Store Choice and Consumer Behavior in Food Deserts: An Empirical Application of the Distance Metric Method

Lauren Chenarides, Pennsylvania State University

Edward C. Jaenicke, Pennsylvania State University

Invited paper presented at the 2017 ASSA Annual Meeting, January 6-8, 2017, Chicago, Illinois.

Copyright 2016 by Lauren Chenarides and Edward C. Jaenicke*. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.*

Store Choice and Consumer Behavior in Food Deserts: An Empirical Application of the Distance Metric Method

Lauren Chenarides*and Edward C. Jaenicke†

Department of Agricultural Economics, Sociology, and Education,
Penn State University,
University Park, PA 16802

November 2016

DO NOT CITE WITHOUT AUTHORS' PERMISSION

Disclaimer: Any opinions, findings, recommendations, or conclusions are those of the authors and do not necessarily reflect the views of the Economic Research Service, U.S. Department of Agriculture. The analysis, findings, and conclusions expressed in this paper also should not be attributed to either Nielsen or Information Resources, Inc. (IRI).

Acknowledgement: This research is funded by the USDA-NIFA-AFRI Predoctoral Fellowship Grant Program number 2015- 67011-22790.

*Electronic address: lec201@psu.edu; Corresponding author
†Electronic address: ecj3@psu.edu

Abstract

While food access is an increasingly studied component of research related to diet and health, consumer behavior and store choice have been relatively overlooked in understanding the dietary health-food access relationship. Especially in areas with high poverty rates, where the proportion of low-access and low-income population persists over time, consumers are faced with shopping at non-traditional stores, which may augment the negative welfare impacts of residing in these areas. Using IRI's Consumer Network Panel, IRI's InfoScan, and Nielsen's TDLinx store characteristics data, this paper develops a structural model of store choice that frames Pinkse, Slade, and Brett's (2002) distance metric (DM) method inside a demand system to model what behaviors drive consumers' store choice decisions, highlighting underserved communities. While the DM method has been used previously to model brand choice, this paper is the first to use it to investigate store choice. Because the store-choice model is based on demand for store attributes (such as relative prices, product assortment measures, store services, and distance between stores), it reveals consumer preferences for store types and provides insight into policy prescriptions that attempt to improve food access.

1 Introduction

In a 2012 report to Congress, the USDA highlights supermarket availability as a key indicator of household food security, directly linking food access to consumer welfare outcomes by implying that consumers who have access to supermarkets will be better able to meet the dietary needs of their household (Ver Ploeg et al. 2012). Especially for households residing in underserved communities, the importance of supermarket access becomes apparent when consumers are limited to shopping at non-traditional stores that offer a limited variety of products or at stores where healthy foods are less likely to be available (Fitzpatrick & Ver Ploeg 2013, Handbury et al. 2015, USDA 2009).

Despite the prominence of research related to supermarket access, few studies test how consumer behavior would adapt given a change in the food retailing landscape, such as the entrance of a new store, that would remedy extreme low access to supermarkets. For example, Cummins et al. (2014) investigate a pilot-study initiative, namely the Pennsylvania Fresh Food Financing Initiative (PFFFI), by evaluating the impacts of opening a new supermarket in Philadelphia. The study finds that, although there is increased access, shoppers do not markedly change the amount of fruits and vegetables consumed. This finding supports an earlier national-level study that indicates that the density of supermarkets in urban areas does not have a significant effect on household fruit and vegetable consumption (Kyureghian et al. 2013). In addition, a recent report by Rahkovsky & Snyder (2015) uses micro-level scanner data to investigate the correlation between at-home food purchases and residency in food deserts. The authors find that while the diet quality of low-access, low-income consumers is measurably different from their counterparts, limited supermarket access and differences in relative prices does not explain the differences in diet. Each of these reports in this emerging line of the literature suggests that accessibility, or proximity, alone is not necessarily a solution to addressing concerns about diet and nutrition, and therefore implies that consumer behavior plays a larger role in appropriately addressing policy-related issues around food access.

While the issue of food access is an important component when focusing on diet and health, the question of *where* consumers choose to shop is perhaps just as important as *what* foods consumers choose to purchase. Recent evidence suggests that regardless of whether households live in food deserts, they will travel outside of their neighborhoods to purchase food items deemed as more healthful, as long as they have the means by which to get there (Ver Ploeg et al. 2015). While the present food retailing landscape indeed plays a major role in determining where households choose to shop, other indicators such as demographics, prices, proximity, characteristics of stores, and underlying preferences also contribute to these decisions. In this paper, by developing and estimating a store-choice model based on store attributes and household characteristics, our goal is to better understand consumer behavior, specifically focusing on the Philadelphia metro-area. Through this research, we address the questions of where and how households shop, and our results will provide insight into the "null" result that Cummins et al. (2014) and others have uncovered.

This paper makes two major contributions to the literature on store choice and consumer behavior. First, we offer an alternative, and theoretically sound, economic model of store choice by modeling the household demand for store attributes. By extending the Distance Metric (DM) demand model of Pinkse et al. (2002) to the problem of consumer store choice, this model can accommodate certain behaviors as drivers of consumers' store-choice decisions, highlighting underserved communities. This technique has been used when modeling product demand by looking at the distance of one product's attributes from another as a way of determining price competition, but our paper is the first to apply the DM model to store choice. Because our store-choice model is based on demand for store attributes such as relative prices, product assortment measures, store services, and physical geographical distance traveled to a store, it reveals consumer preferences on store types and store attributes, and provides insight into policy prescriptions that attempt to improve food access. An important benefit of a store-choice model using the DM method is that it allows multiple store trips, which is widely observed in consumer-level scanner data but not easily incorporated into store-choice

models based on discrete choice methods.

A second contribution of this research is that this analysis integrates several rich data sets, in particular IRI's Consumer Network Panel (household-level scanner data) and InfoScan (store-level scanner data), and TDLinx Store Characteristics data, into the store-choice model. Broadly speaking, the use of these data sources supports a more complete picture of the food retailing environment with extensions and applications in the marketing, health, and policy sectors. When applied to store choice, using these data sources enables two types of store attributes to be constructed and incorporated into the model: namely, store attributes that resemble location and store type, as well as information that reflects attributes observed inside the store. By modeling store choice and focusing on specific store-level characteristics, this research contributes to the growing body of literature that seeks to make creative use of the richness of the micro-level scanner data in modeling consumer behavior.

To preview the results of this paper, the store-choice model generates a number of conclusions about shopping preferences. In general, store attributes on type and location are shown to play an important role in household store switching behavior. In particular, when the price of a store increases, households are more likely to substitute to nearby stores and stores in the same channel. Except for supermarkets, households seem very loyal to their store type (namely, convenience stores, supercenters, and some dollar stores). Moreover, attributes that reflect characteristics inside the store provide additional insights into shopping preferences. These results show that households are less price-sensitive when it comes to the size of the store, which could indicate that one-stop shopping is valued by Philadelphia households. Measures of a store's uniqueness suggest that the relative availability of the products on store shelves is also highly valued. On the other hand, households are more price sensitive at supercenters. Finally, consumers are more likely to switch to a store with a very different product assortment level. This finding could be evidence of the importance of primary and secondary stores that serve complementary roles. These results imply that considerations not only on the right type of store be made when incentivizing new store

entrance to underserved communities, but also on the role that location plays in determining the viability of a new store entrant.

The remainder of this paper is organized as follows. In the following section, we summarize relevant research on the food retailing environment and store choice. Following this overview, we discuss the methodology and extension of the DM model to store choice. Next, we provide detailed descriptive information about the market area and its food retailing environment on which this case study is focused. Finally, we conclude with an overview of results and potential policy implications.

2 Related Literature

The significance of this paper directly relates to efforts made by policymakers who focus on food access generally and more specifically in areas deemed as food deserts. Given the sustained rates of food insecurity in the U.S., research addressing ways in which to mitigate negative welfare outcomes for households living in these communities continues to grow. Therefore, in order to fully motivate this paper, we rely on three major streams of literature that frame this research: (*i*) the role inadequate access plays in shaping household shopping behavior, (*ii*) the nature of the food retailing environment, and (*iii*) current methods for modeling store choice.

2.1 The Significance of Poor Food Access

Over the course of the four years from 2006 to 2010, few changes to food access (i.e., the opening of new supermarkets) have been seen (Ver Ploeg et al. 2012). Especially in urban food deserts, where small-scale stores may face lower entry costs due to their smaller size, the limited selling space also means smaller product assortment of fresh fruits and vegetables (Handbury et al. 2015, USDA 2009). Numerous studies indicate that supermarkets are less prevalent in poorer areas, while fast-food restaurants appear in more concentrated numbers

(Alwitt & Donley 1997, Moore & Diez Roux 2006). The combination of a high density of fast-food stores and the migration of supermarkets to suburban areas may in fact contribute to the disparities in choices among households living in underserved communities and may ultimately compound the hardships faced by these households.

Low-income households, whose presence is more concentrated in rural and urban regions, are faced with shopping at smaller food stores where food prices tend to be higher. Research shows that income has a statistically significant positive effect on fruit and vegetable purchases as well as average store size (Dunkley et al. 2004, Kyureghian & Nayga 2013). The ability to access larger food stores requires higher transportation costs, which present hardships for households living in poor access areas who do not have the transportation means to drive to the nearest supermarket.[1] Faced with higher search costs, these households are often unable to take advantage of the benefits of shopping at larger format stores, such as supermarkets and discount merchandisers, which tend to locate in the suburbs or higher-income areas (Leibtag & Kaufman 2003). Better-off households residing in low-income areas are more likely to own a car, so traveling to a supermarket outside of their immediate neighborhoods allows them to escape the food desert in which they live (USDA 2009).

From the researcher's perspective, food access issues are a multidimensional problem and these studies indicate that other forces may exist within underserved communities that are preventing households from incorporating higher-quality products into their market baskets. Rather than qualifying this issue as one dictated by a *lack* of access, researchers such as Lee (2012) have suggested that it might be an issue of *ease* of access. Yet recent evidence from Ver Ploeg et al. (2015) finds that for low-income, low-access households who are able to travel outside of their neighborhoods for groceries, access only has a small impact on diet quality. Handbury et al. (2015) uncover two findings that unveil some of the underlying behavioral differences among lower income and lower educated households: First, they observe that

[1]In general, big box stores, such as Wal-Mart, are located outside of urban areas, whereas concentrations of low-income and low-access households are highest in urban areas, thus making these transportation costs particularly significant for the affected households (Ellickson & Grieco 2013, Holmes 2011).

households with lower income and education purchase less healthful foods. Second, they further find that the nutritional quality of purchases made by households with low levels of income and education respond very little when new stores enter or when existing stores change their product offerings. Together, their results indicate that policies aimed at improving access to healthy foods in underserved areas will leave most of the socioeconomic disparities in nutritional consumption intact.

