dw \times \tau$	-0.06	-0.08	-0.09	-0.04	-0.17	-0.09	0.01	-0.10	-0.09	-0.11	-0.09	-0.05	-0.10	-0.14	-0.10	-0.12	-0.13	-0.11	-0.10	0.33	0.02
	(0.03)	(0.04)	(0.02)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.02)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.04)	(0.04)	(0.01)	(0.00)
$10^{-2} \times dw \times \tau^2$	0.00	0.02	0.02	0.01	0.04	0.02	0.01	0.03	0.03	0.03	0.02	0.01	0.03	0.03	0.02	0.03	0.04	0.02	0.02	0.09	0.02
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
$dw \times (1/\tau)$	-0.07	-0.07	-0.08	-0.07	-0.10	-0.03	0.01	0.00	-0.04	-0.06	-0.05	-0.01	-0.10	-0.04	-0.05	-0.05	-0.08	-0.04	-0.07	0.34	0.02
	(0.05)	(0.07)	(0.04)	(0.07)	(0.04)	(0.06)	(0.05)	(0.05)	(0.03)	(0.02)	(0.07)	(0.06)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.07)	(0.06)	(0.01)	(0.01)
PROMPROD	0.01	0.02	0.02	0.02	0.01	0.01	0.02	0.00	0.02	0.01	0.03	0.03	0.02	0.02	0.01	0.00	0.01	-0.01	0.01	0.09	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
PROMMEAT	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	-0.01	0.08	0.01
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)
PROMHBC	0.01	0.00	0.02	0.04	0.02	0.03	0.02	0.04	0.01	0.02	0.03	-0.01	0.00	-0.01	0.01	0.02	0.01	0.02	0.01	0.14	0.01
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.00)	(0.00)
PROMGM	0.02	0.06	0.04	0.03	0.04	0.05	0.05	0.05	0.06	0.04	0.02	0.05	0.04	0.06	0.04	0.03	0.05	0.05	0.05	0.12	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.00)	(0.00)
PROMGROC	0.04	0.07	0.07	0.05	0.07	0.05	0.10	0.11	0.11	0.05	0.09	0.10	0.09	0.07	0.09	0.07	0.07	0.05	0.06	0.27	0.01
	(0.03)	(0.04)	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.01)	(0.00)

Table 5: Hierarchical coefficient estimates, duration model (posterior means with posterior standard deviation in parenthesis)

Attribute	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11	d12	d13	d14	d15	d16	d17	d18	d19	s.d.	Z fraction
Constant	3.20 (0.05)	3.30 (0.06)	3.45 (0.03)	3.21 (0.05)	2.96 (0.04)	3.07 (0.05)	3.06 (0.05)	3.19 (0.05)	3.14 (0.03)	2.89 (0.02)	3.34 (0.05)	3.37 (0.05)	3.24 (0.03)	3.43 (0.04)	2.86 (0.03)	2.95 (0.03)	3.02 (0.03)	3.27 (0.06)	3.30 (0.05)	0.61 (0.01)	0.10 (0.01)
dw	-0.07 (0.02)	-0.05 (0.02)	-0.07 (0.01)	-0.04 (0.02)	-0.03 (0.01)	-0.04 (0.02)	-0.04 (0.02)	-0.04 (0.02)	-0.06 (0.01)	-0.04 (0.01)	-0.06 (0.02)	-0.04 (0.02)	-0.04 (0.01)	-0.05 (0.01)	-0.03 (0.01)	-0.04 (0.01)	-0.04 (0.01)	-0.10 (0.02)	-0.05 (0.02)	0.18 (0.00)	0.01 (0.00)
τ	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)	0.02 (0.00)	0.02 (0.00)	0.01 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.01 (0.00)	0.01 (0.00)	0.02 (0.00)	0.01 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.01 (0.00)	0.01 (0.00)	0.04 (0.00)	0.01 (0.00)
promprod	0.02 (0.01)	0.03 (0.01)	0.02 (0.01)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	0.02 (0.01)	0.02 (0.01)	0.03 (0.01)	0.02 (0.01)	0.04 (0.01)	0.03 (0.01)	0.01 (0.01)	0.00 (0.01)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.03 (0.01)	0.04 (0.01)	0.10 (0.00)	0.02 (0.00)
prommeat	0.00 (0.01)	0.01 (0.01)	0.00 (0.01)	0.01 (0.01)	-0.01 (0.01)	0.00 (0.01)	-0.01 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.00)	-0.01 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.00 (0.01)	0.09 (0.00)	0.01 (0.00)
promhbc	0.02 (0.02)	0.02 (0.02)	0.00 (0.01)	0.01 (0.02)	0.01 (0.02)	0.03 (0.02)	0.01 (0.02)	-0.01 (0.02)	0.01 (0.01)	0.02 (0.01)	0.02 (0.02)	-0.01 (0.02)	-0.01 (0.01)	0.02 (0.01)	0.03 (0.01)	0.03 (0.01)	0.02 (0.01)	0.01 (0.02)	-0.03 (0.02)	0.14 (0.00)	0.02 (0.01)
promgm	0.00 (0.02)	0.05 (0.02)	0.04 (0.01)	0.02 (0.02)	0.01 (0.01)	0.04 (0.02)	0.05 (0.01)	0.06 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.02)	0.04 (0.02)	0.03 (0.01)	0.03 (0.01)	0.02 (0.01)	0.03 (0.01)	0.03 (0.01)	0.03 (0.02)	0.04 (0.02)	0.12 (0.00)	0.02 (0.01)
promgroc	-0.01 (0.04)	-0.01 (0.05)	0.04 (0.02)	0.01 (0.04)	0.06 (0.03)	-0.03 (0.04)	0.02 (0.04)	0.00 (0.04)	-0.01 (0.02)	0.00 (0.02)	-0.01 (0.04)	0.01 (0.04)	0.03 (0.03)	0.03 (0.03)	-0.03 (0.02)	-0.01 (0.02)	0.05 (0.02)	0.05 (0.05)	0.00 (0.04)	0.22 (0.01)	0.03 (0.01)

Table 6: Hierarchical coefficient estimates, Log basket size (posterior means with posterior standard deviation in parenthesis)

