Journal of Retailing and Consumer Services 48 (2019) 16–27

Contents lists available at ScienceDirect

Journal of Retailing and Consumer Services

journal homepage: www.elsevier.com/locate/jretconser

# A "hidden" side of consumer grocery shopping choice

Aidin Namin[a,*], Yashar Dehdashti[b]

[a] College of Business Administration, Loyola Marymount University, One LMU Drive, MS 8385, Los Angeles, CA 90045, USA
[b] School of Business Administration, Texas Wesleyan University, AMB 323, 1201 Wesleyan St., Fort Worth, TX 76105, USA

ARTICLE INFO

*Keywords:*
Consumer grocery shopping choice
Consumer packaged goods
Finite Mixture Models
Latent class analysis
Consumer demographics
Retailing

ABSTRACT

This study identifies hidden classes of grocery shoppers and their choice of different items on different days of the week. Following the literature on consumer grocery shopping, three major groups of products are considered: food/drink, cleaning, and personal care. Applying Finite Mixture Modeling to a rich scanner dataset, latent classes of customers and their choice of grocery items on different days of the week are discovered and empirically validated. The model controls for consumer unobserved heterogeneity and demographic characteristics through mixing probabilities. Results uncover latent classes of grocery shoppers and their day of the week shopping day, their sizes, their product choices, mixing probabilities, and demographics. Findings offer retail promotion targeting guidelines for the identified latent classes in the food/drink, cleaning, and personal care groups. Analysis outcome provides marketing and managerial implications in identifying grocery store segments, handling store traffic, managing store promotion and pricing, and improving store layout.

## 1. Introduction

Imagine a grocery store which customizes your shopping experience by offering personal deals to you the second you walk in the store! The idea might seem far-fetched at first since the store manager would need to know *which* customer is interested in deals on *what* items and *when* she is going to do her grocery shopping and purchase those items. Upon deeper contemplation, nonetheless, the store manager might be able to come up with an answer to this question provided that he has access to the right data and is able to employ the appropriate tools. The present research focuses on answering this question by unveiling the characteristics of grocery shoppers who buy specific products on particular days of the week in an effort to help the grocery store managers micro-target their promotional activities towards these different groups of customers at the right time and for the right product categories. The results of this research have important implications for academicians and practitioners who are interested in particularities about grocery shoppers segments' characteristics, basket composition, and day of purchase, and how this information could lead to better micro-marketing strategies.

Although grocery shopping behavior is an important aspect of studying consumer choice behavior, few studies have focused on investigating the behavior of grocery shoppers search and choice using multiple determinants for individual customers (e.g. characteristics of an individual customer, her basket composition, grocery shopping day, etc.) (Bawa and Ghosh, 1999; Koch, 2013; Park et al., 1989; Putrevu and Lord, 2001). The importance of grocery products in consumers' retail shopping basket can hardly be exaggerated. In the U.S., grocery shopping continues to be a prominent aspect of retail shopping. For instance, in 2014, virtually every household in the U.S. bought consumer packaged goods in grocery stores (Statista, 2014). In Dec 2017, U.S. grocery stores sold a staggering $582,891 million just in food and beverages while at the same time health and personal caring stores sold about $300,353 million (U.S. Census Bureau, 2017).

Various angles of grocery shopping behavior have been studied, mostly by focusing on one determinant of search or choice at a time. These studies have focused on areas such as visit frequency (e.g., Kahn and Schmittlein, 1989), expenditure per trip (e.g., Blaylock, 1989), price awareness (McGoldrick and Marks, 1987), household characteristics (e.g., MacKay, 1973), grocery outlet selection (Yavas and Tuncalp, 1984), basket characteristics (e.g., Vartanian et al., 2007), cognitive vs. chronological age of customers and its effect on perception (Teller et al., 2013), in-store advertising (e.g., Roggeveen et al., 2016), nature and sequence of visit to different parts of the store (Brown, 1988), etc. Although these studies have produced impressive results, most of them have concentrated on one set of homogeneous factors (e.g., demographics) to study grocery shoppers' choice patterns. Therefore, the literature still lacks a comprehensive study of grocery store customers incorporating a heterogeneous host of variables (e.g., demographics, basket composition, purchase timing, etc.) to shed light on more subtle

* Corresponding author.
*E-mail addresses:* aidin.namin@lmu.edu (A. Namin), ydehdashti@txwes.edu (Y. Dehdashti).

https://doi.org/10.1016/j.jretconser.2019.01.009
Received 29 June 2018; Received in revised form 10 December 2018; Accepted 10 January 2019
0969-6989/ 

patterns of choice. For instance, while the grocery shopping literature is replete with studies about price awareness among grocery shoppers, there is no information on the characteristics of these customers, the grocery items to which they are price-sensitive, when they plan to visit to store to purchase those items, etc. For instance, while we know what days of the week are more popular with the consumers in general (Fox and Hoch, 2005; Kahn and Schmittlein, 1989), we do not know which segments of these consumers (based on their demographics) are more likely to buy on those days. This means that we do not have a clear-cut answer to the question '***who*** buys ***what*** and ***when*?**' in the major groups of grocery products. More specifically, consider the case of a workaday, mildly overweight grocery shopper, let's say *John* for the sake of this example, who treats himself to a nice cheat meal every Friday evening, after a week of hard work and harsh dieting, by devouring a meat-lovers pizza. It goes without saying that John would be interested in a nice, chilled Coke to complement this feast. John is a bargain hunter and will naturally seek the best price for this purchase. Now, if the grocery store manager, *Leeann*, is aware of the preferences of customers like John for bargains, knows what pairs of products they might be interested in, and is aware of their purchase timing, she might be able to offer such customers the deal for the right products and at the right time. Similar to this example, a detailed profiling of grocery shoppers based on their characteristics and the day of purchase could improve decisions about promotion timing, pricing, in-store advertising, and store layout.

In other words, while it is established that factors such as buyers' demographics and their day of grocery shopping could influence purchase behavior, there is a gap in the literature as to how these factors could be used to predict the behavior of different segments of heterogeneous buyers for different groups of products on various days of a given week. In order to investigate this issue, and with the purpose of unveiling a hidden aspect of consumer grocery shopping choice, the present study delves deeper into the issues of customer grocery choice, day of the week on which purchase occurs, and customer demographics.

We, therefore, contribute to the literature by shedding light on consumers' grocery shopping choice patterns for different groups of products and days of the week chosen to purchase them. Our results have implications for (1) academicians who are interested in studying grocery shoppers' segments in gory detail and (2) practitioners who seek to develop micro-targeting and promotional strategies based on the characteristics of their customers' segments that predict the nature and timing of choice. Particularly, our study provides guidelines as to when and how the promotional activities should be targeted/implemented towards certain demographic groups of consumers in order to increase likelihood of purchase for certain items, which then, translates to higher levels of sales for the store. We, also, empirically validate the preference of certain segments of consumer demographics to buy more and/or more frequently in the presence of targeted promotions on different days of the week. Specifically, grocery store managers could target the promotions to segments that are sensitive to the presence of such offerings on the days those consumers are more likely to make their grocery shopping choice of products.

To accomplish this, we employ a state-of-the-art model and estimation technique controlling for customer unobserved heterogeneity and its impact on customer grocery shopping choice. Our model allows for a probabilistic setting in which mixing probabilities are guided by buyer demographics, while our estimation procedure accounts for a random effect coefficients setting. In other words, we uncover latent classes (i.e., segments) of heterogeneous grocery shoppers who buy three major groups of grocery products, size of each segment, and mixing probabilities for these classes based on customers' choice of the day of the week to do grocery shopping. The estimated mixing probabilities enable us to clearly identify demographic characteristics of each segment of consumers for each product group. Finally, our latent class analysis results have important marketing managerial implications for better identifying grocery store segments and better handling of store traffic management. They can, also, be used in marketing management, particularly promotion and pricing management, as well as in improving the layout of grocery stores.