2.2 The Food Retailing Environment

Over the past twenty years, the introduction of new food retailer formats, such as supercenters and club stores, has significantly changed the landscape of the food retailing industry. During this time, independent grocery stores began competing for market power with larger merchandisers and chain supermarkets. For example, Wal-Mart supercenters have had one of the fastest growing grocery departments. In 1999, Wal-Mart ranked number five in total U.S. grocery sales but became the top grocery retailer in the U.S. and Canada as of 2011 (Kaufman 2007, Zwiebach 2014).

A number of research studies in the industrial organization literature examines the impacts associated with the introduction of new format stores. In order to stay competitive, firms have to differentiate themselves by creating strategic advantages over their competitors, either through marketing techniques (e.g., pricing strategies) or the control of market channels (e.g., distribution networks) (Clarke 2000). Major changes in the food retailing landscape have been inspired by the idea of "one-stop" shopping. The ability for consumers to shop at a single store to make all of their purchases has been a great success for retailers, supporting the idea that firms with the widest selection prevail (Ellickson 2007). The larger the food retailer is, the easier it becomes for them to spread their fixed costs across a wider assortment of products (Leszczyc et al. 2004). Not only is it efficient for the retailer, but it is also efficient for consumers. Although consumers forgo additional services for convenience, households are able to mitigate the fixed cost of shopping by only frequenting a single location.

As the food retailing industry continues to become more efficient for both retailers and consumers, research has emerged that investigates factors leading to food landscape outcomes, including food deserts in the extreme. Economic theory suggests that variations of fixed and variable costs among types of food retailers significantly affect equilibrium outcomes (Ellickson 2007). The players within the food retailing industry face high fixed costs and a heterogeneous consumer base, so in order to justify entering a market, food retailers must be sure that they can gain a competitive foothold. In growing markets, where the expanse of consumer preferences has an impact on the degree of vertical product differentiation, research indicates that markets will not respond by new firms entering the market to "fill in" the product assortment needs of the consumer base. Instead, existing firms will feel pressure to improve the variety of their products, in effect raising fixed costs and creating barriers for other firms to enter (Shaked & Sutton 1987, Sutton 1991, Ellickson 2007).

This line of current economic research that investigates these behaviors shows that the food retailing landscape is the equilibrium outcome of supply and demand factors (Shaked & Sutton 1987, Sutton 1991, Ellickson 2007, Bonanno et al. 2012, Ellickson & Grieco 2013). Extreme food unavailability in a localized market, i.e., a food desert, is one such equilibrium outcome. In markets where access is limited, retailers may not have an incentive to overcome such high fixed costs and therefore choose to locate in markets with more stable demand. If the demand potential is low, retailers may not be willing to participate in certain markets. In these equilibrium models, in particular those which highlight food accessibility, the most significant indicators of food deserts and other landscape outcomes is the uneven dispersion of consumer types (Ellickson 2007, Bonanno et al. 2012).

Many policy initiatives, such as the PFFFI, are aimed at attracting supermarkets to underserved rural and urban neighborhoods, in particular food deserts. These policies work by providing new entrants a lump-sum subsidy that serves as a means for covering the fixed costs associated with market entry. However, recent studies consistently document that the addition of a store does not cause a direct change in shopping behavior, i.e., the so-called

"null" effect mentioned above (Cummins et al. 2014, Kyureghian & Nayga 2013, Kyureghian et al. 2013). In fact, these studies have begun to suggest that the lack of sustainable consumer response may be less about the store itself and, rather, more about what characteristics the store encompasses, such as the prices, promotional frequency, assortment, and other attributes the store offers (Ver Ploeg & Rahkovsky 2016). Given a better understanding about what store features consumers value, research, such as this paper, may be able to uncover the bundle of attributes representing the right type of store that would flourish in a food desert.

2.3 Household Store Choice

The line of literature on food store choice is extensive. Much of this literature describes consumer store choice as a discrete process, whereby households make a list or know what items they want to purchase, and then choose the store that best fits their needs. Focusing on the contemporary models of consumer store choice, this research can be summarized by how each model integrates information about the fixed and variable costs of shopping (Bell et al. 1998, Bell & Lattin 1998, Briesch et al. 2004, Hoch et al. 1994, Sinha 2000). Fixed costs represent the inherent costs associated with visiting a store, such as amenities or assortment levels, as well as travel costs. On the other hand, variable costs are those that represent the costs that might be incurred during the shopping occasion. For example, using household-level scanner panel data, Bell et al. (1998) examine how certain factors (e.g., travel distance) affect the fixed costs of shopping, and how a store's pricing format (EDLP versus Hi-Lo) affects the variable costs of shopping. They find that shoppers with bigger market baskets prefer lower variable costs, reflected by the types of amenities, and higher fixed costs, reflected by certain store types or by a larger selling space, because fixed costs can be spread over more items. This behavior suggests that the likelihood of choosing between two stores that are the same distance from a household is not equal. Briesch et al. (2004) consider an alternative decision making sequence where the ultimate store choice depends on the store

that offers the highest utility. Rather than focus on pricing format, the authors estimate a model of grocery store choice with assortment, convenience, price, and feature advertising as predictors. Preferences for lower prices and shorter travel distances are consistent across all consumers, yet they find that unobserved heterogeneity is greatest for assortment.

One relevant feature of these studies is that they all rely on a discrete choice framework as the theoretical underpinning for explaining how households make the decision on where to shop. The major benefit of using a discrete choice method is its ability to deal with the dimensionality issues arising from many store choices by projecting the characteristics of the stores themselves onto attribute space (McFadden 1973). However, these attribute-based modeling approaches are limited by the independence of irrelevant alternatives (IIA) assumption, which suggests that substitution between two stores is proportional to the shares, rather than a function of attribute proximity (Pofahl & Richards 2009). In other words, if a consumer is choosing between two stores, the odds of choosing one store over the other is not affected by the presence of a third store, regardless of how similar the third store is to either of the first two. In some cases, this property may be valid; however, in the case of consumer demand for stores, the IIA assumption may not be realistic.

Taylor & Villas-Boas (2016) develop an empirical household store-choice model using the multinomial mixed logit. The multinomial mixed logit relaxes the IIA assumption, while simultaneously allows for preferences to be a function of store characteristics, specifically outlet type and distance between a household's home and the outlet choice. In addition, they are able to incorporate household heterogeneity through the use of household demographics. Their approach yields more realistic substitution patterns between store types and provides a framework for measuring willingness to pay for outlet attributes.

Nonetheless, while this framework has many attractive benefits, households are ultimately restricted to make a decision between two alternative stores. This restriction may be mostly unimportant in the product-choice world where households may only purchase one product within a certain category of food items. However, in the case of store choice, households

spend a non-trivial portion of their budget at multiple stores and therefore it does not seem realistic to make this restriction a priori.[2] Although the time period during which the decision occurs can be shortened, say to one day or one week intervals, the inability for discrete choice models to account for multiple trips is a strong limitation and ultimately motivates the decision to use an expenditure share-based demand model in this paper. In addition, the empirical model used in this paper centers on estimating demand for store attributes, thereby preserving similar interpretations as the multinomial mixed logit. We describe this model and its properties in the following section.

3 An Application of the Distance Metric Method to Store Choice

We adopt a linear approximation of the Almost Ideal Demand System (LA/AIDS) of Deaton & Muellbauer (1980) and incorporate the DM method of Pinkse, Slade, and Brett (hereafter "PSB") (2002) into this framework. A number of studies (Rojas & Peterson 2008, Rojas 2008, Pofahl & Richards 2009, Bonanno 2012) have applied DM method to brand-choice models, but this study is the first to apply it to a consumer store-choice model. The extension from brand choice to store choice is relatively straightforward: instead of the product brands, food retail stores (or store chains) comprise the consumer's choice set; instead of product attributes, store attributes help differentiate the stores in the choice set; and instead of each brand's expenditure share, each store's expenditure share will ultimately be the dependent variable in the store-choice model. As briefly noted above, a store-choice model using the DM method will retain the strengths of discrete choice models, namely the focus on attributes, while also allowing multiple store trips in a time period.

Let i denote the household, j denote the store within the choice set, and t denote the

[2]Smith (2004) acknowledges this behavior in shopping frequency by calling attention to households' primary shopping trips and secondary shopping trips.

month. Therefore, the expenditure share function for a household i shopping at store j in month t would resemble:

$$w_{ijt} = a_{ij} + \sum\nolimits_j \gamma_{jk} \log p_{kt} + \beta_{ij} \log \left\{ x_{it} / P_{it}^L \right\} + \epsilon_{ijt} \tag{3.1}$$

where $w_{ijt} = q_{ijt} p_{ijt} / x_{it}$ represents the expenditure share for household i's total food purchases at store j in month t; p_{kt} represents the store-level price index of store j in month t; and x_{it} represents the total food expenditure by household i in month t. To linearize the price index term P_{it}^L, Moschini (1995) proposed to approximate this term with a log-linear analog of the Laspeyres index such that $\log P_{it}$ is replaced by $\log P_{it}^L \equiv \sum_{j=1}^{J} w_{ij}^0 \log P_{jt}$ and w_{ijt}^0 is store j's base share for household i with $w_{ij}^0 \equiv T^{-1} \sum_{t=1}^{T} w_{ijt}$ where $t \in (t, ..., T)$ represents the month. The base share of store j for household i represents a yearly average of household i's purchase shares at store j. The parameters a_{ij}, γ_{jk}, and β_{ij} are to be estimated and e_{ijt} is an error term. Each store share equation represents the share of total food expenditure a household allocates to the stores within their choice set.

Rather than estimate a demand system of $J - 1$ equations and $J(J-1)/2$ cross-price parameters, the cross-price coefficients γ_{jk} are specified as a function of distance measures between stores j and k (PSB, 2002). These distances (δ_{jk}) are spatial measures, but they could be measured in terms of physical space or store-attribute space, such that $\gamma_{jk} = g(\delta_{jk})$. This specification indicates that the level of substitutability depends on the "closeness" of attributes between store j and store k. Store attributes may be discrete (δ_{jk}^d) or continuous (δ_{jk}^c). Stores can be neighbor measures in attribute space if discrete attributes are identical (for example, if both stores are the same type) or if continuous attributes are close in level measures (for example, if two stores sell a similar number unique UPCs). The closer the two stores are in observable characteristics, the more likely they will be considered substitutes for one another. Conversely, the farther apart in attribute space, the less likely the two stores

will be considered substitutes (Bonanno 2012).