Region	Average percentage change in monthly expenditure	monthly number of trips	basket size
1	-0.24	-0.20	-0.06
	(0.29)	(0.28)	(0.11)
2	-0.18	-0.16	-0.03
	(0.26)	(0.23)	(0.11)
3	-0.27	-0.24	-0.05
	(0.26)	(0.25)	(0.11)
4	-0.18	-0.16	-0.02
	(0.26)	(0.24)	(0.11)
5	-0.17	-0.16	-0.02
	(0.22)	(0.22)	(0.11)
6	-0.15	-0.12	-0.03
	(0.24)	(0.23)	(0.11)
7	-0.16	-0.14	-0.03
	(0.24)	(0.22)	(0.12)
8	-0.18	-0.17	-0.02
	(0.24)	(0.22)	(0.12)
9	-0.19	-0.16	-0.04
	(0.22)	(0.21)	(0.11)
10	-0.16	-0.14	-0.02
	(0.26)	(0.25)	(0.13)
11	-0.21	-0.18	-0.04
	(0.26)	(0.24)	(0.11)
12	-0.20	-0.18	-0.02
	(0.27)	(0.26)	(0.11)
13	-0.14	-0.12	-0.03
	(0.24)	(0.22)	(0.11)
14	-0.23	-0.21	-0.(04
	(0.25)	(0.23)	(0.11)
15	-0.18	-0.17	-0.01
	(0.26)	(0.23)	(0.14)
16	-0.16	-0.15	-0.02
	(0.26)	(0.25)	(0.13)
17	-0.15	-0.13	-0.03
	(0.25)	(0.23)	(0.12)
18	-0.37	-0.32	-0.08
	(0.25)	(0.24)	(0.10)
19	-0.21	-0.18	-0.03
	(0.26)	(0.24)	(0.12)
total	-0.19	-0.17	-0.03
	(0.26)	(0.24)	(0.12)

Table 7: Average Wal-Mart entry effects by region (with standard deviation in parenthesis)

VARIABLE	ESTIMATE	SE
INTERCEPT	-1.5569	0.5222
% RENT	0.2447	0.1793
MEDINC	0.0004	0.0008
MEDAGE	0.0015	0.0053
AVGHHSZ	0.0491	0.1277
BABY	0.4589	0.0452
PET	0.3787	0.0323
% WEEKEND	0.3064	0.1058
% 9-5 TRIPS	-0.3134	0.0888
% PRODUCE	-0.6112	0.2818
% SEAFOOD	-2.3161	0.4313
% HMR	-2.2864	0.3978
% MEAT	-0.1657	0.2002
% PRIVATE	1.1086	0.3234
BASKET-SIZE	0.013	0.0006

Table 8: Probit Regression: Profiling the Defectors to Wal-Mart.

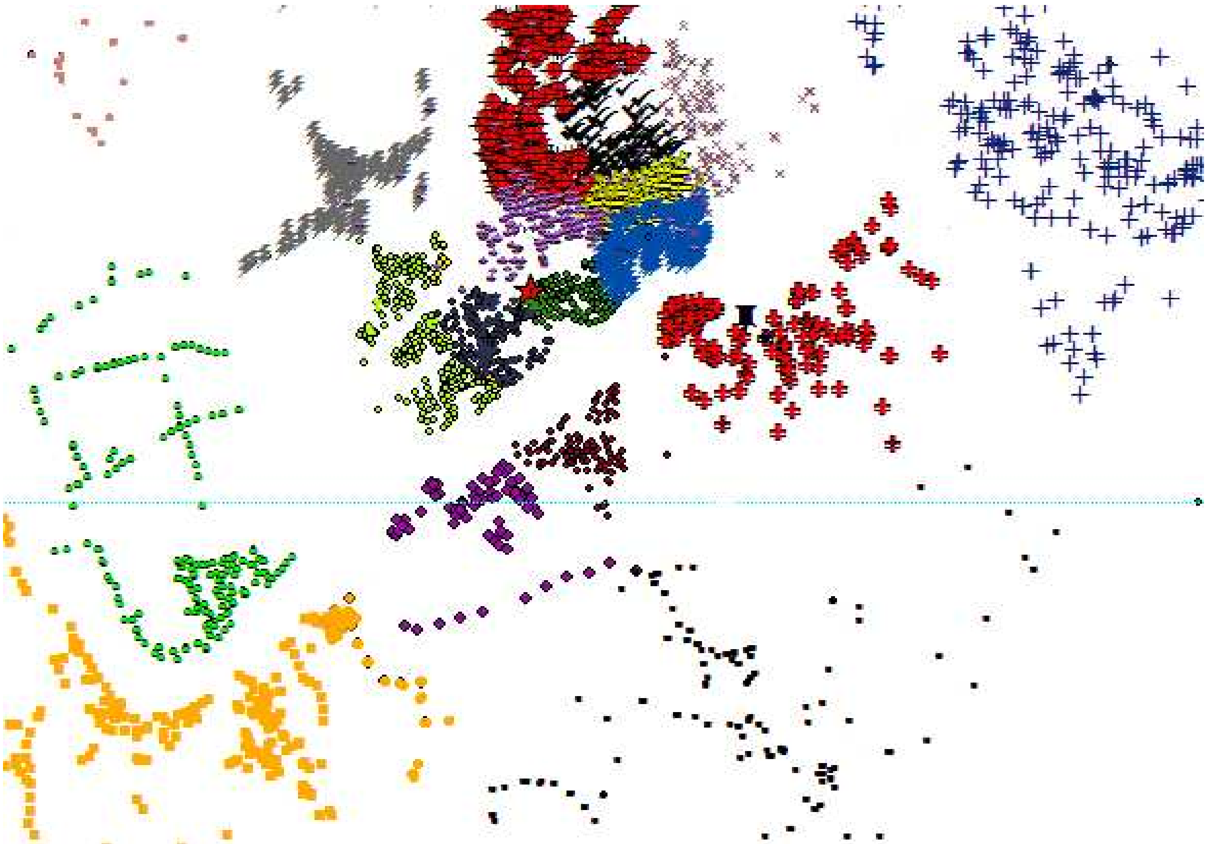


Figure 1: Household locations.