We develop our model using two major streams of grocery shopping literature. One stream deals with the issue of choosing the day of the week to do grocery shopping. The other stream deals with the issue of the major groups of grocery shopping products consumers buy in their grocery shopping trips. Specifically, considering that consumers generally buy products in three major groups, i.e., food/drink products, cleaning products, and personal care products (Farhangmehr et al., 2000), we use Finite Mixture Modeling (FMM) to investigate the latent (i.e., hidden) classes of buyers in each group based on their demographics and the day of the week when shopping occurs. We empirically validate the model using scanner panel data of two major metropolises in the U.S. Our model, also, accounts for unobserved heterogeneity on the demand side in a random effects coefficient probabilistic setting. In summary, accounting for a large set of consumer demographics, our main results show that there are two different-in-size latent classes in the food/drink group. The larger segment (81% in size) is more likely to spend money on food/drink on Friday, Saturday, and Sunday while the smaller segment (19% in size) is more likely to spend money on food/drink on Friday. Additionally, we find empirical evidence for two relatively similar-in-size classes in the cleaning products group. The first class (54% in size) is more likely to spend money on cleaning products on Saturday while the second class (46% in size) is more likely to spend money on cleaning products on Sunday. Finally, we show that there are three classes in the personal care products group. The first class (67% in size) is more likely to buy personal care products on Saturday and Sunday. The second class (21% in size) is more likely to spend money on personal care products on Monday. The third class (12% in size) does not show any significant evidence of having preference for any day of the week. Additionally, for all three groups of products, a full set of demographic variables guide the mixing probabilities for each class and determine their sizes (i.e., probabilities) and specifications. These demographic variables are then used to describe the properties of each identified segment. Further, it is well-established that offering promotions in the grocery shopping context could lead to an increase in sales (see, for instance, Gönül and Srinivasan, 1993). This increase in sales is brought about as a result of some customers buying more and/or more frequently due to the presence of promotions. Therefore, changes in the offering of promotions might lead to an increase in the purchase of particular grocery items. Given that one of the results of our study is the preference of each of the segments for grocery shopping timing, grocery store managers can increase the store's sales by targeting promotional activities to each segment of consumers at the right time, i.e., on the day when they make their grocery shopping choice. Therefore, we show that in the food and carbonated beverage category, when grocery store managers offer promotions on Fridays, Saturdays, and Sundays, Class 1 responds better to these promotions. In other words, based on our results, it is a better strategy for the grocery store manager to target promotions to this segment on Fridays, Saturdays, and Sundays. In the cleaning products category, however, customers in Class 2 respond better to promotional offerings on Sundays, and therefore, it is a better strategy for the grocery store manager to target promotions to this segment on Sundays. Finally, in the personal care products category, Class 1 is more likely to purchase on Saturdays and Sundays, and therefore, it is better to target promotions to this segment on those days.

To the best of our knowledge and based on extensive review of the literature, this is the first study in this domain that examines latent classes of grocery shoppers and their day of the week choice to such an extent. In other words, as elaborated upon in the Section 2 of this paper, while there is no scarcity of articles that examine grocery shopping behavior using one or sometimes two variables at a time, we could not find any studies that take into consideration the three determinants of grocery shopping choice at the same time: (1) who buys (characteristics of the segment of shoppers) (2) what (characteristics of the basket of shoppers) and (3) when (characteristics of shoppers' trip timing).

We are, to a large extent, limited in performing our analyses to the observed variables in the dataset and to the information it contains. While we are aware of the limitations imposed on us due to the nature of our dataset, to the best of our knowledge, our study is unique in using and combining these three levels of variables to uncover latent segments in the grocery store industry. The rest of this paper is organized as follows: In the following sections, we present a review of the relevant literature, followed by the theoretical background and modeling procedure, which is then followed by the description of our data and estimation results. We finish the paper by offering conclusions, managerial implications, and future research.

## 2. Literature review

### 2.1. Grocery shopping behavior and the choice of the day of the week

Consumer grocery shopping is an important aspect of consumer behavior (Bawa and Ghosh, 1999; Park et al., 1989). Certain characteristics differentiate between a regular shopping context and grocery shopping. For instance, multiple objectives (Park et al., 1989) need to be achieved during a given grocery shopping episode. Further, the behavior has to be repeated after a period of time. (Park et al., 1989). Despite its importance, consumer grocery shopping behavior seems to not possess an excellent track record of attracting research studies (Bawa and Ghosh, 1999; Koch, 2013).

One of the areas of research on grocery shopping behavior is frequency, i.e., the choice of (1) when and (2) how many times in a certain period (such as a week) grocery shopping is to be done. This is largely motivated by the fact that store traffic has been identified by retail store managers as the most important success factor (Kim and Park, 1997). In fact, in a grocery store context, it has been found that only thirty percent of shoppers visit stores at fixed intervals (Kim and Park, 1997). This finding is in line with research done by Kahn and Schmittlein (1989) which categorized customers into the 'Quick Trips' and 'Regular Trips' segments. This categorization is based on the amount of money spent; i.e., quick trips are those where the customer spends a small amount of money, and regular trips are those where the customer spends a large amount of money. They, also, found evidence for strong preference for shopping on a specific day of the week. In other words, the authors found that shoppers plan their grocery shopping time based on when in the week they prefer to do it and not when in the week the products they are going to buy will be used up. Further, a study of grocery shopping behavior in England and Wales looked at the proportion of buyers who might have developed 'shopping habits' and found that sixty one percent of shoppers go to grocery stores on a particular day of the week and sixty seven percent of them do so at a particular time of the day (East et al., 1994). Similarly, Fox and Hoch (2005) found Sunday to be the most popular grocery shopping day of the week, with Friday and Saturday being the second and the third most popular days. In line with the frequency issue, Nordfalt (2009) examined the 'unplanned' grocery shopping behavior. He found that large trips (i.e., regular trips, in the language of Kahn and Schmittlein, 1989) are "well-defined" planned trips whereas small (i.e., quick) trips are "contingency-based".

Some of the studies in the realm of grocery shopping frequency investigated the relationship between frequency and demographic variables. MacKay (1973) defined low demand families as those where the number of children is less than two, the husband's job is neither professional nor technical, and the annual family income is lower. High demand families are those where the number of children is equal to or more than two, the husband's job is either professional or technical, and the annual family income is higher. Low disposable time families are defined as those where the youngest child is younger than six years of age, and the wife works. High disposable time families, on the other hand, are those where the youngest child is six years old or older and the wife does not work. Low-demand families who have low disposable time were found to have regular shopping behavior. High-demand families with high disposable time, on the other hand, were found to make more quick trips. Doti and Sharir (1981), also, developed and empirically tested a model of grocery shopping behavior using a variety of demographic variables. They found positive evidence for the effect of these demographic variables on grocery shopping behavior. Using a similar approach, Blaylock (1989) found that race, household size, and age are the most important factors that affect grocery shopping behavior. Along the same line, Bawa and Ghosh (1999) found that larger families which have older heads and more unemployed members tend to shop groceries more frequently.

The various forms of promotion used by grocery stores (e.g., coupons, point of purchase displays, etc.) have been known to influence the timing of product purchases (Gupta, 1988, 1991; Gönül and Srinivasan, 1993; Helsen and Schmittlein, 1992, 1993; Jain and Vilcassim, 1991; Neslin et al., 1985). Gupta (1988) studied the "bump" in sales during promotions by breaking down the sales into brand switching, purchase time acceleration, and stockpiling. He found that eighty four percent of this "bump" in sales comes from brand switching while fourteen percent of it comes from purchase timing acceleration. Neslin et al. (1985) broke purchase acceleration down into acceleration in quantity purchases and acceleration in purchase timing. They found positive evidence of purchase timing affecting the increase in purchase. Jain and Vilcassim (1991) delved deeper into the methodological side and tested the effect of using various probability distributions for modeling interpurchase timing, while replicating the results of previous studies regarding the effect of promotions on purchase timing. Similarly, Gönül and Srinivasan (1993) replicated these results while improving upon the previous modeling approaches used and taking new econometric directions. The knowledge of the day of the week used for grocery shopping could help grocery store managers improve the timing of offering promotions by making it coincide with the customers' purchase timing. Specifically, in this paper, we make a contribution to this literature by identifying which segments of consumers choose their items of purchase on certain days of the week. Based on this, grocery store managers could target their promotional activities to those segments on the particular days of the week in order to motivate such consumers to buy more.

### 2.2. Grocery shopping behavior and the choice of paired products

The general grocery shopping basket comprises several products. Based on the study by Farhangmehr et al. (2000), these products could be grouped into food and drink products, personal care products, and cleaning products groups. Following the literature on network effects, Slater and Olson (2002) emphasized the importance of 'complementors' in the low-tech industry (e.g., razors and shaving cream). In their view, these complementors are an essential aspect of profitability for firms. Further, the issue of cross-category pairing becomes especially important as firms try to increase their brand equity through brand extensions (Kamakura and Kang, 2007). These brand extensions could be accomplished by introducing a new product (e.g., a brand of toothbrush with toothpaste). Based on these rationales, it is increasingly more essential for firms to understand product pairing nuances. In line with this view, different studies have addressed various forms of pairing in the grocery shopping context.

Arora (2011) studied pairing in the context of mouth hygiene/personal care products (i.e., toothbrush, toothpaste, dental floss, and mouthwash). He found positive evidence for the value of pairing these products in the view of the buyer. Simonin and Ruth (1995) similarly described the components of paired items in terms of a 'new' and a 'tie-in' product. They found positive evidence for the favorability of pairing toothbrush and toothpaste.

Soft drink consumption is a highly visible and important aspect of food choice behavior. Studies in the realm of human health have found that there is a positive correlation between the consumption of soft drinks and high-calorie foods such as pizza and hamburgers (see Vartanian et al., 2007 for a meta-analysis). Janiszewski and Cunha (2004), similarly, emphasized the pairing of these products in the form of a 'focal' product (e.g., pizza) and a 'tie-in' product (e.g., coke).

In a study of complementary and substitute products, Chintagunta and Haldar (1998) studied the pairing of powder and liquid detergent products. They found evidence that some of the consumers buy these two items together. Similarly, Song and Chintagunta (2006) investigated category-related factors for buying powder and liquid detergent and found statistically significant correlations between these cleaning items.