The local measure of closeness, δ^d_{jk}, as

$$\delta^d_{jk} = \begin{cases} 0 & if \quad |z^d_j - z^d_k| = 0 \\ 1 & if \quad |z^d_j - z^d_k| = 1 \end{cases} \tag{3.2}$$

and, using the same function of Euclidean distance as the aforementioned literature,

$$\delta^c_{jk} = \frac{1}{1 + 2\sqrt{\sum_l (z^c_j - z^c_k)^2}} \tag{3.3}$$

where z^d_j is a discrete attribute of store j and z^c_j is a continuous attribute of store j.

Using the distance measures δ^d_{jk} and δ^c_{jk}, the cross-price parameter can be written as

$$\gamma_{jk} \log p_{kt} = \sum_{d=1}^{D} \left(\lambda^d_j \sum_{k \neq j}^{J} \delta^d_{jk} \log p_{kt} \right) + \sum_{c=1}^{C} \left(\lambda^c_j \sum_{k \neq j}^{J} \delta^c_{jk} \log p_{kt} \right) \tag{3.4}$$

In addition, the constant term a_{ij}, the own-price coefficient γ_{jj}, and the coefficient on the price index β_{ij} may be written as functions of household i's demographics and store j's attributes, such that

$$a_{ij} = a_0 + \sum_{l=1}^{L} a_l z^a_{jl} + \sum_{h=1}^{H} \phi_h h_{ih} \tag{3.5}$$

$$\gamma_{jj} = \gamma_0 + \sum_{m=1}^{M} \gamma_m z^\gamma_{jm} \tag{3.6}$$

$$\beta_{ij} = \beta_0 + \sum_{n=1}^{N} \beta_n z^\beta_{jn} \tag{3.7}$$

where store j's characteristics are represented by z^a_{jl}, z^γ_{jm}, and z^β_{jn}, and household i's characteristics are represented by h_{ih}. By construction, the characteristics z^θ, where $\theta \in (a, \gamma, \beta)$, and h_{ih} are each be represented separately. Therefore, imposing 3.5, 3.6, and 3.7 into equation

3.1, the specification of the DM-LA/AIDS model is as follows:

$$
\begin{aligned}
w_{ijt} = a_0 &+ \sum_{l=1}^{L} a_l z_{jl}^{a} + \sum_{h=1}^{H} \phi_h h_{ih} + \left(\gamma_0 + \sum_{m=1}^{M} \gamma_m z_{jm}^{\gamma}\right) \log p_{jt} \\
&+ \sum_{d=1}^{D} \left(\lambda_j^d \sum_{k \neq j}^{J} \delta_{jk}^{d} \log p_{kt}\right) + \sum_{c=1}^{C} \left(\lambda_j^c \sum_{k \neq j}^{J} \delta_{jk}^{c} \log p_{kt}\right) \\
&+ \left(\beta_0 + \sum_{n=1}^{N} \beta_n z_{jn}^{\beta}\right) \log \frac{x_{it}}{P_{it}^{L}} + \epsilon_{ijt}
\end{aligned}
\tag{3.8}
$$

Equation 3.8 above shows that store attributes, the z's, can enter the model in several ways: as intercept shifters, through own-price interactions, through interactions with the cross-price term, or through the expenditure term. Drawing from a broad set of store characteristics, substitutability between stores is modeled as a function of the relative distance between the stores along several attribute-space dimensions, such as attributes that reflect store type and distance between stores as well as attributes that reflect what is inside a store (e.g., assortment, service levels, and relative prices). In equation 3.8, the sign and significance of λ_j^d and λ_j^c characterize consumer's store-switching behavior given a change in price. If a coefficient estimate is positive and significant, then this implies that consumers respond to an increase in price at store j by switching to another store k with the same or similar attributes. In other words, store switching would not involve much distance in attribute space.

Lastly, because the parameters in equation 3.8 reflect attributes levels or relative levels rather than brands, the J shares can be stacked so that all parameters can be recovered from a single estimation. Because consumers do not generally shop at all stores in the choice set in one time period, many shares may equal zero for a particular consumer a single estimation; therefore, this method makes it computationally easy to handle a censored version of equation 3.8.

4 Data and Variables

4.1 Consumer Network Panel and Choice Set

The application of this store-choice model to observable market data relies on the Information Resources, Inc. (IRI) household-level store scanner dataset, called the Consumer Network Panel (CNP), for the year 2012.[3] Spanning the U.S., these data document food purchase transactions at food retail outlets by households selected to compose a representative sample. The IRI data geographically link each household to a census tract, and each shopping trip by a household is linked to a retail chain. It is important to note, however, that the specific location of these retail outlets is not known; only the identity of retail chain is known (Sweitzer et al. 2016). For example, if a household shops at Supermarket A in census tract m and records purchases for that store, only the Retail Chain A is observed, of which Supermarket A is perhaps but one location. In this paper, "stores" refer to different retail chains, rather than to unique store locations. Additional data from the 2012 TDLinx Store Characteristics data provides more specific geographical data on store locations for the retail chains.

The store-choice model application uses the Philadelphia metro-area as a case study.[4] Rather than focus solely on Philadelphia County (FIPS code 42101), which is irregularly shaped, we expand the food retailing market area around Philadelphia County slightly to include parts of Bucks (42017), Delaware (42045), and Montgomery (42091) counties by creating a geographical convex hull. Creating the convex hull minimizes the likelihood that two stores, one inside the market area and one outside the market area, share the same set of customers.[5] More importantly, the convex hull identifies our study area and the households

[3]The year 2012 is chosen due to the fact that necessary household demographic information can be linked with the purchase data for this year. Demographic information for earlier years is not available in IRI data obtained in collaboration with USDA. For more information, see Sweitzer et al. (2016). Analysis for additional years following 2012 will be extended for future research.

[4]The motivation behind choosing Philadelphia is so the results can provide context for the Cummins et al. (2014) "null" result.

[5]This may not be a realistic assumption for a household living in a border census tract and might be willing to shop at a store inside the convex hull or outside the convex hull with equal probability and is therefore a limitation of this analysis.

and stores inside that area. Using the TDLinx information, we are able to match 26 food retail chains, encompassing 242 unique store locations, in the Philadelphia metro-area with the retail chains where study-area households report food transactions in the CNP.[6]

The final choice set of retailers in the Philadelphia study area is comprised of 21 food retail chains. These 21 stores capture 68% of sales reported in the CNP transactions data and 55% of average weekly volume reported in TDLinx. The criteria for selecting these stores are based on whether (1) the CNP-retail chain appears in the TDLinx panel, (2) the total expenditure reported at that food retail chain is substantially greater than zero, and (3) monthly store price information for that store is complete (see section 5.2 for details about store price information). The final choice set is made of up twelve grocery stores, two convenience stores, three mass merchandisers, one supercenter, and three dollar stores.

Household-level store expenditure shares, w_{ijt}, are constructed as follows. The numerator ($q_{ijt}p_{ijt}$) represents the total food expenditure by household i at store j in month t. The denominator (x_{it}), or base expenditure, represents the total expenditure a household spends across all stores in the choice set and excludes any food purchases that households made at stores outside of the choice set.[7]

According to the USDA's Food Access Research Atlas, $\sim 34\%$ of the households residing in Philadelphia County live in a census tract deemed as low-income and low-access (at a half mile from the nearest supermarket). A closer look at the shopping frequency patterns of households in the Philadelphia study area during the year 2012 shows the following statistics (n=267 households): First, the average number of trips to any food store made in a month

[6]Only the following channels are considered as viable food retail outlets: Grocery, Convenience, Mass Merchandisers, Supercenters, and Dollar Stores. Using ArcMap version 10.4, we plot the coordinates of the TDLinx stores in Philadelphia, Bucks, Delaware, and Montgomery counties, and overlay the Philadelphia metro-area convex hull, which is presented in figure 1. Using the 'Clip' Geoprocessing tool, we are able to isolate the stores that fall within the market area and this is our basis for constructing the retailer-level store attributes. Of the 26 stores that match, 22 are in the top 27 revenue-generating food retail chains, excluding club stores, reported in the CNP.

[7]Using this method, the sum of the shares equals one, and does not allow for an outside option. Alternative expenditure shares allowing for an outside option have also been calculated; however, store attributes, such as price, would be required to generate the store-level characteristics for all other stores (i.e., the outside option). Therefore, the results from our analysis are specific to this case study area with these set of 21 stores.

is just under eight, while the average number of unique food retail chains visited within a month is over four. Roughly 70% of households, on average, visited more than three unique chains in a given month. Excluding stores outside of the choice set, the primary store receives approximately, on average, 54% of the household's monthly food expenditure, followed by a second and third store totaling close to 22%. Subsequent stores account for the remaining 24%.

Table 1 summarizes select household demographic characteristics and expenditure shares for the 267 households in the Philadelphia County. In general, the average household size (*HHsize*) for the sample is just over two, with a median annual income (*MedInc*) of $56,000. The sample is older, with the average age of the household head at sixty years old (*AgeHead*) and roughly 57% of the household race is identified as white, while 37% of households identify as black. A majority of households report that there is a female head (76%), and 63% of households live in a census tract deemed as low access at half a mile. Only ten percent of the households report having achieved a college degree (*College*).

4.2 Store Characteristics

Aside from price, additional store characteristics are incorporated into the model by using both the IRI data on transactions and the 2012 TDLinx Store Characteristics data. As mentioned above, information about the fixed and variable costs of shopping have traditionally been used to motivate the selection of store characteristics in a store-choice model. Similar to this framework, the approach taken in this paper categorizes store attributes according to two distinct classes. The first class are those attributes that motivate a household to choose the store based on type and location. These include the channel type (namely, supermarket, dollar store, convenience store, supercenter, or mass merchandiser) and geographical distance measures (distance between stores and distance between household location and store). The second class of attributes reflects those attributes inside the store that might influence a household's store choice decision, including factors that might influence how much of their

budget they allocate to each store. These characteristics include price, product assortment levels, variety of product offerings, square footage, and amenities. Each of these measures motivates a household to favor one store over another and ultimately plays a role in influencing store choice (Bonanno & Lopez 2009, Smith 2004, Taylor & Villas-Boas 2016). Table 2 summarizes a list of store attributes from the available dataset.

The attributes that represent type and location are channel type and physical distance between stores, as well as physical distance between household census tract and store. Channel type indicators are *Grocery*, *Convenience*, *MassMerchandiser*, *Supercenter*, and *DollarStore*. The channel type enters into the model as a discrete distance metric measure (*DM_Channel*) and indicates how likely households are to switch to a different store type given a change in price.[8] If households are more loyal to a certain channel type as price changes, then the the coefficient on this variable will be positive; however, if households are not loyal to channel type as price changes, then the coefficient on *DM_Channel* will be negative. In addition, we interact *Supercenter* with own-price to generate *LNPxSupercenter*. This variable represents the price sensitivity of households are shop specifically at supercenters. Traditionally, the type of consumer who frequents a supercenter is more price elastic, so it is expected that the coefficient on *LNPxSupercenter* will be negative (and significant).

