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

Lauren Chenarides
Department of Agricultural Economics, Sociology, and Education
Penn State University
Email: LEC201@psu.edu

Edward C. Jaenicke
Department of Agricultural Economics, Sociology, and Education
Penn State University
Email: ECJ3@psu.edu

WORKING PAPER

Abstract

Shopping and store-choice decisions are intertwined with firms' decisions to enter or exit a market, as well as with heterogeneous consumer demographics. The importance of food access becomes apparent in determining where households choose to purchase food, as consumers residing in underserved areas are faced with shopping at non-traditional stores that may result in negative welfare outcomes. Research regarding consumer purchasing behavior has traditionally looked at store choice as a nested discrete choice decision; however, we propose an alternative approach that models consumer store choice preferences for store attribute bundles, including product assortment, store services, and price via the Distance Metric (DM) method of Pinkse, Slade, and Brett (2002). Methodologically, 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 described to create a more flexible model of store selection within different markets across the U.S.

Keywords: store choice, scanner data, food access, food deserts, distance metric method

Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31-August 2

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

In a 2012 report to Congress, the USDA highlights supermarket availability as an 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). The importance of food access becomes apparent in determining where households choose to purchase food, as consumers residing in underserved areas are faced with shopping at non-traditional stores that offer a limited variety of products or at stores where healthy foods are less likely to be available (Chenarides et al., 2015; Handbury et al., 2016). Few studies have tested how consumer behavior would adapt given a change in the food retailing landscape that would remedy these negative outcomes. For example, Cummins et al. (2014) investigate a pilot-study initiative, namely the Pennsylvania Fresh Food Financing Initiative, 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, which 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). This evidence suggests that accessibility 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 relationship between consumer behavior and store choice has been relatively overlooked in understanding where households choose to shop. Especially in areas with high poverty rates, where the proportion of low access and low income population persist over time, consumers are faced with shopping at non-traditional stores, which may augment the negative welfare impacts of living in food deserts. On the other hand, according to classic entry models, marketing strategies and location are choices made by the firm (e.g., Berry, 1992; Bresnahan and Reiss, 1991). Applied to the realm of food retailing, these firm decisions are expected to be related to the food retailing landscape in which the firm operates. Therefore, the process of modeling store choice assumes that the household's decision is intertwined with firms' decisions to enter or exit a market, as well as with heterogeneous consumer demographics. By developing and estimating a store-choice model based on store attributes and household characteristics, our goal is to better understand consumer behavior in underserved areas. We specifically focus on the questions of where and how households shop, and our results will provide insight into the Cummins et al. (2014) “null” result.

This paper makes two major contributions to the literature on store choice and consumer behavior. First, we extend the Distance Metric (DM) demand model of Pinkse, Slade, and Brett (2002) to the problem of consumer store choice to model what behaviors drive consumers’ store-choice decisions, highlighting underserved communities. This technique traditionally 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. Because our store-choice model is based on demand for store attributes (such as relative prices, product assortment measures, store services, and market coverage), it reveals consumer preferences on store types and provide insight into policy prescriptions that attempt to improve food access. Applying this method to consumer store choice provides an alternative to discrete choice methods by allowing the model to reflect multiple store trips, a limitation in existing store-choice models.

A second contribution of this research is that our analysis integrates several rich data sets, namely the IRI Consumer Network scanner data, TDLinx store-attribute data, and the IRI Store scanner data. 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.

The remainder of this paper is organized as follows. In the following section, we summarize relevant literature on the food retailing environment and store choice. Following this overview, we discuss our 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 our case study is focused. Finally, we conclude with an overview of our results and potential policy implications.

Related Literature

The significance of this paper directly relates to efforts made by policymakers who focus on food access, specifically areas deemed as food deserts, which are more likely to remain food deserts over time (Dutko and Ver Ploeg, 2013). Therefore, three major streams of literature frame our research: (i) the nature of the food retailing environment, (ii) the role inadequate access plays in shaping household shopping behavior, and (iii) current methods for modeling store choice.

The Food Retailing Environment

Over the past few decades, the introduction of new food retailer formats, such as supercenters and club stores, has significantly changed the landscape of the food retailing industry. 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 and, as of 2011, became the top grocery retailer in the U.S. and Canada (Kaufman et al., 2000; Supermarket News, 2013). This major triumph over traditional food outlets can be attributed in large part to the innovation of automated distribution and procurement systems (Ellickson, 2004). Although innovation brings with it sunk costs, the investment in new technology for large chain stores means reduced costs in the long-run due to better tracking mechanisms of their inventory, as well as the expectation that stores could offer more products to their patrons.

A number of research studies examine 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 (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, 2006). 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 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 respond not by new firms entering the market to "fill in" the product assortment needs of the consumer base, but rather 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 and 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 (e.g., Shaked and Sutton 1987; Sutton 1991; Ellickson 2007; Ellickson and Grieco 2013; Bonanno, et al. 2012). 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 economies. 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 (e.g., Ellickson, 2007; Bonanno, 2012).