Our work adds to the extant literature by filling the gap in consumers' grocery shopping choice patterns for different groups of products on different days of the week. In a probabilistic setting, we use an advanced latent class analysis technique to identify segments of buyers for the three main groups of grocery products. In each of our segments, mixing probabilities are determined based on buyers' demographic characteristics; therefore, they enable us to discover the day of the week when buyers shop for different product groups.

## 3. Theoretical background and model

As discussed in previous sections of this paper, we aim at investigating the unobserved heterogeneity among consumers' grocery purchase choices for different items bought on different days of the week. More specifically, as summarized earlier, Farhangmehr et al. (2000) show that most consumers buy cleaning, personal care, and food/drink products during their grocery store shopping trips. We, therefore, focus on these major groups of grocery products choice and select products which belong to each of these three product groups. Hence, in our empirical analysis, by uniting different groups of a grocery shopping trip, we investigate food and drinks (Raijas, 2002), personal care (Yeon Kim and Chung, 2011), and cleaning (Farhangmehr et al., 2000) as the three main groups of products consumers buy in a shopping trip.

Further, we select a pair of products (Chintagunta and Haldar, 1998) from each aspect to perform our analysis. Choice of products within each pair is supported by two main rationales: (1) in our scanner data, items within each pair are the most commonly purchased ones by the consumers for each aspect and show very strong correlation for being bought together and (2) a review of the literature shows that items in each pair are purchased together from a grocery store. Based on previous studies, pizza (frozen) and carbonated beverages (De Castro, 2009; Coon et al., 2001; Paeratakul et al., 2003; Vartanian et al., 2007; Bruening et al., 2014; Mathias et al., 2013), toothbrush and toothpaste (Arora, 2011; Kamakura and Kang, 2007; Simonin and Ruth, 1995), and powder and liquid detergents (Chintagunta and Haldar, 1998; Song and Chintagunta, 2006) are items which are most frequently bought together.

We, therefore, develop and estimate an advanced latent class model which allows controlling for consumer unobserved heterogeneity through demographic variables and helps unveil consumer grocery shopping choice on different days of the week for each of those three groups of products separately. In order to do so, we use a Finite Mixture Model (FMM) to identify the latent classes for each pair of products. Such a setting enables us to identify probability distribution of subpopulations (i.e., segments) within our data. It also enables us to estimate the size of each latent class in the data. An advantage of using a more advanced latent class analysis model (i.e., FMM) over the well-known Kamakura and Russell (1989) is that it provides more flexibility in determining the mixing probabilities as it allows them to be a function of consumer demographics, while it lets the days of the week impact share of money spent on each pair. Consumer demographics are then used as descriptors for each latent class. Following similar studies that tackle a segmentation problem (see, for example the very well-cited, Furse et al., 1984; Kamakura and Russell, 1989), we do not use a hypothesis-testing structure in this research. Similar to other segmentation studies, we take a descriptive approach to developing our model where shoppers are assigned to various segments based on their estimated probabilities of membership for each latent class (i.e., mixing probabilities). Controlling for unobserved heterogeneity, these mixing probabilities are estimated based on the demographic information for each customer and promotional offerings by the retailer. As explained at length in our Section 4, the scheduling of these promotions varies over time.

In our FMM setting, we set the dependent variable as the share of money spent on each pair relative to the total transaction dollar amount per shopping trip. Our FMM setting is based on a well-established latent class analysis and segmentation marketing approach introduced by Rossi et al. (2012). In our setting, latent classes (i.e., segments) are identified through a number of multivariate normal distributions. Each multivariate normal distribution (i.e., mass point) has a vector of means as well as a variance-covariance matrix. We aim at identifying the proper number of such mass points and estimating their mean and variance-covariance matrices. Additionally, the model allows controlling for unobserved heterogeneity through a random coefficient setting via a class-specific vector of $\beta s$, slopes, which are to be estimated. The unobserved heterogeneity and latent classes in our data come from consumer demographic characteristics, which impacts size and membership probability for each latent class for each pair of products.

Eq. (1) presents the joint density for the FMM model:

$$f(y_j;\, \pi_1, .\pi_2, ..,\pi_w, \mu_1, \mu_2, \ldots ,\mu_w, \Sigma_1, \Sigma_2, \ldots ,\Sigma_w) = \pi_1 g_1(y_j;\, \mu_1, \Sigma_1) + \pi_2 g_2(y_j;\, \mu_2, \Sigma_2) + \ldots + \pi_w g_w(y_j;\, \mu_w, \Sigma_w) \tag{1}$$

In Eq. (1), $f(.)$ is the conditional probability density function and $y_j$ is the ratio of dollar value spent on each pair of products to the total transaction dollar amount, which represents the dependent variable. $\pi_i$ s are the mixing probabilities, which are between zero and one and add up to 1 across all latent classes (i.e., $\sum_{l=1}^{w} \pi_l = 1$ for each pair) and have multinomial logistic distribution. $g_i(y_j;\, \mu_i, \Sigma_i)$ are multivariate normal distributions with means $\mu_i$ and variance-covariance matrix of $\Sigma_i$ for each class. Additionally, $w$ represents the number of latent classes to be determined. Finally, $l$ is the number of random coefficients (i.e., slopes) in each class, which is six in our model and controls for seven days in a week. The goal is to identify the proper number of latent classes (segments) for each pair and estimate the vector of $\beta$ coefficients for each class. In other words, our investigation unveils the latent segments of customers based on their demographics who choose to buy certain items on certain days of the week on their grocery shopping trip.

More specifically, in our setting, we use days of the week as variables in our main equation, while mixing probabilities are guided by a list of 21 variables from seven categories. As indicated in the Section 2, the choice of our demographic variables is guided by previous studies done in this domain. We, therefore, use the demographic variables as family size, ethnicity, income, homeownership status, employment status, and age. We, next, control for promotions and price discounts that might have been offered to consumers during their grocery shopping trip in our mixing probabilities equation.

In other words, Eqs. (2) and (3) present the main and mixing probability equations for each product pair in our model; we, therefore, have three sets of equations, one set per pair.

$$y_{ijkw} = Mon_{ijkw} + Wed_{ijkw} + Thu_{ijkw} + Fri_{ijkw} + Sat_{ijkw} + Sun_{ijkw} + \varepsilon_{ijkw} \tag{2}$$

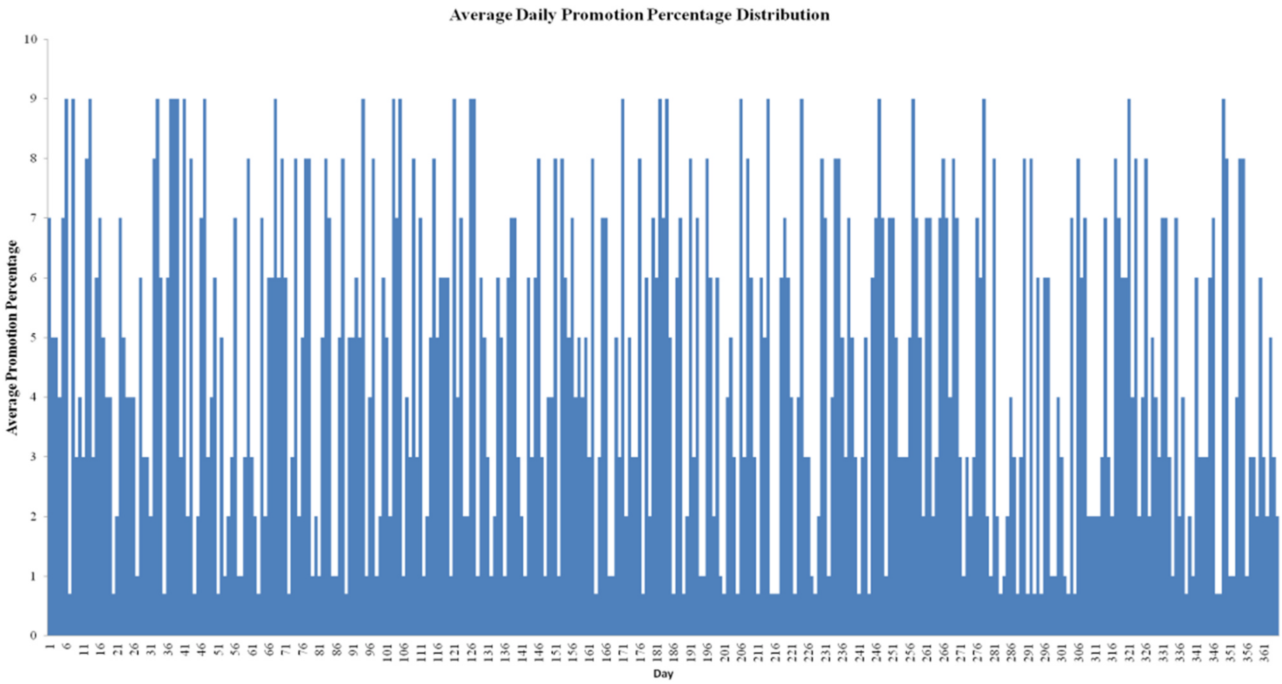


**Fig. 1.** Distribution of average daily promotion percentage.

$$\begin{aligned}\pi_{jw} = \; & Fam_{ijkw} + Asian_{ijkw} + Hisp_{ijkw} + White_{ijkw} + Pmtion_{ijkw} \\ & + Incl20_{ijkw} + Incl2030_{ijkw} + Incl3040_{ijkw} + Incl4050_{ijkw} \\ & + Incl6080_{ijkw} + Incl80100_{ijkw} + Incl100125_{ijkw} + Incl125150_{ijkw} \\ & + Incl150200_{ijkw} + Incg200_{ijkw} + Homeown_{ijkw} + Workft_{ijkw} \\ & + Homemakr_{ijkw} + Workpt_{ijkw} + Rtired_{ijkw} + Age_{ijkw} + \varepsilon_{ijkw}\end{aligned} \tag{3}$$