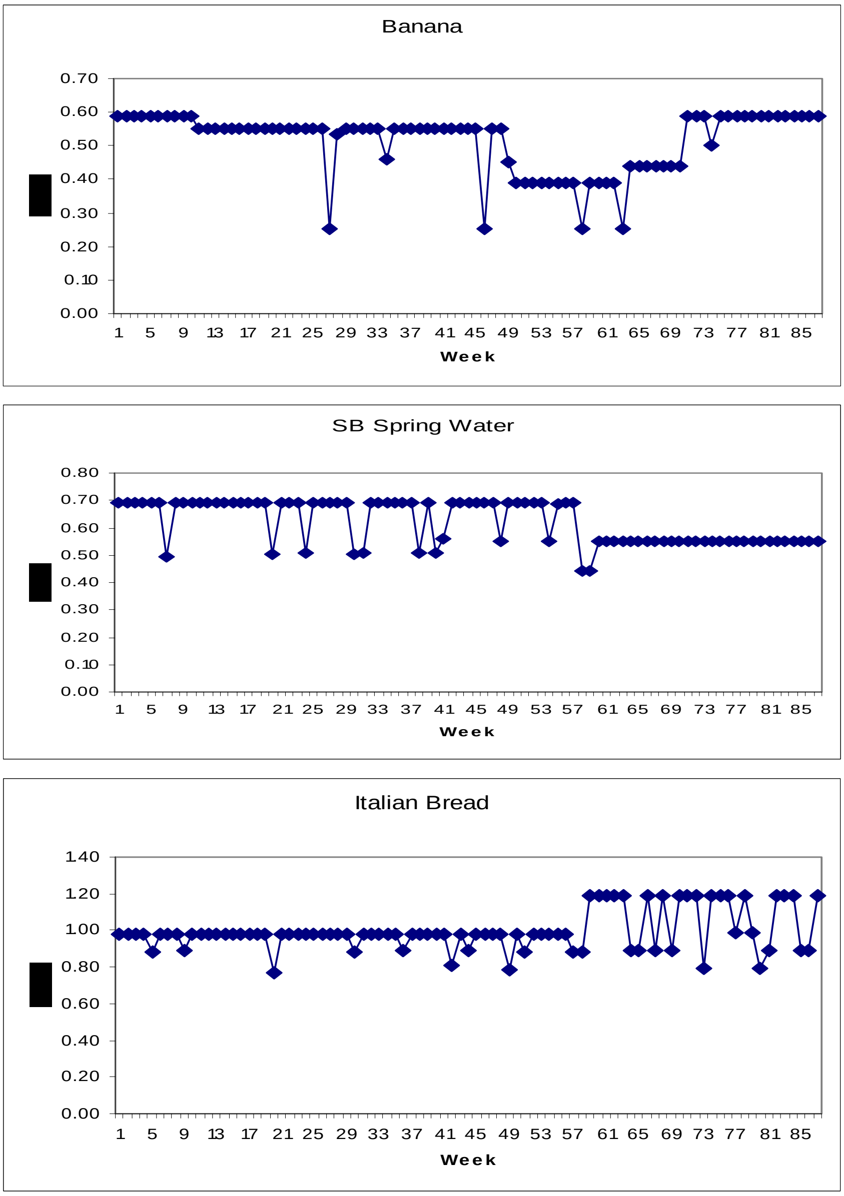


Figure 2: Pricing series for three products.

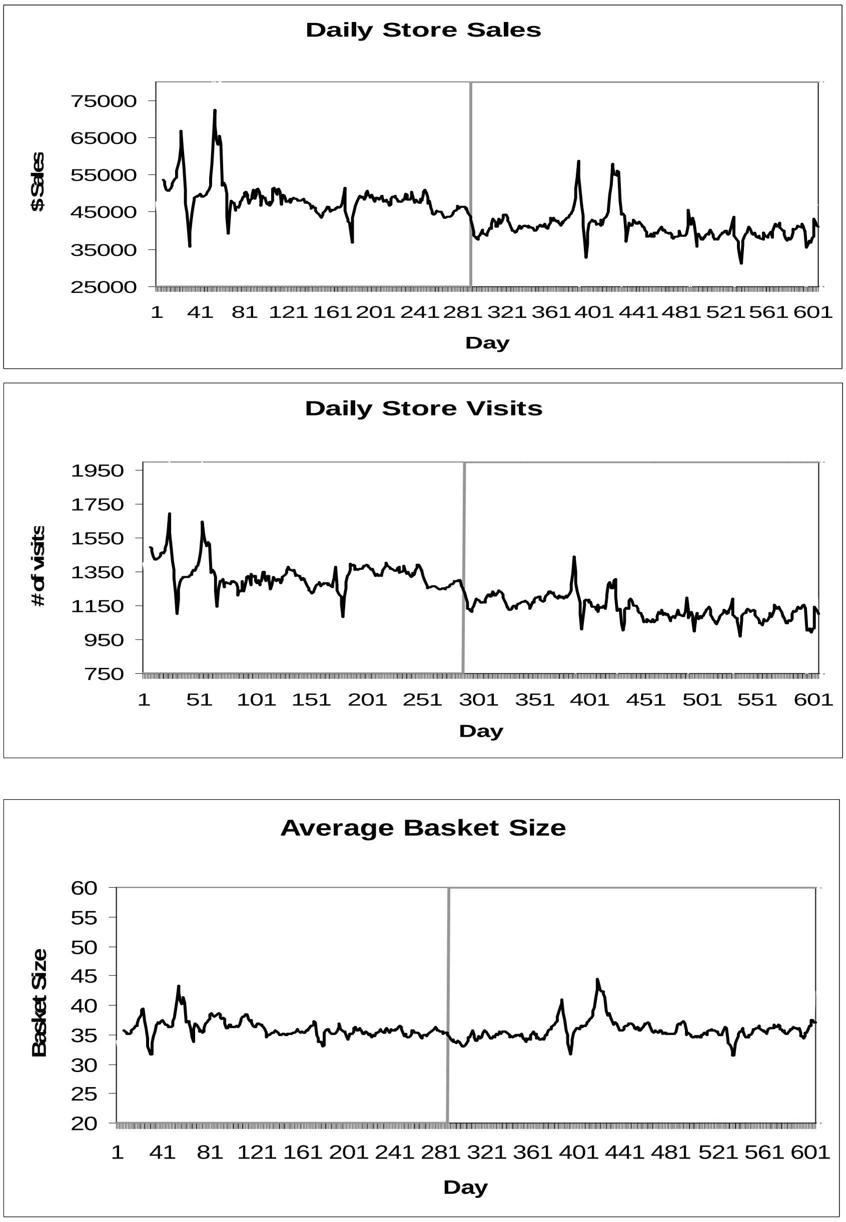


Figure 3: Aggregate sales, visits and basket size for the focal store.