In addition to channel type, we use the coordinates of each unique store from TDLinx to calculate the average physical distance between store j and all other stores in the choice set (*StoretoStoreDist*), such that $j \neq k$. *StoretoStoreDist* enters the model as a continuous distance measure (*DM_SDIST*). We use the inverse of *StoretoStoreDist* so store distance is translated to physical "closeness." It is expected that if two stores, j and k, are close in physical space, a household will substitute to a store k closer in proximity to their original store j, given a chance in price j. Put another way, travel distance between stores is a fixed cost, so in order to minimize these fixed costs, households are perhaps more likely to switch

[8]For continuous distance metric variables, the inverse Euclidean distance measures are computed and stored in "weighting" matrices, where the element in each matrix corresponds to the "closeness" between two stores' characteristics. For discrete distance metric variables, a '0' indicates that the two are of different channel types, for example, or '1' if they are both the same.

between stores located closer in physical space.

Likewise, the average distance between a household's census tract and store is also calculated. *HHtoStoreDist* is measured by taking the sum of the miles between each census tract centroid and each unique store coordinate[9] in TDLinx, and then dividing by the number of stores within that retail chain to create census tract-store distance pairs. This "farness" measure represents the average distance a household living in census tract m travels to get to any store within a given retail chain. *HHtoStoreDist* is interacted with the expenditure term and enters into the model as *LNEXPxHHDIST*. The sign on estimated coefficients for expenditure-attribute interactions indicate the expenditure sensitivity to those attributes. It is expected that if the sign on *LNEXPxHHDIST* is negative and significant, that households will be more expenditure sensitive the further they have to travel from their home to visit a store.

To capture the diversity of attributes that reflects characteristics inside the store, such as breadth and the variety of product offerings available at each retailer, we construct several additional measures that may be new to store-choice models. *Breadth* reflects the number of unique UPCs carried in store j, and is a straightforward measure of product offerings. To scale this variable, *Breadth* is divided by the value of assortment from the store with the highest number of UPCs ($\max(Breadth)$). This measure *PctMaxAssort*, which takes a value from zero to one, can be easily compared between stores in the choice set. This variable enters the model as a continuous distance measure (*DM_PctMaxAssort*) and represents how consumers might substitute between stores with similar levels of product availability, given a change in price. For example, if a household typically shops at a retail chain that sells a wide variety of distinct UPCs, observing a positive and significant estimated coefficient would indicate that a household is more likely to switch to a store with a similar quantity of distinct UPCs, given a change in price.[10]

[9]Here a unique store from TDLinx is considered to be contained in the set of retail chains from IRI.

[10]We also calculate this assortment measure by strictly looking at the fruits and vegetables categories, as well as ethnic foods. Fruits and vegetable UPCs are taken from the following aisles: {Frozen Fruits & Vegetables, Produce, Shelf Stable Fruit, Shelf Stable Vegetable}; Ethnic UPCs are taken from the following

As an alternative to these assortment-type measures, we look for attributes that might necessitate, or attract, a household to visit more than one store. For example, one store might be the only retailer to carry a certain set of products, or even a specific item. Therefore, we look to create a retail-level "uniqueness score" that represents the share of UPCs that are sold only at the specific retailer (*UniqueScore*), and scale this number by the total number of unique UPCs at the store (*Breadth*).[11] In addition, since some stores are valued for the quality of their private label offerings, this measure may also capture the number of private label products or specialty products only available at certain stores, and not at others. This measure enters as an interaction term with own-price (*LNPxShUniqueUPC*). If the estimated coefficient on this variable is positive, then this suggests that households, on average, are less price-sensitive to stores where a higher proportion of their products are unique to that store. In other words, given an increase in own price, a household would continue to shop at the store since they know no other store would carry that item, therefore perhaps supporting the notion of continuing to shop at that store out of necessity. Alternatively, if the sign on *LNPxShUniqueUPC* is negative and significant, then this might indicate that the uniqueness of the store's product offerings acts as a measure of how many "luxury-type" items (e.g., gourmet chocolate or specialty foods) this store carries, which are generally associated with more elastic behavior.

Finally, we use information about the total square footage among all stores within each retail chain from TDLinx to construct an average square foot measure.[12] Similar to the *PctMaxAssort* calculation, we normalize the average square footage by dividing by the

aisle: {Ethnic}. However, in the final model specification, these measures are not used.

[11] Furthermore, using the information about the store aisles in the IRI data, we calculate this same measure by looking specifically at fruits and vegetables (*UniqueScore_FVonly*) and ethnic foods (*UniqueScore_ETHonly*), which represent items within these categories that can only be purchased at one store in the choice set. However, in the final specification, these measures are not used.

[12] Additional store attributes that were considered include the following: *AssortFVratio* reflects the total number of UPCs for each store that are in the fruit and vegetable aisles divided by the total number of UPCs outside the fruit and vegetable category. *AssortFVratio* is therefore meant to describe how prevalent fruit and vegetable products (considered to be healthy options) are relative to other products. It captures the ability to find a specific fruit or vegetable brand amidst all other choices in a census tract. Other amenities that can be captured via TDLinx are the presence of a pharmacy (*Pharm*) and whether a store has a gas station (*GasStation*).

maximum square footage of a store in the choice set, so $AvSqft$ takes a value between zero and one. This measure enters the model as an interaction term with own price ($LNPxAVSQFT$) and captures the price sensitivity of households to the overall store footprint. In general, larger stores might have lower overall prices, so including this variable can capture to some extent how price sensitive households are to having more space during their shopping period, controlling for prices.

5 Estimation

While the use of the DM-LA/AIDS model may prove to be a tractable and realistic technique to model store choice, there are three empirical concerns that are discussed before estimating the model: (*i*) censored data, (*ii*) store price construction and potential endogeneity, (*iii*) and market coverage limitations.

5.1 Censoring

Although the use of household-level scanner data offers a considerable amount of desirable information over other data sources, it introduces the issue of censoring. Despite restricting the choice set to 21 stores, a disproportionate number of zero expenditure shares exists. In the Philadelphia study area, the highest number of unique store shopping trips by a household during a month is nine. In some cases, households do not shop at any store during a month, and therefore the observable total monthly food expenditure of w_{ijt} equals zero. The number of households who do not make any purchases at stores in the choice set varies anywhere between 10% and 20% in a given month. While the use of traditional LA/AIDS demand models may not be practical to resolve this dimensionality issues due to the large number of integrals, the construction of the DM method reduces the estimation into a single equation and store expenditure shares are estimated using a Tobit model (Li et al. 2013, Rojas & Peterson 2008). Therefore, the ability of the DM method to easily accommodate the censored

nature of the data is another attractive benefit.

Similar to Li et al. (2013), we use a Tobit model and treat w_{ijt} as a latent variable w^*_{ijt}, where the observed share is assumed to be equal to the latent share whenever the latent share is greater than zero (Tobin 1958), such that:

$$w_{ijt} = \begin{cases} w^*_{ijt} & if \quad w^*_{ijt} > 0 \\ 0 & if \quad w^*_{ijt} \leq 0 \end{cases} \tag{5.1}$$

where

$$\begin{aligned} w^*_{ijt} = a_0 &+ \sum_{l=1}^{L} a_l z^a_{jl} + \sum_{h=1}^{H} \phi_h h_{ih} + \left(\gamma_0 + \sum_{m=1}^{M} \gamma_m z^{\gamma}_{jm}\right) \log p_{jt} \\ &+ \sum_{d=1}^{D} \left(\lambda^d_j \sum_{k \neq j}^{J} \delta^d_{jk} \log p_{kt}\right) + \sum_{c=1}^{C} \left(\lambda^c_j \sum_{k \neq j}^{J} \delta^c_{jk} \log p_{kt}\right) \\ &+ \left(\beta_0 + \sum_{n=1}^{N} \beta_n z^{\beta}_{jn}\right) \log \frac{x_{it}}{P^L_{it}} + \epsilon_{ijt} \end{aligned} \tag{5.2}$$

and $\epsilon_{ijt} \sim N(0, 1)$ The empirical results presented in this paper show results using a Tobit model as shown here in equation 5.2.

5.2 Store-Level Prices

Store choice, or in this case the share of total food expenditures a household allocates to a given store, depends on the prices faced by the household as well as other store-level attributes, such as product availability. To create a store-specific price, we use the observed price from the household-level transactions data (CNP) to create a store-level price index. This price index is modeled after the Fisher Ideal price index:

$$p_{jt} = \sqrt{\frac{\sum p_{kjt} q_{k0}}{\sum p_{k0} q_{k0}} \frac{\sum p_{kjt} q_{kjt}}{\sum p_{k0} q_{kjt}}} \tag{5.3}$$

where p_{kjt} and q_{kjt} are the price and purchase quantity for UPCs in category k at store j

during month t, respectively, and p_{k0} and q_{k0} are the market-level average price and quantity, or base price and base quantity, for UPCs in category k, respectively. Since not all stores sell the same categories of products, if a store does not sell any UPCs in a specific category, we treat the terms p_{kjt} and q_{kjt} as missing.[13]

Construction of a single price index for each store may present an empirical concern, specifically the potential for price endogeneity, insofar as the error term, ϵ_{ijt}, in equation 5.2 represents information other than price, such as taste preferences, that cannot be quantified by the data but has an impact on demand. Likewise, unobservable factors that influence store location can also shift the supply curve. Without being able to identify the direction in which these factors shift the demand or supply curves, the orthogonality assumption under OLS between the error term and regressors in equation 5.2 is violated. Therefore, it cannot be determined that a change in price is due to a demand shift or a supply shift, resulting in a need to adopt an appropriate identification strategy.

While we acknowledge that this price-specification approach has limitations, the Fisher Ideal price index has one major benefit over other price indexes. That is, the Fisher Ideal price index helps overcome substitution bias (Rahkovsky & Snyder 2015, Zhen et al. 2013), which may be especially important when estimating store choice. That is, if the overall price of the goods in a store increases resulting in decreased demand at that store, then using a price index that can accommodate this behavior is an attractive property. The Fisher Ideal price index in equation 5.3 is the geometric mean between the Laspeyres price index, $\frac{\sum p_{kjt} q_{k0}}{\sum p_{k0} q_{k0}}$, and the Paasche price index, $\frac{\sum p_{kjt} q_{kjt}}{\sum p_{k0} q_{kjt}}$. Both the Laspeyres and Paasche prices indexes report changes in price levels over time to indicate how much a basket of goods would cost in the base period. The Laspeyres price index uses the price in time period zero as its base price, while the Paasche price index uses the price in the current time period.