In Eq. (2), $y_{ijk}$ is the ratio of dollar amount spent on pair $j$ to the total bill amount by customer $i$, on date $k$ for class $w$, in which $j = 1,2,3$. Each of the $j$'s represents one pair of products related to each major group of grocery products (food/drinks, cleaning, and personal care) and is an indicator of one of the three sub-datasets used for this analysis. The independent variables are dummies controlling for different days of the week, keeping Tuesday as the baseline. Finally, $\varepsilon_{ijk}$ is an error term with standard normal distribution.

In Eq. (3), $\pi_{jw}$ is the mixing probability for segment $w$ of pair $j$. $Fam_{ijkw}$ is the number of people in the household of customer $i$. The next three variables are ethnicity dummy variables (Asian, Hispanic, and White), using Black as the baseline. $Pmtion_{ijkw}$ represents the dollar amount of discount received in purchasing pair $j$. Our next 12 variables measure different levels of customer income, starting from less than \$20,000 to over \$200,000 annual income, using income between 50,000 and 60,000 as the baseline. $Homeown_{ijkw}$ is a dummy variable which takes on the value of 1 if the customer is a homeowner. The next four variables are dummies which control for working status of the customer as working full-time, working part-time, being a homemaker, or retired, using unemployed as the baseline. Finally, $Age_{ijkw}$ is the customer age. Details on how $w$ is determined through empirical tests are explained in the Section 5 of the paper.

## 4. Data

The data used for this study are scanner panel data, which include customer grocery shopping data of two large metropolises in the U.S. in year 2007 with over 3.61 million total observations, covering 36,237 unique items. Our scanner dataset comes from a large grocery store chain in the U.S., where it is a major player in the grocery shopping industry and holds a sizable market share in the grocery store market in the country; hence, it may be used as an appropriate representative of the grocery store market. Also, our data cover consumer grocery shopping choice in two large metropolises in the U.S., where both areas are heavily populated and incorporate very diverse demographics in a wide range of aspects. Therefore, the data, to a large extent, may serve as a representative of consumers who reside in other urban areas and their grocery shopping behavior. These two important attributes of our dataset make it a reasonable tool for the purpose of our empirical analyses and make us believe that findings developed based on our data serve as a reliable representative of urban consumers' grocery shopping choice and behavior. That, would in turn, make our findings generalizable to other urban consumers and grocery store locations.

Our data contain information from multiple grocery items, cover details of each transaction by each customer as he/she checked out of the store, and is marked anonymously with a customer ID. For instance, if a customer purchases five items in a single shopping trip to the store, the data shows this information on five separate rows marked by the dates of corresponding purchases for that particular customer. Additionally, we have information on customer demographics, type and number of items purchased, price paid for each item, whether the customer used any form of promotion on each item within the transaction, date of purchase, and total dollar amount of the bill. It is noteworthy that, in line with grocery stores' policies, the pattern of promotions in our dataset does not remain constant. In other words, grocery store managers use offering schedules that vary over time. Particularly, according to our data, average daily promotion depth (i.e., percentage) varies between 0.7% and 9.3%, with the mean of 4.6, while average weekly promotion depth is between 2.6% and 6.7%, with the mean of 4.7. Figs. 1 and 3 illustrate the distribution of average daily and weekly promotion percentages across the entire 365-day period, respectively. As both figures show, no specific pattern of promotion depth is evident in our data at the daily or weekly levels; hence, no scheduled pattern of promotion depth offerings has been exercised by our retailer at either time periods. Additionally, to investigate variations in promotion counts over time, we looked into our data for number of unique products on sale for each day and week by the focal retailer. We observe that between 1.4% and 3.7% of total unique items sold by our store were on sale (i.e., on promotion) each day with the mean of 2.5%.

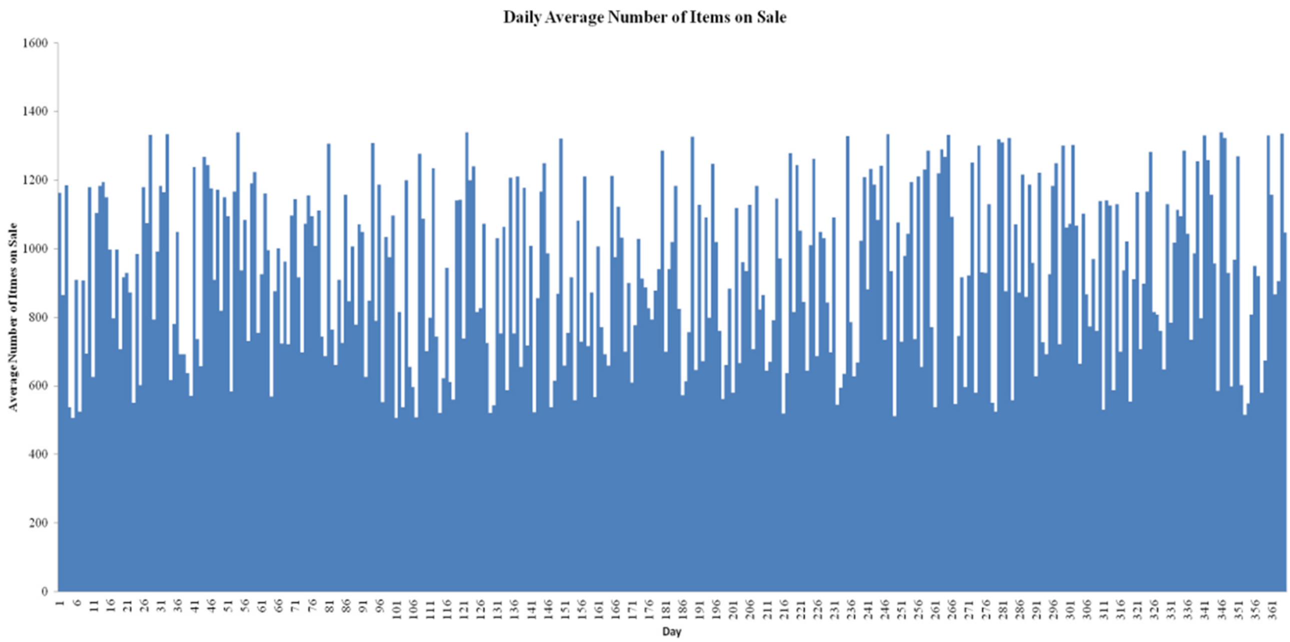


**Fig. 2.** Distribution of daily average number of items on sale.

Similar measures at the weekly level are 2.1% and 2.8%, followed by a mean of 2.4. Figs. 2 and 4 summarize the distribution for number of unique items on sale on each day and on each week across the entire time period, respectively. Following such analyses, no specific pattern of promotion counts schedule is observable in our data either.

Further, to ensure that this policy is not unique to our dataset and is indeed a policy opted for by all grocery store managers, we conducted interviews with a number of grocery store managers including those who provided the data. These managers confirmed that an offering schedule varying based on speculations about demand is indeed what is used in practice. Further, an article on promotion planning by Cohen et al. (2017) delves deep into the current promotion scheduling of grocery stores and offers a quantitative procedure for planning promotions. Based on this article, the current policy of grocery store managers is to plan the promotions' schedules manually using "what-if scenarios". These scenarios are based on speculations about shifting demand and try to predict the potential response of customers to promotion planning. In other words, managers' decisions on offering a promotion on an item (vs. not offering promotion), offering promotions on a pair of items, and how to schedule and order these offerings are based on "what-if scenarios" using managers' intuition of customers' demand and habit (Cohen et al., 2017; Bogomolova et al., 2017; Perakis, 2017). In other words, temporal tweaking of the promotions' schedules is a way for managers to fine-tune their promotional offerings (Dun and Bradstreet First Research, 2013, 2018). Using the transaction dates and developing a computer program, we identified day of the