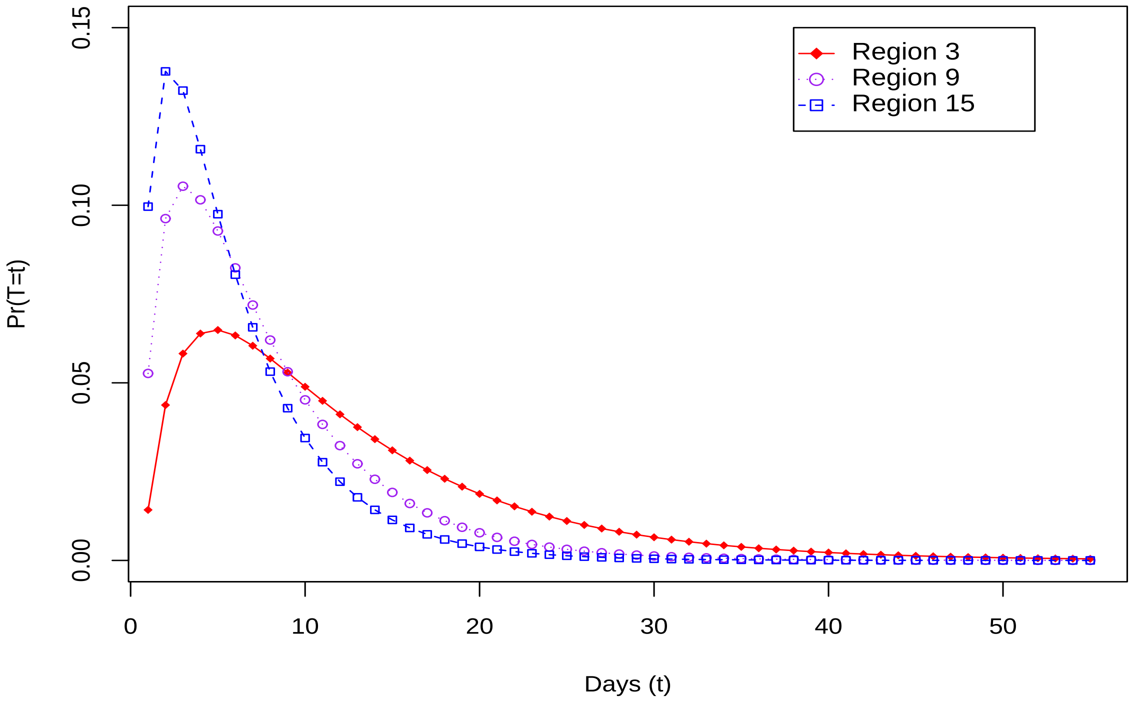


Figure 4: Distributions of inter-store-visit times for three regions prior to Wal-Mart entry.

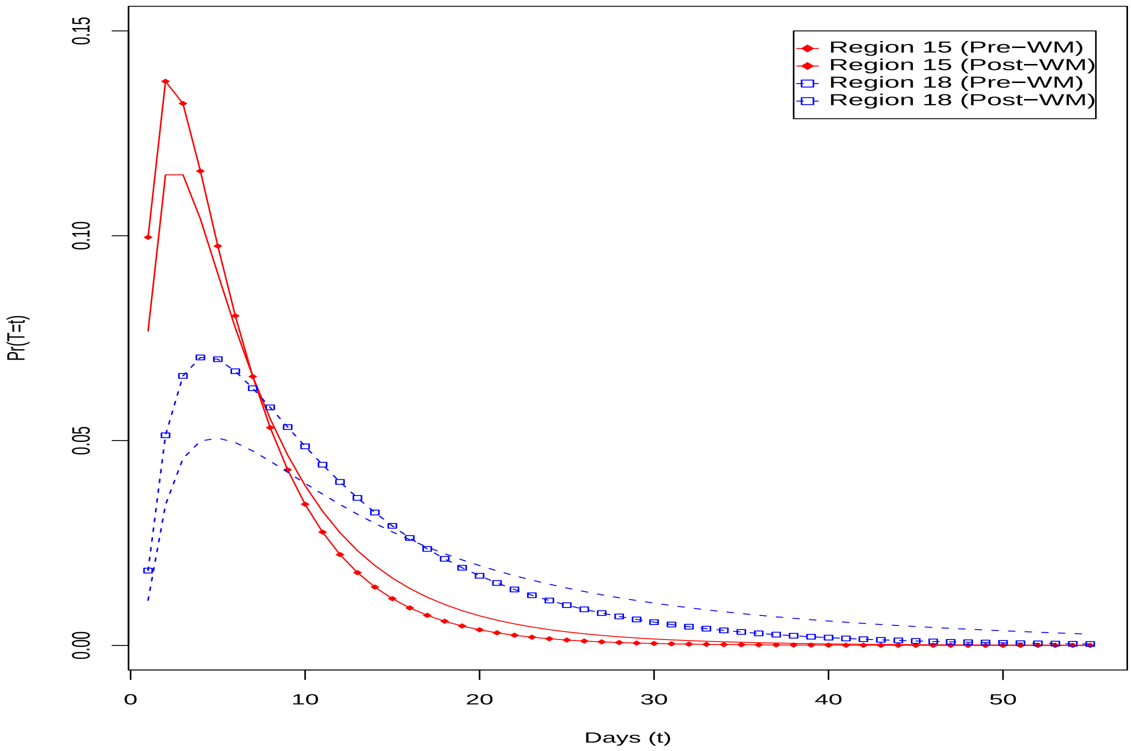
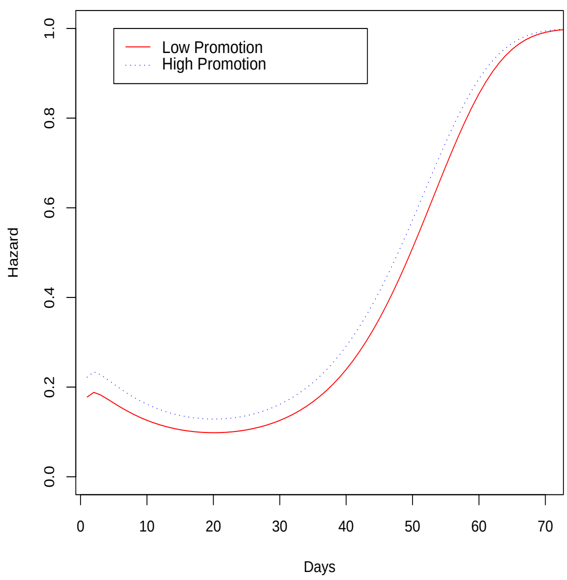
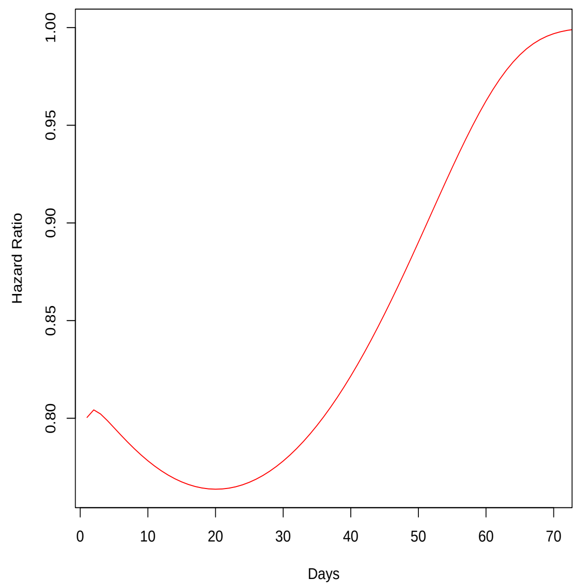