[13]We calculate two alternative price indexes. The first looks at all product categories ($StorePriceIndex_{all}$). The second (and the one used in this paper) constructs the price index that relies only on six categories ($StorePriceIndex_6$): {Fresh Eggs, Bottled Water, Carbonated Beverage, Cookies, Salty Snacks, and Milk}. Estimation was run using both index measures and showed similar results. Figure 2 graphically depicts the store-specific price index variation.

In effect, the Laspeyres price index may overstate inflation, and the Paasche price index may understate inflation. Since this price index is associated with each retailer j, rather than varying by household, fluctuations or shocks should be "averaged out." Using this price index may, therefore, minimize the correlation with the error term and alleviate potential endogeneity issues.

5.3 Market Coverage Limitations

IRI's CNP covers household transactions at stores that carry UPC-coded products. Therefore, information is not available for purchases from establishments that do not have UPCs such as farmers markets or food-away-from-home venues. Although this analysis fails to capture the full set of possible food sources across the households' food retail environment, what is observed are all food retailers that sell UPC-coded items.[14] Among those food retailers, the retail chains in the 21-store choice set receive at least 68% of all purchases recorded in the CNP transactions data. In addition, a wide variety of store types and variation in store attributes is captured in the choice-set data, thereby allowing the estimation to uncover policy-relevant food purchasing behaviors from the choice-set and study-area data.

6 Empirical Results

6.1 Selection of Model Specification

Following the approaches of Bonanno (2012) and Li et al. (2013), we follow two steps in selecting our preferred model specification. One challenge in estimating equation 5.2 is to incorporate a full set of policy-relevant store attributes while mitigating the risk of multicollinearity, as some of the store attributes from Table 2 may be more likely to shift the model's parameters than others. Therefore, since multicollinearity may be a concern due

[14]Note that independent stores are grouped into an "OTHER GROCERY" category and may appear to have a higher percentage of the market as a whole; however, individually, each store may only have a small percent. This can be verified using TDLinx.

to the number of demographics, store characteristics, and distance measures that can be included in the store-choice model, we measure the Variance Inflation Factor (VIF) for over 20 model specifications, varying the sets of dependent variables that interact with own price, product attributes, and expenditure. The final specification results in an average VIF of 2.42, suggesting the degree of multicollinearity is trivial (O'Brien 2007).

In addition, we generate elasticity estimates for eight of the twenty best-performing specifications.[15] Model specifications with highly unstable elasticities are discarded, yielding one preferred specification with an average own-price elasticity of -1.53.[16] The results for the preferred specification are presented in the following section and are based on the following attributes: three attributes that are interacted with own price, including the normalized average square footage ($LNPxAVSQFT$), share of unique UPCs ($LNPxShUniqueUPC$), and supercenter dummy ($LNPxSupercenter$); three distance measures, including store-to-store distance (DM_SDIST), channel type ($DM_Channel$), and percentage of maximum assortment ($DM_PctMaxAssort$); and one expenditure interaction, which is the household-to-store distance ($LNEXPxHHDIST$).

6.2 Results for Preferred Specification

We estimate equation 5.2 via a Tobit model and the results are presented in Table 3. The set of demographic variables includes the following: household size, age of the head of the household, median and income, as well as household-level fixed effects for SNAP eligibility, race identifiers (white, black, Asian), ethnicity, presence of a dependent under age 18, presence of an adult over age 65, whether the household has a female head, and whether either head of household received a college education. In addition, we include a variable that indicates whether the household lives in a low-income tract, as defined by the USDA's Food Access

[15] Elasticity equations for the censored store-choice model are presented in Appendix A.

[16] The range of own price elasticities is $[-4.77, 0.09]$ with a median of -1.22 and coefficient of variation of -0.76. The average elasticities by channel type are: convenience stores (-0.23), mass merchandisers (-3.95), dollar stores (-1.11), grocery stores (-1.32), and supercenter (-0.58).

Research Atlas.[17]

We first discuss the set of results involving the own price and attributes that are interacted with own price. Despite the suspected presence of endogeneity in the model, Table 3 shows that the estimated coefficient on own-price ($LNP = -0.2703$) is negative and significant, indicating that as the price of the store increases, the expenditure share decreases as expected. The coefficient on the interaction term between price and supercenter is also negative and significant ($LNPxSupercenter = -0.1559$). Households who shop at supercenters are generally more price elastic, so a negative and significant result is expected. The coefficient on the interaction term between price and square footage is positive and significant ($LNPxAVSQFT = 0.5698$), implying that, all else constant, households are less price sensitive to stores with higher square footage. This result may be linked to one-stop shopping behavior: i.e., households may be less sensitive to price in larger stores if, in a one-stop shopping trip, they are willing to accept some higher prices in order to purchase a larger market basket in one shopping trip.

Next, the coefficient on the interaction term between price and share of unique UPCs is positive and significant ($LNPxShUniqueUPC = 0.4437$). A positive and significant sign on this variable suggests that households are less price-sensitive to a store offering a higher proportion of unique products. This result is consistent with a number of shopping behaviors: First, if a household can shop at only one store to purchase a specific set of items with very limited substitutability, then a positive sign supports the notion that store uniqueness leads to more inelastic demand. Second, because this measure also captures private label products, a positive and significant sign could imply that households are loyal to the set of retail-specific private label products. Finally, households who regularly purchase high-end or specialty items only available perhaps at one store in the choice set may be overall less price elastic.

Another set of observations focuses on the parameter estimates associated with the DM terms. These coefficients can be interpreted as the households' response to price changes as

[17]Household demographics are included as intercept shifters terms, for which only some are significant.

stores become more competitive and similar in attributes. A positive coefficient, for example, implies that households facing a price increase would tend to switch to stores with similar attributes. The estimated coefficient associated with store closeness in terms of physical proximity is positive and significant ($DM_SDIST = 0.0921$), supporting the notion that if the price of one store increases, households will substitute to another store that is close in physical distance, rather than travel to a store farther away. By suggesting that a store's location relative to other stores is a key element in understanding the underlying behavior of consumers shopping habits, this result may provide an important policy insight. In other words, convenience matters, at least as it pertains to distance. In addition, households are more likely to switch to a store within the same channel ($DM_Channel = 0.0033$). Both these results suggest that stores strongly compete for shoppers using location and channel type, and these relationships are important for investigating impacts to consumer welfare.[18]

The substitutability metric that measures the expanse of product offerings compared to all other retailers in the choice set is negative and significant ($DM_PctMaxAssort = -0.0240$), implying that households will switch to a store that has a very different level of overall product assortment. This result has two intriguing potential explanations. First, this result suggests that that food retailers use their product assortments to strategically position themselves to avoid direct competition. In other words, when faced with a price increase, consumers are nudged to switch to stores that do not carry the same set of products. In addition, the switching result is also consistent with the notion that households have primary and secondary stores that are different in terms of product assortment. Faced with a price increase, consumers may switch to their secondary store instead of seeking out a store that is similar in product assortment to their primary store. The ability to understand this type of substitution behavior appears to be a direct potential benefit of the DM model allowing for multiple shopping trips.

[18] An extension of this work will be to develop a supply side model to test changes to the food retailing landscape by simulating the addition of a new store of the a certain attribute bundle, including channel type and store distance, to investigate welfare changes for consumers and food retailers.

The final set of coefficients in Table 3 involves the expenditure term and interaction between expenditure and household distance. The estimated parameter for the expenditure term is positive and significant ($LNEXP = 0.4350$), implying that as a household's budget increases, so does its share of expenditure at that store. When expenditure is interacted with average household distance to a store, the estimated coefficient is negative and significant ($LNEXPxHHDIST = -0.0236$). Adding this result to the coefficient on $LNEXP$ suggests a dampening effect, meaning the positive effect that the food budget has on store shares decreases as the distance to the store increases. Conversely, this result also means that when a household's food-at-home budget decreases, the household may spend a larger share on food stores that are farther away. This result is, therefore, consistent with low-income households traveling substantial distances to get to a potential primary store, which might be a one-stop, low price store. Following from the result above, physical distance plays a key role in understanding consumer behavior.

Using the estimated parameters, we calculate and present the conditional own-price elasticities in Table 4. There are three major points to be made about these results and their possible implications: First, mass merchandisers have substantially larger elasticities than other channels (-3.95 for mass merchandisers versus -0.81 for all others). This result suggests that households who shop at mass merchandisers tend to be more price-elastic compared to all other channels. Second, convenience stores and supercenters have the lowest elasticities (-0.23 and -0.58, respectively). This finding indicates that demand for both convenience stores and supercenters are inelastic, which may be motivated by a consumer's valuation for convenience and one-stop shopping. Coupled with the results discussed above that indicate store-to-store closeness significantly contributes to switching behavior, these findings support the idea that convenience is an important factor in store choice, regardless of the price implications. Finally, specialty, or high-end, supermarkets have higher elasticities when compared with other supermarkets.

From a policy standpoint, the estimation and elasticity results could imply that programs

whose aim is to subsidize new supermarket entry may need to highlight the strategic nature of supermarket location. In other words, given the inelastic demand for convenience stores, for example, the success of a new supermarket might be impeded by the nature of the extant food retailing environment, in particular competing for loyal convenience-store customers. On the other hand, if a new entry were large enough (as measured by square footage) or unique enough (in terms of UPCs), then perhaps proximate location is less important, whereby this type of new store could be a located farther way from convenience stores or other incumbents.

7 Policy Implications and Discussion

As the food retailing environment continues to evolve, the aim of this research is to expound on the relationship between consumer behavior, shopping decisions, and food access. With the emergence of recent literature indicating that simply increasing access is not a full-proof solution to solving issues surrounding limited food access, there is evidence that other factors, such as the prices households face, assortment levels, and other attributes the store offers are at play. This paper attempts to understand some of the underlying preferences that might explain consumers' shopping decisions, including what store features households value, especially in low access communities, and to contribute to the ongoing discussion of identifying the most appropriate policy prescriptions.

To model store choice in this paper, and in effect examine demand for store attributes, we employ the DM method of Pinske, Slade, and Brett (2002). The use of the DM method offers a straightforward way to measure substitution patterns between stores with similar attributes. In addition, the importance of product assortment, store services, and price can be incorporated into a flexible model of store selection within the Philadelphia metro-area. The relationship between differentiated stores and their attributes is captured by the relative distance between the stores along several dimensions (e.g., assortment, service levels, and price), as well as geographical distance. Via the DM method, households are able choose the

stores that possess the most desirable set of attributes and substitute between stores that are relatively closer in proximity across the set of characteristics. In addition, by combining several rich data sets – i.e., Nielsen's TDLinx store-attribute data, the IRI Consumer Network scanner data, and the IRI Store scanner data – this analysis allows a look inside stores and therefore supports an investigation into both new and long-standing food-policy questions. The use of these data sources supports a more complete picture of both the food environment and consumer behavior, and it is our hope that these methods and results generate significant interest and discussion with applications in marketing, health, and food policy.