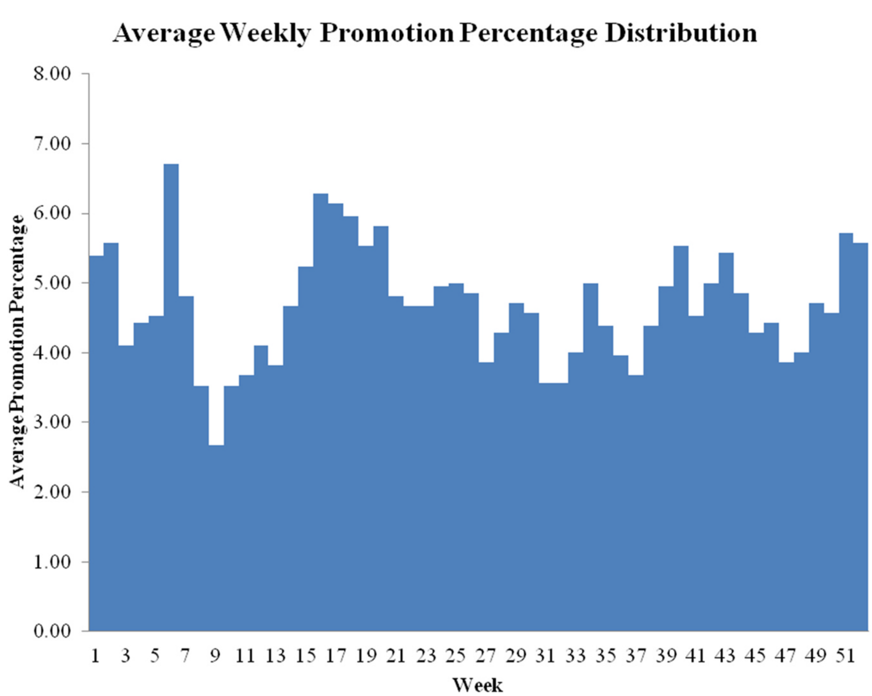


**Fig. 3.** Distribution of average weekly promotion percentage.

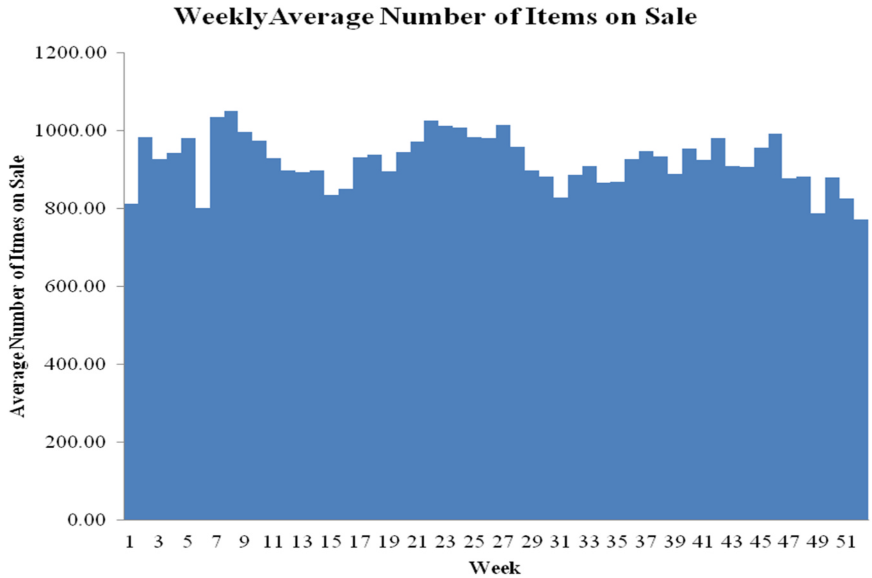


**Fig. 4.** Distribution of weekly average number of items on sale.

week on which each transaction was recorded throughout the dataset.

In order to investigate the impact of the day of the week on customers' purchasing behavior, we had to reshape the data choosing customer as the basis. We performed that by developing codes using Visual Basic for Applications (VBA). In our coding, we made sure that different shopping transactions for each customer on different dates are treated separately. We find that during each shopping trip, each customer purchased on average about 27 items from the grocery store and that there are over 2631 unique customers in our dataset. We, next, divided our main dataset into three subsets. Each subset of data includes a pair of products, representing one of the three major groups of grocery products, as explained in the previous section of the paper.

For each pair of products, we specifically look into the correlation of transactions between the incidents during which at least one of the two items are purchased. The estimated correlations for all three pairs are significant and are presented in Table 1. The high correlations, in-line with the literature, confirm the meaningful grouping of those items for each pair.

We further analyze the data in terms of percentage of trips relative to total number of trips for each customer purchasing each of our pairs; next we average those percentages for each pair across all customers. Table 2 summarizes the average percentage of trips for each pair of products across all customers during which at least one item within each pair was purchased.

**Table 1**
Correlation results for purchased incidents.

| Product group | Correlation |
|---|---|
| Pizza and Carbonated beverages | 0.872a |
| Powder and liquid detergent | 0.814b |
| Tooth-brush and Tooth-paste | 0.893a |

**Table 2**
Percentage of shopping trips.

| Product group | Percentage of shopping trips |
|---|---|
| Pizza and carbonated beverages | 46.3% |
| Powder and liquid detergent | 38.5% |
| Tooth-brush and tooth-paste | 32.2% |

## 5. Estimation, results, and discussion

As it is common to use for panel datasets, our model controls for unobserved heterogeneity through random coefficients. Following Rossi (2014), for such data, a fully non-parametric approach would not be applicable; instead a semi-parametric approach should be used, which in our case is a mixture of multivariate normal distributions. Specifically, the random coefficient element of the model uses an unrestricted multivariate normal model. In a finite mixture of normals, a standard data augmentation with the indicator of the normal component can be used (Rossi, 2014). Using this indicator enables us to rely on a normal prior for each cross-section (Rossi, 2014).

Our advanced latent class model is estimated for each of the three pairs of products using maximum likelihood estimation. The likelihood estimation is based on the assumption that each response variable is iid distributed across the estimation sample conditional on each latent class. Maximum likelihood estimation is performed by combining each latent class conditional likelihood, using probabilities as weights, and is conducted in two stages: first, we estimate the probability-weighted conditional likelihoods for each latent class within each pair, and next, we take the summation of those numbers to estimate the overall likelihood value for each group. We employ the AIC (Akaike, 2011), the BIC (Schwarz, 1978), and the Delta Method (Oehlert, 1992) as our criteria in identifying the number of latent classes for each product pair.[1] Additionally, for each latent class within each product pair, we estimate (1) the class-specific random effects coefficients per variable impacting the main model, (2) the class-specific coefficients per variable impacting the mixing probabilities, and (3) the class sizes. In the next part, we present and discuss the results for each product pair. Eqs. (4) and (5) present the two-level likelihood functions. $f_{jwt}^{(1)}(y|X, \beta)$ represents the conditional likelihood of customer $j$ at time period $t$ for each class $w$.

$$f_{jw}^{(2)}(y|X, \beta) = \int \prod_{t=1}^{Tj} f_{jwt}^{(1)}(y|X, \beta) g_w(\beta; \mu, \sum) d\beta_{jw} \tag{4}$$

$$L_w(y, X, \beta) = \prod_{j=1}^{N} f_{jw}^{(2)}(y|X, \beta) \tag{5}$$

Further, for each latent class $w$, class size (i.e., probabilities), is estimated using Eq. (6):

[1] A detailed and step-by-step process of these applications is provided in the Appendix A.

$$\pi_w = \frac{\exp(z_w)}{\sum_{l=1}^{W} \exp(z_l)} \quad (6)$$

where $z_w = X'\tau_w$ and $\tau_w$ is the vector of coefficients. Taking the first latent class as the base class, we will have $\tau_1 = 0$, hence, $z_1 = 0$.

Finally, we use the *fmm* command in Stata to perform our analysis. Following the estimation procedure for this command, we, first, run a model in which mixing probabilities are not guided by demographics in order to apply the estimation results for the complete model. For the initial model, in order to get the starting values, our estimation process employs the EM algorithm for the starting values. It assumes all values are *observable* for the latent classes, hence developing a full likelihood function and estimating it based on that. Then, it uses the result of the simplified model as an input for the full model, which uses the mixing probabilities guided by demographics, to maximize log likelihood and continues the estimation as an iterative procedure for the rest.

### *5.1. The food and beverage group of grocery products (pizza and carbonated beverages)*

Using the BIC and Delta Method, we identify two latent classes (segments) with sizes of 81% and 19% for pizza and carbonated beverages. In the larger segment, we find that more money relative to total expenditure is spent on Friday, Saturday, and Sunday for the food and beverage group. Further, we find that customers who belong to this larger segment are more likely to increase their relative spending in presence of promotions and are more likely to work full-time. We also find that customers of this segment are less likely to be Asian and are less likely to make over $150,000 of annual income. For the smaller segment, however, we find that these customers are more likely to buy pizza and carbonated beverages on Friday but are less likely to buy those on Thursday. Also, we find that people in this segment have larger family sizes, are more likely to be white, make less than $40,000 a year, and work part-time.