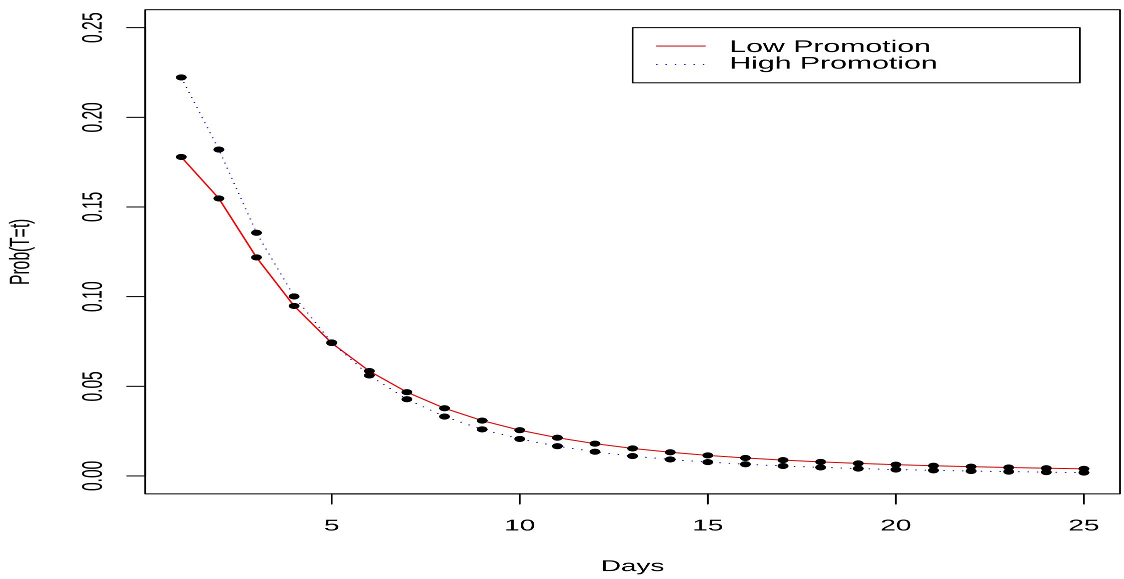
Figure 5: Distributions of inter-store-visit times for two regions, pre and post Wal-Mart entry.



(a) Hazard functions, Household #3572



(b) Hazard ratio, Household #3572



(c) Distributions of Inter-store-visit times, Household #3572

Figure 6: Promotion Effects for Household #3572.

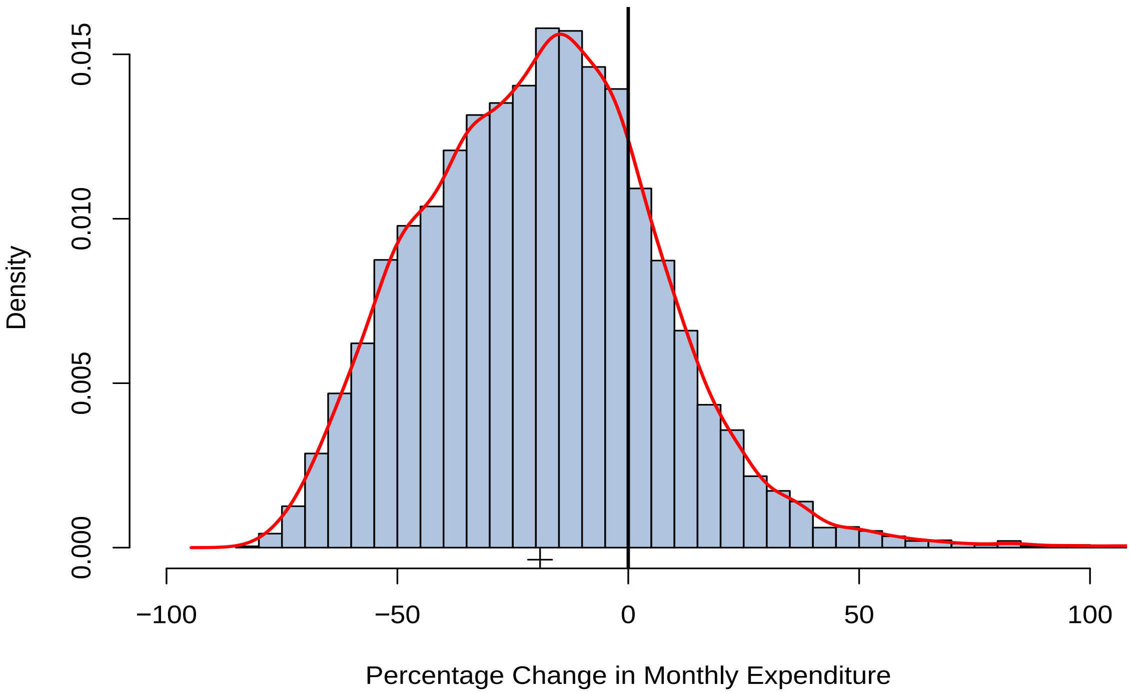


Figure 7: Distribution of Wal-Mart entry effect (relative change), 9852 households.