The flexible DM store-choice model generates a number of conclusions about shopping preferences: Households are less price-sensitive when it comes to the size of the store. Hence, one-stop shopping is valued by Philadelphia households. The relative uniqueness of the products on store shelves is also highly valued. On the other hand, households are more price sensitive at supercenters. When a price increase induces a household to switch stores, then switches are more likely to occur at nearby stores and stores in the same channel. However, consumers are more likely to switch to a store with a very different product assortment level. This finding could be evidence of the importance of primary and secondary stores that serve complementary roles. Lastly, households seem very loyal to convenience stores, supercenters, and some dollar stores.

These conclusions generate some important policy-relevant implications: To succeed, a new store entry first and foremost has to compete with convenience stores, supercenters, and dollar stores, and this competition intensifies with proximate distance. Large stores that could presumably fulfill one-stop shopping needs are also intense competitors. A number of these findings relate to overall convenience, a part of the shopping experience on which any new entry must focus. The new entrant's product assortment, however, is a more complicated strategic concern. Uniqueness may also help a store compete, but consumers more easily switch to stores with very different levels of total product assortment. Policy interventions that are focused on improving the variety at specific stores, such as the Healthy Corner

Store Initiative in Philadelphia or the Staple Food Ordinance in Minneapolis, could be an alternative to the lump-sum subsidy approach as a cost-effective solution.

While this paper may offer insights to address underlying behavioral consumer shopping preferences, there are limitations. First, our results are valid under the assumption that they are consistent and not affected by endogeneity bias. We acknowledge the possibility that our store-level price indexes are correlated with unobserved drivers of household store expenditure shares. This is particularly relevant for drawing insights on policy recommendations around consumer purchasing behavior and store choice. Finding an appropriate price instrument to use in our analysis presents several challenges, as our analysis is limited in geographic scope (i.e., specifically the Philadelphia metro-area) and variation in time (i.e., our analysis spans twelve months). Future research focusing on accounting for multiple markets and/or a longer time period could provide enough variation in the data to allow for solutions in the endogeneity problem. Nonetheless, we stand by the usefulness and novelty of our approach, which presents an application to looking at consumers' responsiveness to store attributes that could be used in other contexts.

In addition, as previous research has suggested, policies aimed only at improving access to healthy foods in underserved areas may not alleviate the persistent socioeconomic disparities as it relates to addressing diet quality. Further research may be needed to determine an integrated, cost-effective policy solution. In addition, consumer knowledge about nutrition and unobservable food preferences, which are both difficult to measure, are not incorporated into the findings of this paper. Nonetheless, this research leads to some important conclusions about store-type and store-attribute preferences that may not be attainable with less flexible store-choice models.

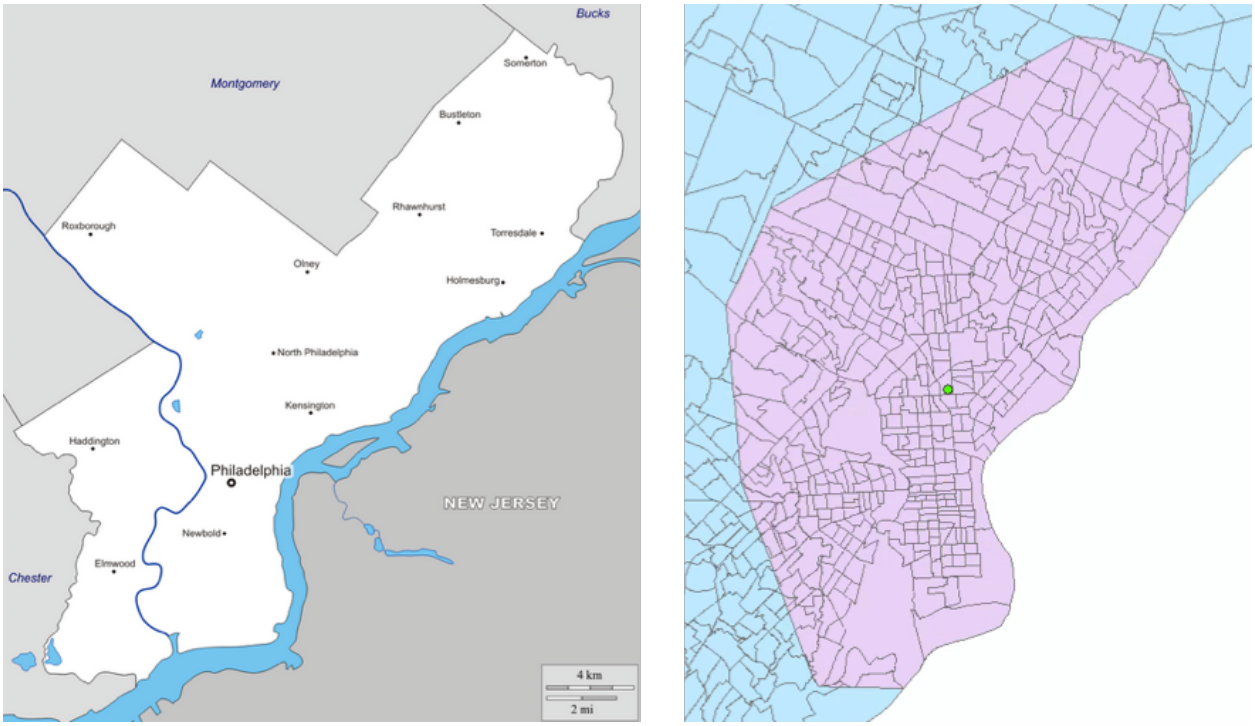

Figure 1: Philadelphia Metro-Area

Figure 1 presents a side-by-side depiction of Philadelphia county (L) and the Philadelphia metro-area convex hull (R) on which this analysis is focused.

Table 1: Household Summary Statistics – Select Demographics (n=267)

Variable	Overall		Low-Income		Non Low-Income	
Demographics						
$HHSize$	2.38	(1.42)	2.62	(1.84)	2.30	(1.24)
$AgeHead$	60.37	(12.83)	61.06	(14.57)	60.14	(12.20)
$MedInc$ ($00s)	564.68	(337.84)	204.33	(129.35)	687.80	(296.38)
$SNAPelig$ (share)	0.25	(0.43)	–		–	
$White$ (share)	0.57	(0.50)	0.51	(0.50)	0.59	(0.49)
$Black$ (share)	0.37	(0.48)	0.41	(0.50)	0.36	(0.48)
$Asian$ (share)	0.02	(0.14)	0.01	(0.12)	0.02	(0.14)
$Hispanic$ (share)	0.04	(0.21)	0.06	(0.24)	0.04	(0.20)
$FemaleHead$ (share)	0.76	(0.43)	0.75	(0.44)	0.76	(0.43)
$ChildUnder18$ (share)	0.07	(0.26)	0.04	(0.21)	0.08	(0.27)
$SeniorCit$ (share)	0.37	(0.48)	0.43	(0.50)	0.35	(0.48)
$College$ (share)	0.10	(0.31)	0.03	(0.17)	0.13	(0.34)
Expenditure and Shopping Behavior						
$ExpenditureShare(w_{ijt})$	0.37	(0.34)	0.37	(0.34)	0.38	(0.34)
$TotalMonthlyFoodExpenditure$ ($)	168.03	(173.08)	149.00	(164.04)	174.54	(176.00)
$LNEXP$	2.16	(0.49)	2.14	(0.51)	2.17	(0.48)
$TripsPerMonth$	7.82	(5.37)	8.70	(6.55)	7.52	(4.86)
$NoUniqueRetailersVisited$	4.30	(2.49)	4.47	(2.80)	4.24	(2.37)
Food Environment						
$LowAccessTract, 1/2mi$ (share)	0.63	(0.48)	0.56	(0.50)	0.65	(0.48)
$LowIncomeTract$ (share)	0.58	(0.49)	0.72	(0.45)	0.53	(0.50)
$NoVehicleTract$ (share)	0.34	(0.47)	0.37	(0.49)	0.33	(0.47)
$HHtoStoreDist$ (mi)	7.46	(2.65)	7.38	(2.72)	7.49	(2.62)
Household Count	267		68		199	
Share of Households	–		0.25		0.75	

Note: Standard errors reported in parentheses.

Table 2: Store-Level Summary Statistics – Select Attributes (j=21)

Attribute	**Mean**	**St.Dev.**	**Min**	**Max**
Continuous Variables				
$StorePriceIndex_{all}$	1.02	(0.35)	0.51	2.92
$StorePriceIndex_{6}$	1.01	(0.39)	0.40	3.07
$Breadth$ (00s)	8.98	(12.23)	0.13	53.14
$PctMaxAssort$	0.19	(0.25)	0.003	1
$PctMaxAssort_FV$	0.18	(0.26)	0	1
$PctMaxAssort_Ethnic$	0.16	(0.24)	0	1
$UniqueScore$	0.44	(0.09)	0.21	0.70
$UniqueScore_FVonly$	0.06	(0.04)	0	0.17
$UniqueScore_ETHonly$	0.01	(0.01)	0	0.04
$AssortFVratio$	0.06	(0.04)	0	0.21
$AvSqft$	0.28	(0.26)	0.2	1
$StoretoStoreDist$ (mi)	7.58	(0.88)	6.56	10.67
Discrete Variables	Counts			
$Grocery$	12		–	
$Supercenter$	1		–	
$MassMerchandiser$	3		–	
$Convenience$	2		–	
$DollarStore$	3		–	
$Pharmacy$	9		–	
$GasStation$	4		–	
Store Count	21			