These interesting findings can guide the store manager in deciding what other items should be promoted as well as pricing techniques employed for other products in the store. For instance, the store can benefit from offering promotions on pizza and carbonated beverages on Friday and Saturday to draw wealthier customers (class 1) to the store. Then, with those customers present in the store, managers can offer higher prices on other items, hence, improve their revenue and profitability. For the second segment, given their importance of family sizes and lower levels of income, the store can increase the sales of other products through offering lower priced items such as snacks or larger packs of items, i.e., “value packs” on Thursday to benefit from those customers’ presence in the store while they consider buying pizza and beverages.

Further, offering promotions has a significant effect on the purchase decision of Class 1. In contrast, promotions do not have a significant effect on the purchase decision of Class 2. Considering that Class 1 is more likely to make their purchase decisions on Fridays, Saturday, and Sundays, it is better to target promotions to attract customers in this class on those days.

Tables 3a and 3b summarize the estimation findings for pizza and carbonated beverages.

### *5.2. The cleaning group of grocery products (powder and liquid detergent)*

The BIC and Delta method guided us in identifying two segments (classes) for powder and liquid detergents; i.e., the cleaning group of grocery products. Unlike pizza and carbonated beverages, which had two classes with significant differences in size, this group turned out two classes of relatively the same size: 54% and 46%. For the first class (54% in size), the cleaning pair is purchased on Saturday. Customers who make more than $150,000 a year and those who work full-time are more likely to belong to this class. Customers in the second class (46% in size), however, purchase more cleaning items on Sunday. People in this class respond more positively to promotions, make less than $30,000 annually, and are less likely to be homeowners.

While the sizes of these two classes for the cleaning group are more similar, our findings can be used to shed light on managerial decisions for retail managers. According to the results of our model, while customers in the first class may not significantly respond to promotions, they are more likely to belong to a more affluent segment of customers. Therefore, the retailer can avoid offering promotions for the cleaning group on Saturday and generate higher cash flows through selling those items at higher prices. The retailer can also offer more premium products with less/zero discounts to generate more revenue from this wealthier segment of the market. For the second class, however, the retailer should offer more discounts on the cleaning group items on Sunday, and since the customers in this segment are not so well-off, offering cheaper products and/or different types of promotions (coupons, discounts, BOGO, etc.) could be implemented on those days to increase demand from this segment.

Unlike the previous grocery category of grocery products, in this category, offering promotions has a significant effect on the purchase decision of Class 2, but not for Class 1. Given that customers in this class are more likely to make their purchases on Sundays, it is a better

**Table 3a**
Pizza and carbonated beverages day of a week.

| Day of the week | Class 1 | Class 2 |
|---|---|---|
| Monday dummy | −0.531 | 0.205 |
| Wednesday dummy | 1.394 | −2.108 |
| Thursday dummy | 1.117 | −3.228[a] |
| Friday dummy | 6.281[a] | 6.164[b] |
| Saturday dummy | 4.937[b] | 7.314 |
| Sunday dummy | 5.064[b] | −0.683 |

Log likelihood = -6349.652.
Number of observations = 39,274.
[c] $p < 0.10$.
[a] $p < 0.01$.
[b] $p < 0.05$.

**Table 3b**
Pizza and carbonated beverages mixing probabilities.

| Variable | Class 1 | Class 2 |
|---|---|---|
| Family Size | 6.379 | 6.731[a] |
| Asian | −1.348[a] | 4.628 |
| Hispanic | −0.360 | 0.028 |
| White | 0.983 | 1.791[b] |
| Promotion | 2.537[b] | 1.374 |
| 0–19,999 | 1.934 | 7.649[c] |
| 20,000–29,999 | 6.331 | 5.791[a] |
| 30,000–39,999 | 5.974 | 8.437[a] |
| 40,000–49,999 | 0.096 | 1.008 |
| 60,000–79,999 | 3.098 | 2.172 |
| 80,000–99,999 | 2.361 | 1.083 |
| 100,000–124,999 | 1.369 | 2.317 |
| 125,000–149,999 | 0.248 | 3.065 |
| 150,000–199,000 | −3.649[c] | 0.054 |
| 200,000 + | −5.694[b] | 6.379 |
| Own | 2.061 | 3.361 |
| Work Full Time | 7.367[a] | 2.317 |
| Homemaker | 1.942 | 6.810 |
| Work Part Time | 6.614 | 6.791[b] |
| Retired | 4.120 | 3.781 |
| Age | 1.129 | 9.617 |
| Class size | 81% | 19% |

AIC = 12,753.304.
BIC = 12,984.919.
[a] $p < 0.01$.
[b] $p < 0.05$.
[c] $p < 0.10$.

**Table 4a**
Powder and liquid detergent day of a week.

| Day of the week | Class 1 | Class 2 |
|---|---|---|
| Monday dummy | 6.339 | 9.046 |
| Wednesday dummy | 7.596 | −2.311 |
| Thursday dummy | 3.466 | 5.380 |
| Friday dummy | −4.788 | 1.536 |
| Saturday dummy | 1.338[a] | −9.124 |
| Sunday dummy | 0.349 | 2.273[b] |

Log likelihood = -4412.937.
Number of observations = 27,757.
[c] $p < 0.10$.
[a] $p < 0.01$.
[b] $p < 0.05$.

**Table 4b**
Powder and liquid detergent mixing probabilities.

| Variable | Class 1 | Class 2 |
|---|---|---|
| Family Size | 2.064 | 1.057 |
| Asian | −0.081 | −2.007 |
| Hispanic | 1.349 | 0.021 |
| White | −0.080 | −0.251 |
| Promotion | 0.271 | 6.319[a] |
| 0–19,999 | 2.647 | 3.661[a] |
| 20,000–29,999 | 3.581 | 9.034[b] |
| 30,000–39,999 | 1.073 | −6.057 |
| 40,000–49,999 | 3.578 | −2.314 |
| 60,000–79,999 | 0.093 | −8.317 |
| 80,000–99,999 | −12.371 | 10.564 |
| 100,000–124,999 | 1.930 | 1.082 |
| 125,000–149,999 | 5.592 | −9.349 |
| 150,000–199,000 | 4.338[a] | 2.904 |
| 200,000+ | 2.148[c] | 8.827 |
| Own | 1.060[a] | −2.247[b] |
| Work Full Time | 5.611[b] | −9.012 |
| Homemaker | 4.189 | 3.629 |
| Work Part Time | 0.015 | 0.003 |
| Retired | 3.824 | 4.560 |
| Age | 0.573 | 1.279 |
| Class size | 54% | 46% |

AIC = 8879.874.
BIC = 9102.118.
[a] $p < 0.01$.
[b] $p < 0.05$.
[c] $p < 0.10$.

strategy for the grocery store manager to target promotions to attract this segment on Sundays.

Tables 4a and 4b illustrate the estimation results for the cleaning product group.

### *5.3. The personal care group of grocery products (toothbrush and toothpaste)*

For the personal group of grocery products, the BIC and Delta Method helped us identify three latent classes (segments) of shoppers; the first class is significantly larger (67%) while the other two have smaller sizes of 21% and 12%.

The first class comprises customers who buy more personal care items on Saturday and Sunday. Larger family sizes are more likely to belong to this class. Also, customers in this class respond more positively to promotions and are more likely to make less than $40,000 a year. Customers in this class also are more likely to work both full-time and part-time but are less likely to make between $60,000–80,000 a year. Hence, it may be fair to conclude that customers in this segment are the upper-middle class segment of the market.

Our second class represents customers who buy more personal care items on Monday, but less of those on Wednesday. These people are more likely to make more than $80,000 and are more likely to work full-time or be retired. For our third class, we did not find any evidence that the days of the week have a significant impact on customers' purchase decision on the personal care pair. These customers, who seem to be insensitive to the day of the week for shopping personal care items, are more likely to be Asian or white and respond more positively to promotions. They are, also, less likely to make over $150,000 a year. Offering promotions in this category affects the purchase decision of Class 1 and Class 3 only. However, promotions have a stronger effect on the purchase decision of Class 3 than Class 1. Class 1 is more likely to purchase personal care items on Saturdays and Sundays, and therefore, promotions should be targeted to attract this segment on those days. In contrast, while Class 3 is promotion-sensitive, it does not show any preference for any particular day of the week. Therefore, it is not possible for the grocery store manager to plan the timing of the promotions for this segment.

Managers can use our findings to make in-store decisions regarding promotion and pricing of their items. For the first class, which is larger in size, less affluent, and more price sensitive, mangers can offer different forms of promotions on Saturday and Sunday for personal care items and make a wider range of lower priced products available for this group on those days. For the second class, which does more shopping on Monday and less on Wednesday, and is wealthier, managers can avoid offering promotions and discounts on either day and, at the same time, provide a larger number of higher priced items on those days. Given the lower price sensitivity of this class, such decisions could help the store manager in generating more revenue on these days from this group of customers. Finally, while the third class does not seem to have a specific day for their personal care shopping, considering their positive response to promotions, managers could pick a day and offer deeper than usual promotions on that day to create a conditioning effect. They might observe an increase in store traffic from this group on the selected day of the week after some time.

Tables 5a and 5b show a summary of estimation results for the personal care product group.