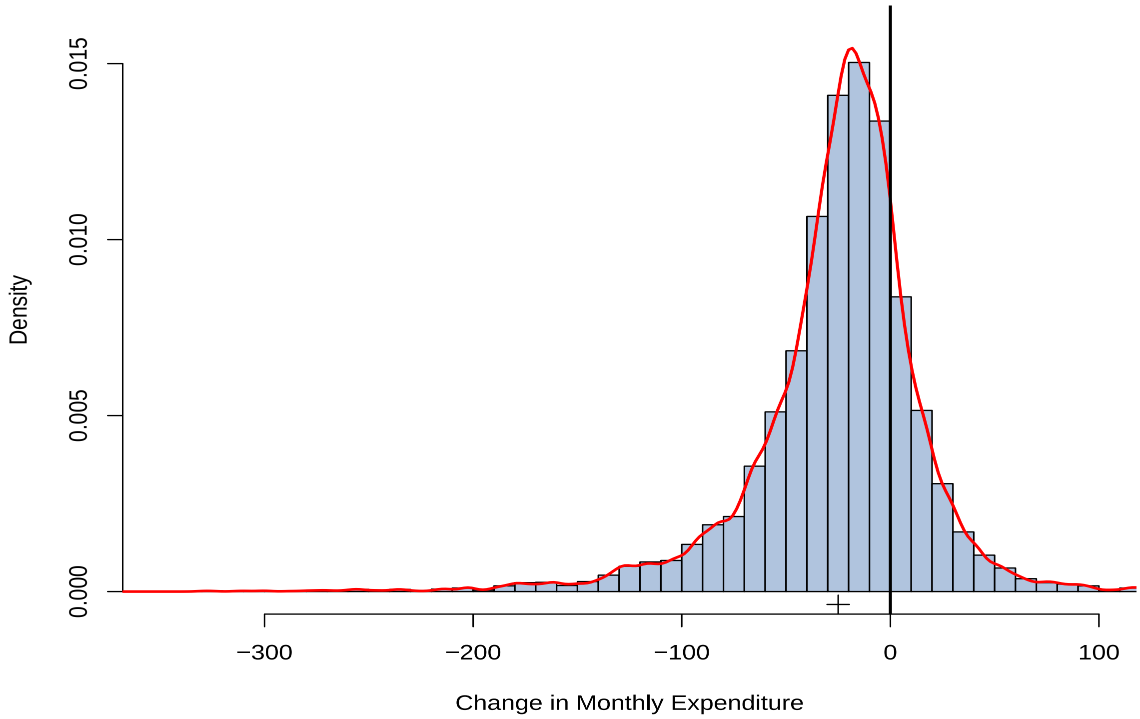


Figure 8: Distribution of Wal-Mart entry effect (absolute change), 9852 households.

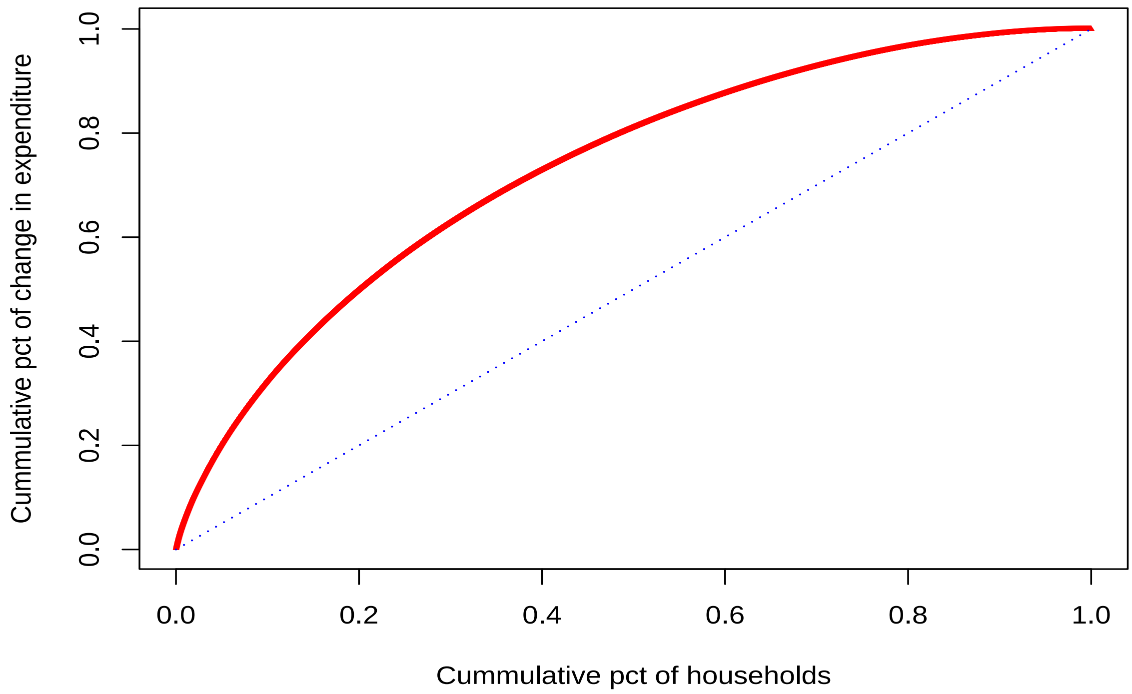
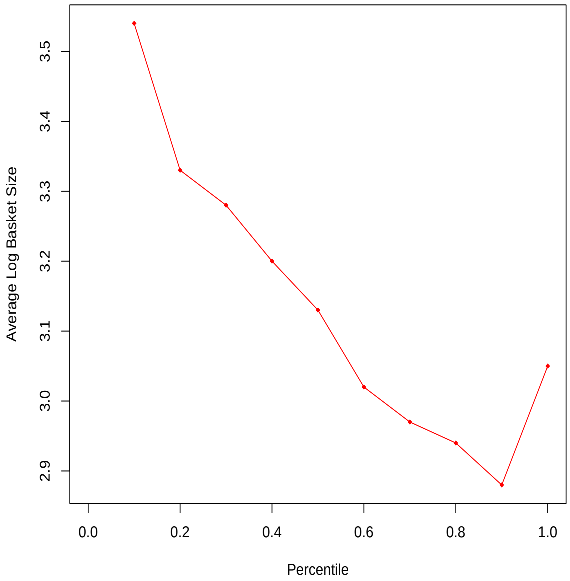
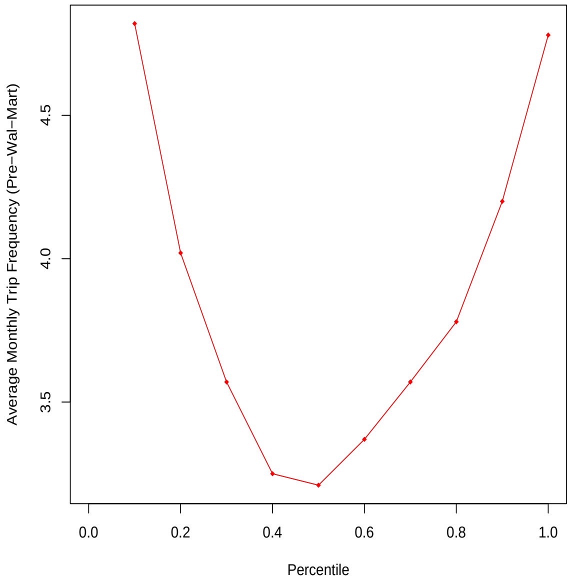