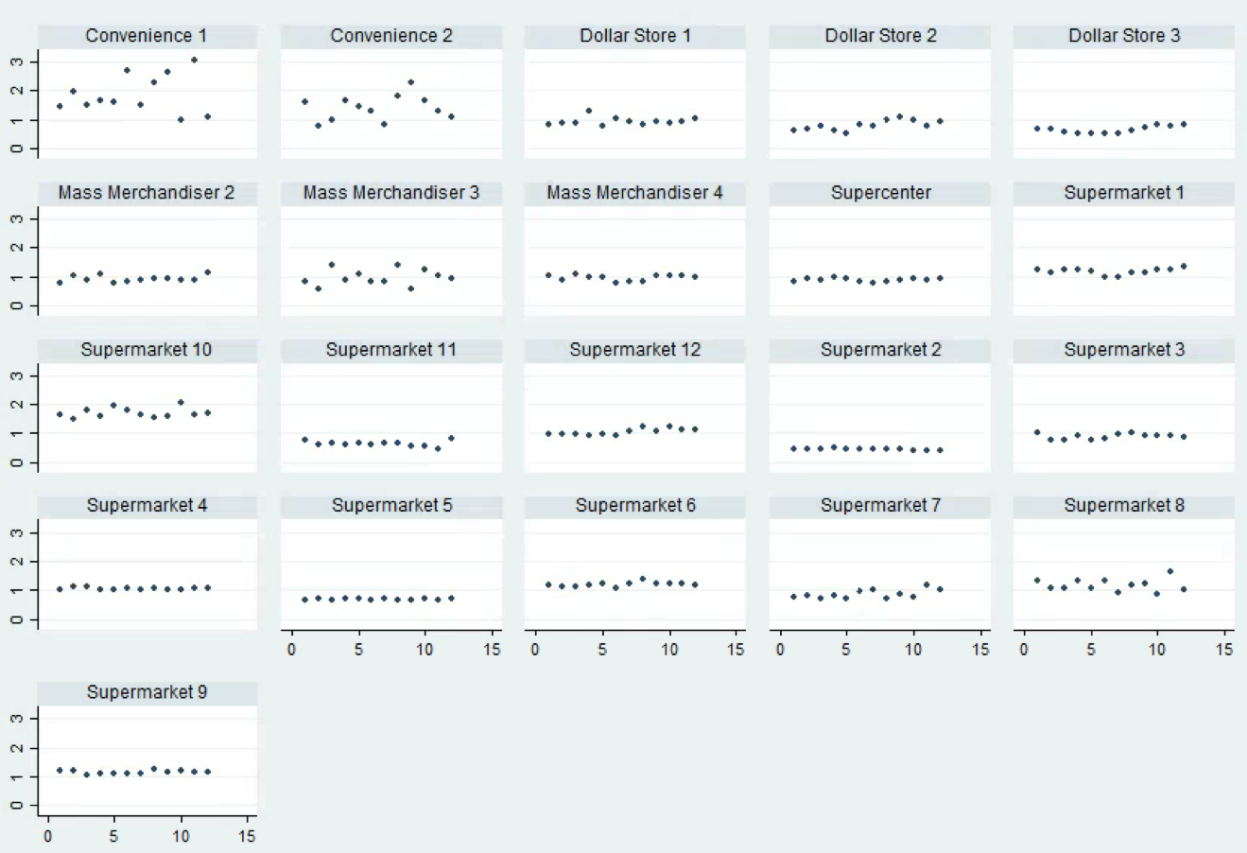

Figure 2: Store-Level Price Indexes

Figure 2 shows store-level price indexes ($StorePriceIndex_6$) for each store (j=21), spanning the twelve months in 2012 (horizontal axis) and the corresponding value for the price index (vertical axis).

Table 3: Estimated Parameters for Preferred Model Specification

Variables	**Coefficient**		**Standard Errors**
$HHSize$	−0.0069	*	0.0039
$AgeHead$	−0.0001		0.0005
$MedInc$	0.0000		0.0000
$SNAPelig$	0.0357	**	0.0357
$White$	0.0572	**	0.0289
$Black$	0.0449		0.0292
$Asian$	−0.0513		0.0474
$Hispanic$	−0.0036		0.0241
$FemaleHead$	0.0470	***	0.0108
$ChildUnder18$	−0.0180		0.0188
$SeniorCit$	0.0347	**	0.0141
$College$	−0.0207		0.0160
$LowIncomeTract$	−0.0514	***	0.0101
LNP	−0.2703	***	0.0494
$LNPxSupercenter$	−0.1559	***	0.0378
$LNPxAVSQFT$	0.5698	***	0.0295
$LNPxShUniqueUPC$	0.4437	***	0.0876
DM_SDIST	0.0921	***	0.0059
$DM_Channel$	0.0033	***	0.0003
$DM_PctMaxAssort$	−0.0240	***	0.0005
$LNEXP$	0.4350	***	0.0107
$LNEXPxHHDIST$	−0.0236	***	0.0010
$Constant$	−1.1985	***	0.0901
VIF	2.42		
No. of Obs.	67,284		
No. of Households	262		

*, **, *** represent 10%, 5%, and 1% significance levels, respectively.

Table 4: Own-Price Elasticities

Retail Chain	**Own-Price Elasticity**		**Standard Errors**
Grocery 1	−1.12	***	0.1203
Grocery 2	−1.16	**	0.4746
Grocery 3	−2.07	***	0.3528
Grocery 4	−1.33	***	0.0802
Grocery 5	−0.70		0.4261
Grocery 6	−1.41	***	0.1870
Grocery 7	−0.64	***	0.2129
Grocery 8	−1.22	***	0.3490
Grocery 9	−1.07	***	0.0479
Grocery 10	−1.71	***	0.4048
Grocery 11	−1.84	***	0.5325
Grocery 12	−1.59	***	0.3874
Supercenter	−0.58		0.4896
Mass Merchandiser 1	−3.84	***	0.2408
Mass Merchandiser 2	−3.23	***	0.3705
Mass Merchandiser 3	−4.77	***	0.3250
Convenience 1	−0.55	***	0.2098
Convenience 2	0.09		0.4433
Dollar Store 1	−1.13	***	0.3190
Dollar Store 2	−0.75	*	0.3857
Dollar Store 3	−1.48	***	0.5090
Min	−4.77		
Max	0.09		
Mean	−1.53		
St. Dev.	1.15		
CoV	−0.76		
Mean Own-Price Elasticity by Channel Type			
Grocery	−1.32		
Supercenter	−0.58		
Mass Merchandiser	−3.95		
Convenience	−0.23		
Dollar Store	−1.11		

A Elasticity Formulas for Censored Demand

A general definition of the uncompensated elasticities of demand for LA/AIDS from Green and Alston (1990) is:

$$\begin{aligned}\eta_{jk} &= \frac{d\ln Q}{d\ln P_k} \\ &= \frac{d\ln w_j}{d\ln P_k} - \frac{d\ln P_j}{d\ln P_k} \\ &= -\delta_{jk} + \frac{d\ln w_j}{d\ln P_k} \\ &= -\delta_{jk} + \frac{dw_j}{d\ln P_k} * \frac{1}{w_j}\end{aligned} \tag{A.1}$$

where $w_j = \frac{P_j * Q_j}{X}$ and the price index (P) is defined as $\log P \equiv \sum_{j=1}^{J} w_j^0 \log p_j$.

Conditional and unconditional elasticities are obtained following Bonanno (2010), Yen and Huang (2002), and Li et al. (2016). Recall the censored demand equation 5.2:

$$w_{ijt} = \begin{cases} w^*_{ijt} & if \quad w^*_{ijt} > 0 \\ 0 & if \quad w^*_{ijt} \leq 0 \end{cases}$$

where

$$\begin{aligned}w^*_{ijt} = a_0 &+ \sum_{l=1}^{L} a_l {z^a}_{jl} + \sum_{h=1}^{H} \phi_h h_{ih} + \left(\gamma_0 + \sum_{m=1}^{M} \gamma_m {z^\gamma}_{jm}\right) \log p_{jt} \\ &+ \sum_{d=1}^{D} \left({\lambda^d}_j \sum_{k\neq j}^{J} {\delta^d}_{jk} \log p_{kt}\right) + \sum_{c=1}^{C} \left({\lambda^c}_j \sum_{k\neq j}^{J} {\delta^c}_{jk} \log p_{kt}\right) \\ &+ \left(\beta_0 + \sum_{n=1}^{N} \beta_n {z^\beta}_{jn}\right) \log \frac{x_{it}}{{P^L}_{it}} + \epsilon_{ijt}\end{aligned}$$

Differentiate the conditional expectation of the expenditure share with respect to price k to derive the conditional elasticities. The conditional expectation for store j is as follows:

$$E[w_j | w_j > 0] = f_j(\theta) + \sigma_j \left[\frac{\phi(f_j(\theta)/\sigma_j)}{\Phi(f_j(\theta)/\sigma_j)}\right] \tag{A.2}$$

where

$$E[w_j] = E[w_j|w_j > 0] * Pr(w_j > 0) \tag{A.3}$$

and

$$Pr(w_j > 0) = Pr(\epsilon > -f_j(\theta)) = 1 - \Phi(-f_j(\theta)/\sigma_j) = \Phi(f_j(\theta)/\sigma_j) \tag{A.4}$$

where $w_j = f_j(\theta) + \epsilon$ and $\theta \in (a, \gamma, \beta)$ are the estimable parameters. The conditional elasticity represents the percentage change in expenditure at store j for those households who shop at store j given a percentage change in the prices at store k.

Combining A.1 and A.2, the conditional price elasticity of demand is:

$$\eta_{jk} = -\delta jk + \frac{dE[w_j|w_j > 0]}{d\ln P_k} * \frac{1}{E[w_j|w_j > 0]} \tag{A.5}$$

where

$$\begin{aligned}
\frac{dE[w_j|w_j > 0]}{d\ln P_k} &= f_j'(\theta) + \sigma_j \frac{d\lambda(f_j(\theta)/\sigma_j)}{d\ln P_k} * \frac{1}{\sigma_j} * f_j'(\theta) \\
&= f_j'(\theta)\left[1 + \frac{d\lambda(f_j(\theta)/\sigma_j)}{d\ln P_k}\right] \\
&= f_j'(\theta)\left[1 - \lambda\Big(\frac{f_j(\theta)}{\sigma_j}\Big) * \left(\frac{f_j(\theta)}{\sigma_j} + \lambda\left(\frac{f_j(\theta)}{\sigma_j}\right)\right)\right] \\
&= f_j'(\theta)\left[1 - \frac{\lambda}{\sigma_j} f_j(\theta) + \lambda^2\right] \\
&= f_j'(\theta)\left[1 - \frac{\lambda}{\sigma_j}[f_j(\theta) + \sigma_j\lambda]\right]
\end{aligned}$$

$$= f_j'(\theta)\Big[1 - \frac{\lambda}{\sigma_j}[E[w_j|w_j > 0]]\Big] \tag{A.6}$$

such that $\lambda = \frac{\phi(f_j(\theta)/\sigma_j)}{\Phi(f_j(\theta)/\sigma_j)}$ and $\frac{d\lambda(f_j(\theta)/\sigma_j)}{d\ln P_k} = -\lambda\left(\frac{f_j(\theta)}{\sigma_j}\right) * \left(\frac{f_j(\theta)}{\sigma_j} + \lambda(\frac{f_j(\theta)}{\sigma_j})\right)$ (Hayashi, 2000).