## 6. Conclusion, managerial implications, and further research

Marketing strategies are becoming more and more individualized in various arenas including retail. Grocery shopping, in particular, is moving towards micro-marketing and offering more customized deals to customers (Nielsen, 2015). The present research offers details about hidden segments of grocery shoppers by unveiling *which types* of customers are interested in *what* type of products and *when* they purchase them. In particular, demographic characteristics of grocery shoppers have been known to affect grocery shopping behavior, so has the choice of the day of the week when the purchase is done. However, there is a gap in this literature about identifying the timing of grocery shoppers' purchase of different groups of products guided by demographic variables. In this study, we uncover latent classes of grocery shoppers based

**Table 5a**
Tooth brush and tooth paste day of a week.

| Day of the week | Class 1 | Class 2 | Class 3 |
|---|---|---|---|
| Monday dummy | 0.267 | 1.368[a] | 1.864 |
| Wednesday dummy | 2.364 | −2.702[b] | 4.975 |
| Thursday dummy | −1.773 | 1.461 | 3.648 |
| Friday dummy | −2.015 | −0.907 | 2.134 |
| Saturday dummy | 2.363[a] | 1.945 | −3.647 |
| Sunday dummy | 2.118[b] | 2.684 | −1.349 |

[c] $p < 0.10$.
Log likelihood = -5512.391.
Number of observations = 49,967.
[a] $p < 0.01$.
[b] $p < 0.05$.

**Table 5b**
Tooth brush and tooth paste mixing probabilities.

| Variable | Class 1 | Class 2 | Class 3 |
|---|---|---|---|
| Family Size | 1.736[a] | 3.227 | 1.561 |
| Asian | 0.018 | −0.250 | 2.664[a] |
| Hispanic | 0.204 | 1.337 | 0.334 |
| White | 0.112 | 3.224 | 1.225[b] |
| Promotion | 0.969[b] | 1.281 | 1.134[b] |
| 0–19,999 | 1.664[c] | 1.084 | 2.526 |
| 20,000–29,999 | 1.638[b] | −4.622 | 2.447 |
| 30,000–39,999 | 2.364[b] | 1.637 | 3.228 |
| 40,000–49,999 | 0.047 | 1.327 | 2.168 |
| 60,000–79,999 | −2.355[a] | 2.364 | 3.281 |
| 80,000–99,999 | 2.843 | 1.966[c] | 2.764 |
| 100,000–124,999 | 1.368 | 1.731[c] | 0.089 |
| 125,000–149,999 | 4.263 | 2.446[a] | 1.367 |
| 150,000–199,000 | 3.264 | 3.661[a] | −3.229[b] |
| 200,000+ | 2.367 | 2.350[b] | −2.148[a] |
| Own | 1.367 | 0.937 | 1.005 |
| Work Full Time | 0.086[b] | 0.668[b] | 1.972 |
| Homemaker | 1.295 | −3.648 | 2.167 |
| Work Part Time | 0.058[b] | 2.649 | 0.157 |
| Retired | 1.364 | 0.822[a] | 2.384 |
| Age | 0.364 | 1.237 | −4.974 |
| Class size | 67% | 21% | 12% |

AIC = 11,078.782.
BIC = 11,316.898.
[a] $p < 0.01$.
[b] $p < 0.05$.
[c] $p < 0.10$.

on their demographic characteristics and the day of the week when the purchase took place for three major groups of products that are bought during a grocery shopping trip. Our analyses build the foundation for targeting promotional activities for different segments of consumer demographics on different days of the week. To the best of our knowledge, this is the first comprehensive study which empirically investigates the connection between day of the week and customer grocery shopping behavior in major groups of grocery products: food/drink, cleaning, and personal care. Our state-of-the-art empirical technique enables us to unveil the latent classes of customers within the three grocery items groups which could open up new decision making veins for grocery store managers.

Our study contributes to the literate in several ways. First, from a methodological perspective, we apply an advanced finite mixture latent class model to a customer grocery shopping behavior dataset and empirically validate our model using this dataset which comes from national grocery retailers. Second, our results have important managerial implications for grocery stores. Managers can use our findings to make pricing and promotion targeting decisions on different days of the week for different product groups and expect to improve their control over offering deals which could also assist them with store traffic management. Implications of our findings for promotion targeting decisions, include, but are not limited to, timing, depth, and frequency of promotions. Additionally, managers can use our findings in setting pricing policies for their products and in making product quality decisions for items they offer within each product group. For example, one such decision could be about choosing between premium prices for well-known items vs. more affordable prices for lesser-known items guided by the day of the week when a particular customer with a specific grocery shopping budget visits the store. Furthermore, based on the expected demand on different days of the week, managers can decide on the product assortments they offer, matching the characteristics of the customers who visit the store on that particular day. Third, through our findings, we shed light on promotion decisions which can help managers in planning their store layout, such as, display decisions, more (less) visible and easy (hard) to access shelf/aisle assignment, and location decisions for items which are more/less likely to be purchased together in the store. Finally, our findings have interesting implications for academicians who are interested in better segmentation and targeting schemes of the grocery shoppers market using econometrics tools that allow for controlling for multiple sets of variables describing these shoppers. This is especially interesting as most of the studies in this domain focus on one set of homogenous factors to segment the grocery shoppers market as opposed to various sets of variables which could draw a more detailed picture of these segments and the customers in them.

A major limitation of the current study is the use of data from only two metropolises in the US. Such data may not necessarily be a perfect representative sample for the whole US population, especially, for those customers who live in less urban areas. Demographics of shoppers, their response to promotions, and their shopping trip schedules may not necessarily be the same for a customer who lives in a big city vs. one who lives in a more rural area. Further research, with more geographically diverse data, could replicate our model and validate the generalizability of our findings for multiple locations across the country.

Additionally, a recent study by The Nielsen Company (2015) revealed a trend in which millennial shoppers are more and more drawn to online grocery shopping. While the size of this segment in the U.S. is small (about 3.3%), it is expected to rise to a size of about 11% by 2023 (Melis et al., 2015). Our current dataset does not include any online shopping information. A richer dataset with information about online grocery shopping could increase the quality and generalizability of our results.

## Appendix A

In this appendix, we provide a summary of the processes used to determine the number of latent classes for one of our pairs, pizza and carbonated beverages, as an example. In this step-by-step process, we cover the application of AIC and BIC values, as well as the Delta Method in deciding on the number of latent classes in our setting for this pair. Similar procedures have been performed for the other pairs.

As the first step, and in order to be able to develop initial values for the full model, we run the model with a single latent class in which mixing probabilities are not guided by demographic variables. Next, we use the results of this simplified model as the input for a single latent class model in which mixing probabilities are guided by demographic variables. Further, we calculate the AIC and BIC for the single latent class model. Numbers are shown in Table A1.

**Table A1**
Log likelihood, AIC, and BIC for the single latent class model.

| Criterion | Value |
|---|---|
| Log likelihood | −7523.122 |
| AIC | 15,100.24 |
| BIC | 15,063.87 |

**Table A2**
Log likelihood, AIC, and BIC for the two-latent-class model.

| Criterion | Value |
|---|---|
| Log likelihood | −6349.7 |
| AIC | 12,753.304 |
| BIC | 12,984.919 |

**Table A3**
Log likelihood, AIC, and BIC for the three-latent-class model.

| Criterion | Value |
|---|---|
| Log likelihood | −6432.752 |
| AIC | 12,919.5 |
| BIC | 12,883.13 |

As the next step, we run our model with two latent classes and then calculate the AIC and BIC for the two-latent-class model. Results are shown in Table A2.

Both AIC and BIC values show lower amounts compared with the single-class model. Further, analyzing the latent classes, we calculate the class sizes, the Delta Method standard deviations (Oehlert, 1992), and test to see if their Delta Method variance significantly differs. The results confirm the significant difference between the two numbers. We, then, find the sizes for the two latent classes to be 0.19 and 0.81. Therefore, we can conclude that there are at least two latent classes for the pizza and carbonated beverages pair.

Next, we run the model with three latent classes. Results are summarized in Table A3

Comparing the results of the three-class model with the two-class case, we observe improvement in the BIC, but not in the AIC. We, then, calculate class sizes for the three-class model; results indicate class sizes of 0.86, 0.12, and 0.02. Given the inconsistency in BIC and AIC directions, and the considerably smaller size of one of the three classes, we perform analysis on the Delta Method standard errors. Test results show that standard errors in the second and third classes do not significantly differ; therefore, we conclude that the right number of latent classes for the pizza and carbonated beverages pair is two. A similar procedure has been used for the other two pairs in our analysis.