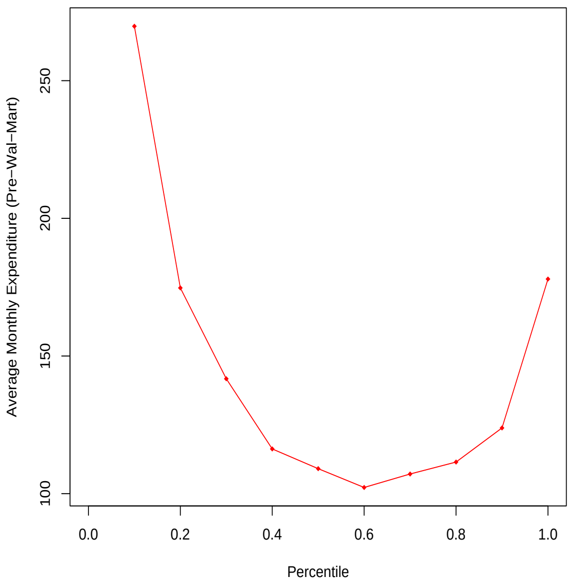
Figure 9: Lorentz curve for Wal-Mart effects distribution.



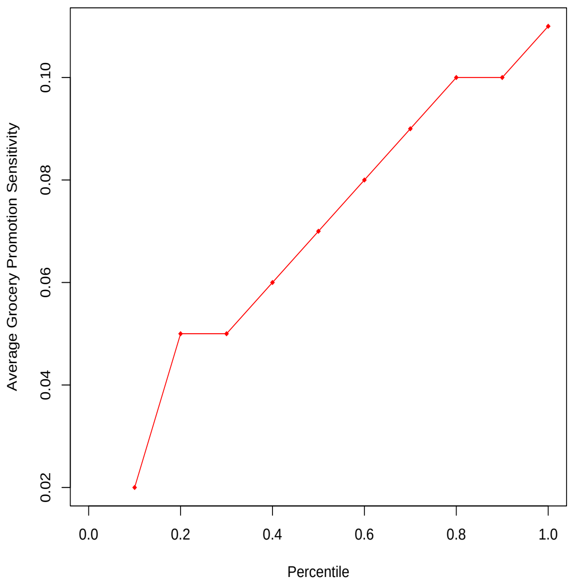
(a) Basket Size



(b) Pre-Wal-Mart Trip Frequency



(c) Pre-Wal-Mart Expenditure



(d) Grocery Promotion Sensitivity

Figure 10: Decile plots of the effects distribution versus shopping behavior metrics.

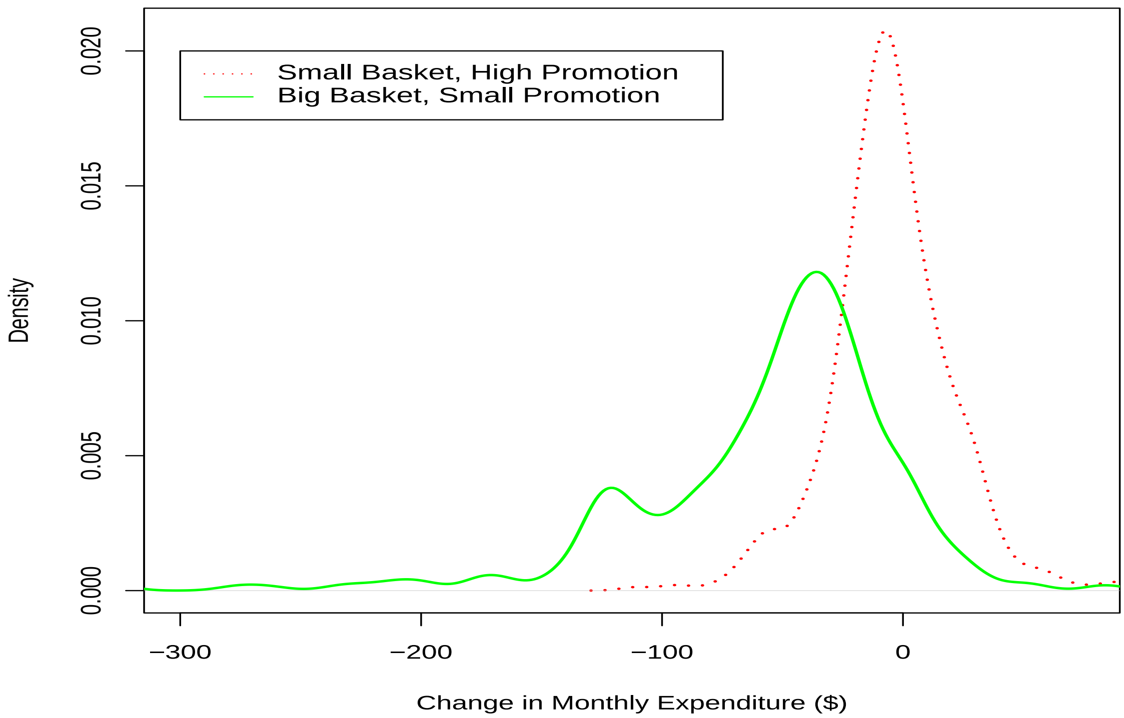


Figure 11: Effects Distribution for Two Segments.