Substituting A.6 into A.5, yields the following:

$$\eta_{jk} = -\delta_{jk} + \left[\frac{1}{E[w_j|w_j > 0]} - \frac{\lambda}{\sigma_j}\right] * f_j'(\theta) \tag{A.7}$$

where

$$\eta_{jk}^C = \begin{cases} -\delta_{jk} + \left[\frac{1}{E[w_j|w_j>0]} - \frac{\lambda_j}{\sigma_j}\right] * \left[(\gamma_0 + \sum_{m=1}^{M}\gamma_m z_{jm}^{\gamma}) - (\beta_0 + \sum_{n=1}^{N}\beta_n z_{jn}^{\beta})w_j\right] & , \quad j = k \\ \left[\frac{1}{E[w_j|w_j>0]} - \frac{\lambda_j}{\sigma_j}\right] * \left[\left(\sum_{d=1}^{D}(\lambda_j^d \delta_{jk}^d) + \sum_{c=1}^{C}(\lambda_j^c \delta_{jk}^c)\right) - (\beta_0 + \sum_{n=1}^{N}\beta_n z_{jn}^{\beta})w_k\right] & , \quad j \neq k \end{cases} \tag{A.8}$$

Following Yen and Huang (2002) and Li et al. (2016), we derive the unconditional elasticity equation, which may be calculated as the sum of the conditional elasticity, η_{jk}^C, and probability elasticity, η_{jk}^P:

$$\eta_{jk}^U = \eta_{jk}^C + \eta_{jk}^P$$

The unconditional price elasticity, η_{jk}^U, measures the overall response of changes in store expenditure share given an increase in price (or other explanatory variable). According to Yen and Huang (2002), the probability elasticity represents the sensitivity of the probability or likelihood that a household will increase (or decrease) their expenditure at a store given a percentage change in the price of the store. The probability elasticity is derived by taking the derivative of $Pr(w_j > 0) = \Phi(f_j(\theta)/\sigma_j)$ with respect to price k:

$$\begin{aligned} \frac{d\ln\Phi(f_j(\theta)/\sigma_j)}{d\ln P_k} &= \frac{d\ln\Phi(.)}{d\Phi(.)} * \frac{d\Phi(.)}{d\ln P_k} \\ &= \frac{1}{\Phi(.)} * \frac{d\Phi(.)}{d\ln P_k} \\ &= \frac{1}{\Phi(.)} * \frac{f_j'(\theta}{\sigma_j}\phi(.) \\ &= \sigma_j^{-1}\lambda * f_j'(\theta) \end{aligned}$$

$$\eta_{jk}^{P} = \begin{cases} \sigma_j^{-1}\lambda_j * \left[(\gamma_0 + \sum_{m=1}^{M} \gamma_m z_{jm}^{\gamma}) - (\beta_0 + \sum_{n=1}^{N} \beta_n z_{jn}^{\beta}) w_j\right] & , \quad j = k \\ \sigma_j^{-1}\lambda_j * \left[\left(\sum_{d=1}^{D}(\lambda_j^d \delta_{jk}^d) + \sum_{c=1}^{C}(\lambda_j^c \delta_{jk}^c)\right) - (\beta_0 + \sum_{n=1}^{N} \beta_n z_{jn}^{\beta}) w_k\right] & , \quad j \neq k \end{cases} \tag{A.9}$$

References

Alwitt, L. F. & Donley, T. D. (1997), 'Retail stores in poor urban neighborhoods', *Journal of Consumer Affairs* **31**(1), 139–164.

Bell, D. R., Ho, T.-H. & Tang, C. S. (1998), 'Determining where to shop: Fixed and variable costs of shopping', *Journal of Marketing Research* pp. 352–369.

Bell, D. R. & Lattin, J. M. (1998), 'Shopping behavior and consumer preference for store price format: Why "large basket" shoppers prefer edlp', *Marketing Science* **17**(1), 66–88.

Bonanno, A. (2012), 'Functional foods as differentiated products: the italian yogurt market', *European Review of Agricultural Economics* p. jbr066.

Bonanno, A., Chenarides, L. & Goetz, S. J. (2012), Limited food access as an equilibrium outcome: An empirical analysis, *in* '2012 AAEA/EAAE Food Environment Symposium, May 30-31, Boston, MA', Agricultural and Applied Economics Association.

Bonanno, A. & Lopez, R. A. (2009), 'Competition effects of supermarket services', *American Journal of Agricultural Economics* **91**(3), 555–568.

Briesch, R. A., Chintagunta, P. K. & Fox, E. J. (2004), Assortment, price, and convenience: Modeling the determinants of store choice, Technical report, Working Paper.

Clarke, I. (2000), 'Retail power, competition and local consumer choice in the uk grocery sector', *European Journal of Marketing* **34**(8), 975–1002.

Cummins, S., Flint, E. & Matthews, S. A. (2014), 'New neighborhood grocery store increased awareness of food access but did not alter dietary habits or obesity', *Health Affairs* **33**(2), 283–291.

Deaton, A. & Muellbauer, J. (1980), 'An almost ideal demand system', *The American economic review* **70**(3), 312–326.

Dunkley, B., Helling, A. & Sawicki, D. S. (2004), 'Accessibility versus scale examining the tradeoffs in grocery stores', *Journal of Planning Education and Research* **23**(4), 387–401.

Ellickson, P. B. (2007), 'Does sutton apply to supermarkets?', *The RAND Journal of Economics* **38**(1), 43–59.

Ellickson, P. B. & Grieco, P. L. (2013), 'Wal-mart and the geography of grocery retailing', *Journal of Urban Economics* **75**, 1–14.

Fitzpatrick, K. & Ver Ploeg, M. (2013), On the road to food security? vehicle ownership and access to food. Paper Prepared for Research on Connections between Health and SES Using PSID Data Survey Research Center, Institute for Social Research, University of Michigan.

Handbury, J., Rahkovsky, I. & Schnell, M. (2015), What drives nutritional disparities? retail access and food purchases across the socioeconomic spectrum, Working Paper Series 21126, National Bureau of Economic Research.

Hoch, S. J., Dreze, X. & Purk, M. E. (1994), 'EDLP, Hi-Lo, and margin arithmetic', *The Journal of Marketing* pp. 16–27.

Holmes, T. J. (2011), 'The diffusion of wal-mart and economies of density', *Econometrica* **79**(1), 253–302.

Kaufman, P. (2007), 'Strong competition in food retailing despite consolidation', *Amber Waves* **5**(1), 5.

Kyureghian, G. & Nayga, R. M. (2013), 'Food store access, availability, and choice when purchasing fruits and vegetables', *American Journal of Agricultural Economics* **95**(5), 1280–1286.

Kyureghian, G., Nayga, R. M. & Bhattacharya, S. (2013), 'The effect of food store access

and income on household purchases of fruits and vegetables: A mixed effects analysis', *Applied Economic Perspectives and Policy* **35**(1), 69–88.

Lee, H. (2012), 'The role of local food availability in explaining obesity risk among young school-aged children', *Social science & medicine* **74**(8), 1193–1203.

Leibtag, E. S. & Kaufman, P. R. (2003), Exploring food purchase behavior of low-income households: How do they economize?, Technical report, USDA, Economic Research Service.

Leszczyc, P. T. P., Sinha, A. & Sahgal, A. (2004), 'The effect of multi-purpose shopping on pricing and location strategy for grocery stores', *Journal of Retailing* **80**(2), 85–99.

Li, J., Jaenicke, E. C. & Anekwe, T. D. (2013), Health-related product attributes and purchasing behavior in the ready-to-eat cereal market: An application with household-level, censored data, *in* '2013 Annual Meeting, August 4-6, 2013, Washington, DC', Agricultural and Applied Economics Association.

Moore, L. V. & Diez Roux, A. V. (2006), 'Associations of neighborhood characteristics with the location and type of food stores', *American Journal of Public Health* **96**(2), 325–331.

Moschini, G. (1995), 'Units of measurement and the stone index in demand system estimation', *American journal of agricultural economics* **77**(1), 63–68.

Pinkse, J., Slade, M. & Brett, C. (2002), 'Spatial price competition: A semiparametric approach', *Econometrica* **70**(3), 1111–1153.

Pofahl, G. M. & Richards, T. J. (2009), 'Valuation of new products in attribute space', *American Journal of Agricultural Economics* **91**(2), 402–415.

Rahkovsky, I. & Snyder, S. (2015), Food choices and store proximity, Economic Research Report 195, U.S. Department of Agriculture, Economic Research Service.

Rojas, C. (2008), 'Price competition in u.s. brewing', *The Journal of Industrial Economics* **56**(1), 1–31.

Rojas, C. & Peterson, E. B. (2008), 'Demand for differentiated products: Price and advertising evidence from the u.s. beer market', *International Journal of Industrial Organization* **26**(1), 288–307.

Shaked, A. & Sutton, J. (1987), 'Product differentiation and industrial structure', *The Journal of Industrial Economics* pp. 131–146.

Sinha, A. (2000), 'Understanding supermarket competition using choice maps', *Marketing Letters* **11**(1), 21–35.

Smith, H. (2004), 'Supermarket choice and supermarket competition in market equilibrium', *The Review of Economic Studies* **71**(1), 235–263.

Sutton, J. (1991), *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration*, MIT press.

Sweitzer, M., Levin, D., Okrent, A., Muth, M. K., Brown, D., Capogrossi, K., Karns, S. A., Siegel, P. & Zhen, C. (2016), Understanding iri household-based and store-based scanner data, Technical report, USDA, Economic Research Service.

Taylor, R. & Villas-Boas, S. B. (2016), 'Food store choices of poor households: A discrete choice analysis of the national household food acquisition and purchase survey (foodaps)', *American Journal of Agricultural Economics* **98**(2), 513–532.

Tobin, J. (1958), 'Estimation of relationships for limited dependent variables', *Econometrica* pp. 24–36.

USDA, E. R. S. (2009), Access to affordable and nutritious food: Measuring and understanding food deserts and their consequences, Report to congress, USDA, Economic Research Service.

Ver Ploeg, M., Breneman, V., Dutko, P., Williams, R., Snyder, S., Dicken, C. & Kaufman, P. (2012), Access to affordable and nutritious food: Updated estimates of distance to

supermarkets using 2010 data, Economic Research Report 143, USDA, Economic Research Service.

Ver Ploeg, M., Mancino, L., Todd, J., Clay, D. M. & Scharadin, B. (2015), Where do americans usually shop for food and how do they travel to get there? initial findings from the national household food acquisition and purchase survey, Economic Information Bulletin 138, USDA Economic Research Service.

Ver Ploeg, M. & Rahkovsky, I. (2016), 'Recent evidence on the effects of food store access on food choice and diet quality', *Amber Waves* p. 1C.

Zhen, C., Finkelstein, E. A., Nonnemaker, J. M., Karns, S. A. & Todd, J. E. (2013), 'Predicting the effects of sugar-sweetened beverage taxes on food and beverage demand in a large demand system', *American journal of agricultural economics* p. aat049.

Zwiebach, E. (2014), 'The growth list: Chains on the fast track', *Supermarket News* .
URL: *http://search.proquest.com.ezaccess.libraries.psu.edu/docview/1520777169?accountid=13158*