## References


Akaike, H., 2011. Akaike's information criterion. In: International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg.
Arora, R., 2011. Bundling or unbundling frequently purchased products: a mixed method approach. J. Consum. Mark. 28 (1), 67–75.
Bawa, K., Ghosh, A., 1999. A model of household grocery shopping behavior. Mark. Lett. 10 (2), 149–160.
Blaylock, J.R., 1989. An economic model of grocery shopping frequency. Appl. Econ. 21 (6), 843–852.
Bogomolova, S., Szabo, M., Kennedy, R., 2017. Retailers' and manufacturers' price-promotion decisions: intuitive or evidence-based? J. Bus. Res. 76, 189–200.
Brown, S., 1988. Information seeking, external search and 'shopping' behaviour: preliminary evidence from a planned shopping center. J. Mark. Manag. 4 (1), 33–49.
Bruening, M., MacLehose, R., Eisenberg, M.E., Nanney, M.S., Story, M., Neumark-Sztainer, D., 2014. Associations between sugar-sweetened beverage consumption and fast-food restaurant frequency among adolescents and their friends. J. Nutr. Educ. Behav. 46 (4), 277–285.
Chintagunta, P.K., Haldar, S., 1998. Investigating purchase timing behavior in two related product categories. J. Mark. Res. 43–53.
Cohen, M.C., Leung, N.H.Z., Panchamgam, K., Perakis, G., Smith, A., 2017. The impact of linear optimization on promotion planning. Oper. Res. 65 (2), 446–468.
Coon, K.A., Goldberg, J., Rogers, B.L., Tucker, K.L., 2001. Relationships between use of television during meals and children's food consumption patterns. Pediatrics 107 (1) (e7–e7).
De Castro, J.M., 2009. When, how much and what foods are eaten are related to total daily food intake. Br. J. Nutr. 102 (8), 1228–1237.
Doti, J.L., Sharir, S., 1981. Households' grocery shopping behavior in the short-run: theory and evidence. Econ. Inq. 19 (2), 196–208.
Dun & Bradstreet First Research, 2013. Grocery stores & supermarkets industry profile. Available at: ⟨https://iue.libanswers.com/data/.../grocery_stores_and_supermarkets_march_2013.pdf⟩ (Accessed 4 October 2018).
Dun & Bradstreet First Research, 2018. Grocery stores & supermarkets industry profile. Available at: ⟨http://www.firstresearch.com/Industry-Research/Grocery-Stores-and-Supermarkets.html⟩ (Accessed 8 October 2018).
East, R., Lomax, W., Willson, G., Harris, P., 1994. Decision making and habit in shopping times. Eur. J. Mark. 28 (4), 56–71.
Farhangmehr, M., Marques, S., Silva, J., 2000. Consumer and retailer perceptions of hypermarkets and traditional retail stores in Portugal. J. Retail. Consum. Serv. 7 (4), 197–206.
Fox, E.J., Hoch, S.J., 2005. Cherry-picking. J. Mark. 69 (1), 46–62.
Furse, D.H., Punj, G.N., Stewart, D.W., 1984. A typology of individual search strategies among purchasers of new automobiles. J. Consum. Res. 10 (4), 417–431.
Gönül, F., Srinivasan, K., 1993. Consumer purchase behavior in a frequently bought product category: estimation issues and managerial insights from a hazard function model with heterogeneity. J. Am. Stat. Assoc. 88 (424), 1219–1227.
Gupta, S., 1988. Impact of sales promotions on when, what, and how much to buy. J. Mark. Res. 342–355.
Gupta, S., 1991. Stochastic models of interpurchase time with time-dependent covariates. J. Mark. Res. 1–15.
Helsen, K., Schmittlein, D.C., 1992. How does a product market's typical price-promotion pattern. J. Retail. 68 (3), 316.
Helsen, K., Schmittlein, D.C., 1993. Analyzing duration times in marketing: evidence for the effectiveness of hazard rate models. Mark. Sci. 12 (4), 395–414.
Jain, D.C., Vilcassim, N.J., 1991. Investigating household purchase timing decisions: a conditional hazard function approach. Mark. Sci. 10 (1), 1–23.
Janiszewski, C., Cunha Jr, M., 2004. The influence of price discount framing on the evaluation of a product bundle. J. Consum. Res. 30 (4), 534–546.
Kahn, B.E., Schmittlein, D.C., 1989. Shopping trip behavior: an empirical investigation. Mark. Lett. 1 (1), 55–69.
Kamakura, W.A., Kang, W., 2007. Chain-wide and store-level analysis for cross-category management. J. Retail. 83 (2), 159–170.
Kamakura, W.A., Russell, G., 1989. A probabilistic choice model for market segmentation and elasticity structure. J. Mark. Res. 26 (4), 379–390.
Kim, B.D., Park, K., 1997. Studying patterns of consumer's grocery shopping trip. J. Retail. 73 (4), 501–517.
Koch, S., 2013. A Theory of Grocery Shopping: Food, Choice and Conflict. Bloomsbury Publishing Plc., London, UK.
MacKay, D.B., 1973. A spectral analysis of the frequency of supermarket visits. J. Mark. Res. 10 (1), 84–90.
Mathias, K.C., Slining, M.M., Popkin, B.M., 2013. Foods and beverages associated with higher intake of sugar-sweetened beverages. Am. J. Prev. Med. 44 (4), 351–357.
McGoldrick, P.J., Marks, H.J., 1987. Shoppers' awareness of retail grocery prices. Eur. J. Mark. 21 (3), 63–76.
Melis, K., Campo, K., Breugelmans, E., Lamey, L., 2015. The impact of the multi-channel retail mix on online store choice: does online experience matter? J. Retail. 91 (2), 272–288.
Neslin, S.A., Henderson, C., Quelch, J., 1985. Consumer promotions and the acceleration of product purchases. Mark. Sci. 4 (2), 147–165.
Nielsen, 2015. The future of grocery e-commerce: digital technology and changing shopping preferences around the world. Available at: ⟨https://www.nielsen.com/content/dam/nielsenglobal/vn/docs/Reports/2015/Nielsen%20Global%20E-Commerce%20and%20The%20New%20Retail%20Report%20APRIL%202015%20(Digital).pdf⟩ (Accessed 30 May 2018).
Nordfalt, J., 2009. Unplanned grocery purchases: the influence of the shopping-trip type

revisited. J. Consum. Behav. 8 (1), 1–13.
Oehlert, G.W., 1992. A note on the delta method. Am. Stat. 46 (1), 27–29.
Paeratakul, S., Ferdinand, D.P., Champagne, C.M., Ryan, D.H., Bray, G.A., 2003. Fast-food consumption among US adults and children: dietary and nutrient intake profile. J. Am. Diet. Assoc. 103 (10), 1332–1338.
Park, C.W., Iyer, E.S., Smith, D.C., 1989. The effects of situational factors on in-store grocery shopping behavior: the role of store environment and time available for shopping. J. Consum. Res. 15 (4), 422–433.
Perakis, G., 2017. A revolutionary model to optimize promotion pricing. Huffington Post. Available at: 〈https://www.huffingtonpost.com/entry/a-revolutionary-model-to-optimize-promotion-pricing_us_579638bee4b0e002a313c6da〉 (Accessed 27 September 2018).
Putrevu, S., Lord, K.R., 2001. Search dimensions, patterns and segment profiles of grocery shoppers. J. Retail. Consum. Serv. 8 (3), 127–137.
Raijas, A., 2002. The consumer benefits and problems in the electronic grocery store. J. Retail. Consum. Serv. 9 (2), 107–113.
Rossi, P., 2014. Bayesian Non-and Semi-parametric Methods and Applications. Princeton University Press.
Rossi, P.E., Allenby, G.M., McCulloch, R., 2012. Bayesian Statistics and Marketing. John Wiley & Sons.
Roggeveen, A.L., Nordfält, J., Grewal, D., 2016. Do digital displays enhance sales? Role of retail format and message content. J. Retail. 92 (1), 122–131.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464.
Simonin, B.L., Ruth, J.A., 1995. Bundling as a strategy for new product introduction: effects on consumers' reservation prices for the bundle, the new product, and its tie-in. J. Bus. Res. 33 (3), 219–230.
Slater, S.F., Olson, E.M., 2002. A fresh look at industry and market analysis. Bus. Horiz. 45 (1), 15–22.
Song, I., Chintagunta, P.K., 2006. Measuring cross-category price effects with aggregate store data. Manag. Sci. 52 (10), 1594–1609.
Statista, 2014. U.S. food shopping behavior - statistics & facts. Available at: 〈https://www.statista.com/topics/1527/food-shopping-behavior/〉 (Accessed 30 May 2018).
Teller, C., Gittenberger, E., Schnedlitz, P., 2013. Cognitive age and grocery-store patronage by elderly shoppers. J. Mark. Manag. 29 (3–4), 317–337.
U.S. Census Bureau, 2017. Advance monthly sales for retail and food services, November 2017. Available at: 〈https://www.census.gov/retail/marts/www/marts_current.pdf〉 (Accessed 30 May 2018).
Vartanian, L.R., Schwartz, M.B., Brownell, K.D., 2007. Effects of soft drink consumption on nutrition and health: a systematic review and meta-analysis. Am. J. Public Health 97 (4), 667–675.
Yavas, U., Tuncalp, S., 1984. Perceived risk in grocery outlet selection: a case study in Saudi Arabia. Eur. J. Mark. 18 (3), 13–25.
Yeon Kim, H., Chung, J.E., 2011. Consumer purchase intention for organic personal care products. J. Consum. Mark. 28 (1), 40–47.