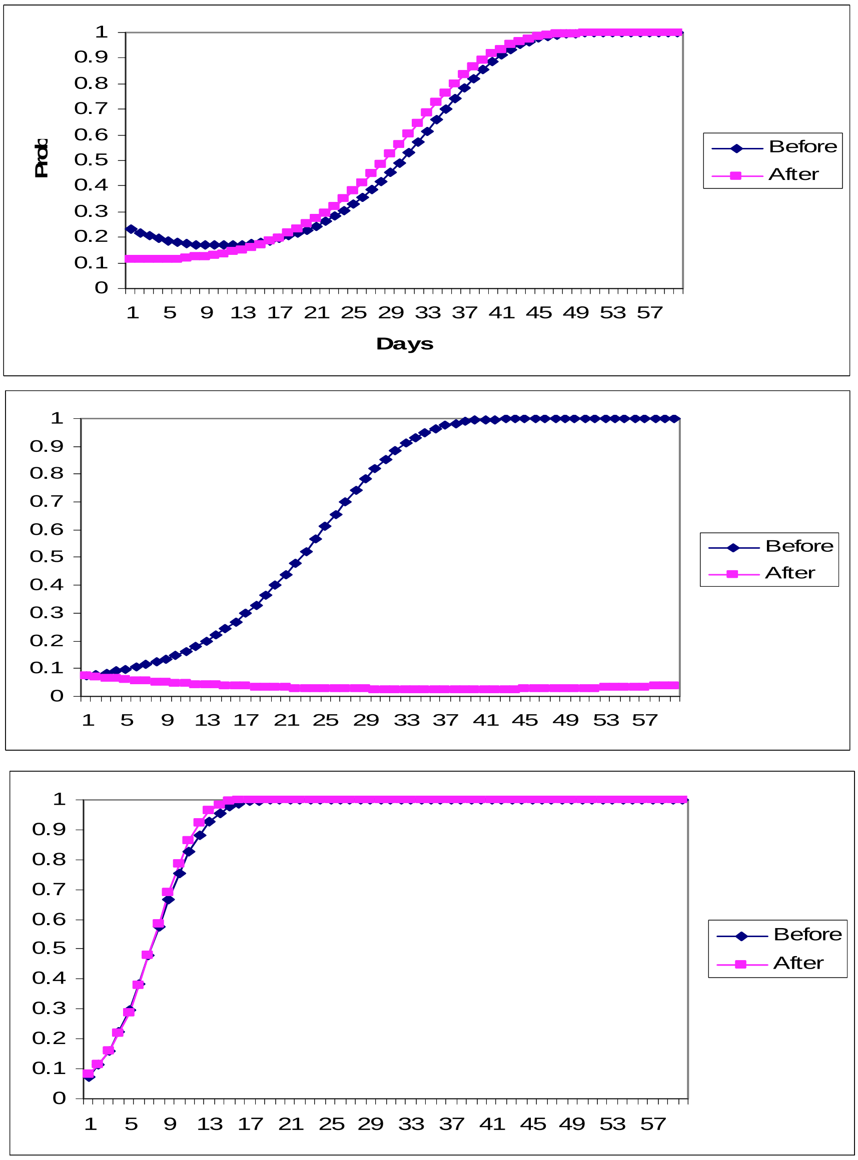


Figure 12: Hazards for three households: Household A: Distance to Store .09 miles, Basket Size: \$32, pct. of (9-5) trips:17, Household B: Distance to WM .21 miles, Basket size: \$51, % of (9–5) trips: 4%, Household C: Distance to WM .21 miles, Basket size: \$9, % of (9–5) trips: 78%.

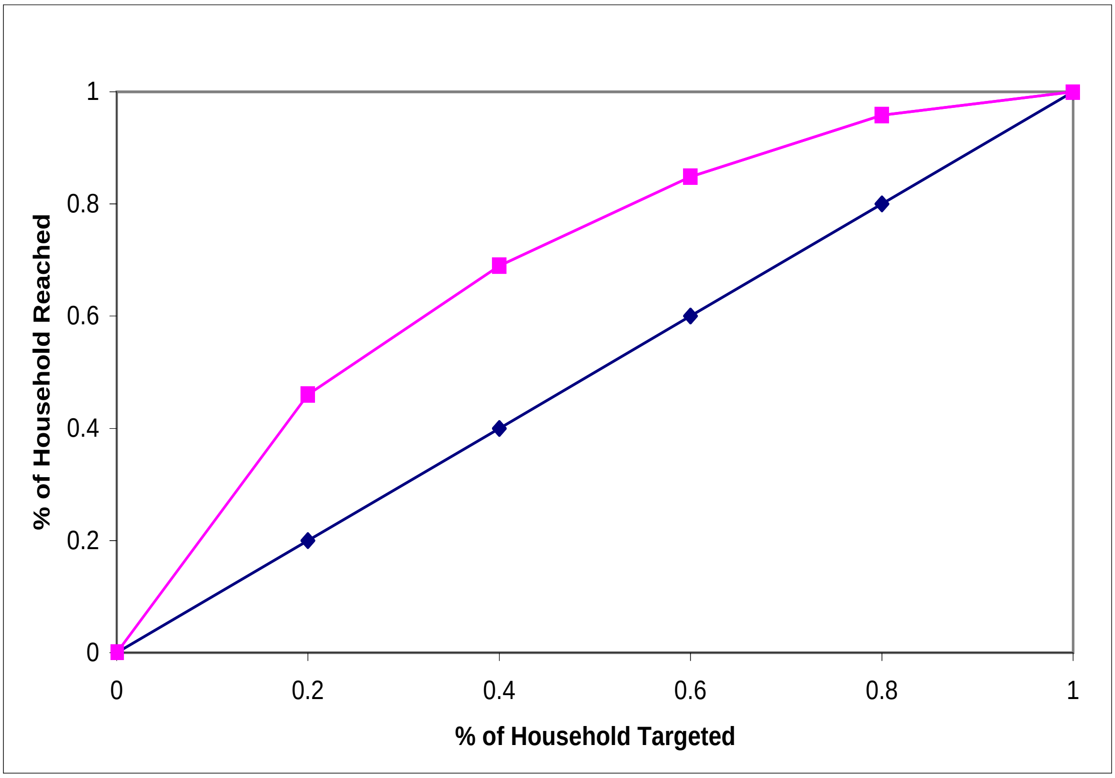


Figure 13: Gains Chart for hold-out sample.