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, for example (Handbury et al., 2015; Ver Ploeg, 2010). Numerous studies agree that supermarkets are less prevalent in poorer areas, while fast-food restaurants appear in more concentrated numbers (Alwitt and Donley, 1997; Moore and 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 (USDA, 1997). Research shows that income has a statistically significant positive effect on fruit and vegetable purchases as well as average store size (Dunkley et al. 2008; Kyureghian et al.

2013). The ability to access larger food stores requires higher transportation costs, which presents hardships for households living in poor access areas who do not have the transportation means to drive to the nearest supermarket.[1] Given the higher search costs they face, 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 and Kaufman, 2003). Households that are better-off may reside in low-income areas; however, these households are more likely to own a car, so traveling to a supermarket outside of their immediate neighborhoods is not considered outside of their means, and are therefore able to escape the food desert in which they live (Ver Ploeg, 2010).

From the researcher's perspective, food access issues are a multidimensional problem as 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 have suggested that it might be an issue of ease of access (Lee, 2012). Handbury et al. (2015) make two findings that support this idea by examining the underlying differences in behavior among lower income and lower educated households. First, they observe that households with lower income and education purchase less healthful foods. 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.

Household Store Choice

The line of literature on store choice is extensive. 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 (e.g., Bell, Ho, and Tang, 1998; Bell and Lattin, 1998; Briesch, Chintagunta, and Fox, 2004; Hoch, Dreze, and Purk, 1994; Sinha, 2000). 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

[1] In general, big box stores, such as Wal-Mart, are located in rural 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 (Holmes, 2011; Grieco and Ellickson, 2013).

best fits their needs. 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 and higher fixed costs, 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. Briesh et al. (2009) 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.

The relevance of calling attention to these studies is the focus on the 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 that arise from estimating demand if consumers face many choices by projecting the stores themselves onto characteristics space (McFadden, 1973). However, these attribute-based modeling approaches still have several limitations.

First, the independence of irrelevant alternatives (IIA) assumption suggests that substitution between two stores is proportional to the shares, rather than a function of attribute proximity (Pofahl and 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 store selection, the IIA assumption may not be realistic. Taylor and Villas-Boas (2016) rely on the mixed logit model that addresses this limitation, accounting for store characteristics as well as household demographics, yielding more realistic substitution patterns.

Nonetheless, within the discrete choice framework, households are restricted to make a decision between two alternative stores. An assumption within these models, and a second limitation, is that consumers may only choose the store that gives the highest utility. While this assumption may be overlooked in the product-choice world, 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] The inability for discrete choice models to account for multiple trips is a strong limitation and ultimately motivates our decision to use an expenditure based demand model.

Store Choice Model

We adopt a linear approximation of the Almost Ideal Demand System (LA/AIDS) of Deaton and Muellbauer (1980) and incorporate the DM method of Pinkse, Slade, and Brett's (hereafter "PSB") (2002) into this framework.[3] Applying this method to consumer store choice provides an alternative to discrete choice methods by allowing the model to reflect multiple store trips, a limitation in existing store-choice models. This estimation specification is supported by Rojas and Peterson (2008), Rojas (2008), and Pofahl and Richards (2009), who apply the DM method to brand-choice models. Let i denote the household, j denote the set of stores, and t denote the month. Therefore, the expenditure share function for a household i shopping at store j in month t would resemble:

$$(1) \qquad w_{ijt} = a_{ij} + \sum_j \gamma_{jk} \log p_{kt} + \beta_{ij} \log\{x_{it}/P_{it}^L\}$$

where w_{ijt} represents the expenditure share for household i's total food purchases at store j in month t, p_{kt} represents the "price" 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 $\log P_{it}$, 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_{ij}^0 is store j's base share for household i with $w_{ij}^0 \equiv T^{-1} \sum_{t=1}^{T} w_{ijt}$ where $t \in (1, \ldots, T)$ 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. Each store share equation represents the share of total food expenditure a household allocates to the stores within their choice set.

[2] Smith (2004) acknowledges this behavior in shopping frequency by calling attention to households' primary shopping trips and secondary shopping trips.
[3] Further development of this research will frame the DM method within the Exact Affine Stone Index (EASI) demand system framework (Lewbel and Pendakur, 2009). The attractive properties and interpretations of EASI model will provide an additional benefit to the store choice problem.

Rather than estimate a demand system of J-1 equations and $J(J$-1)/2 cross-price parameters, we specify the cross-price coefficients γ_{jk} as a function of distance measures between stores j and k (PSB, 2002). These distances (δ_{jk}) are measured in terms of 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}). For example, a discrete attribute indicates that two stores share the same attribute and are considered neighbors in attribute space, while a continuous attribute represents a characteristic that all stores possess yet can vary among them, such as the level of product assortment. Although these two terms both describe distance measures, δ_{jk}^{d} acts as a local measure of closeness (either it takes a value of one if the two stores share the same attribute or zero otherwise), whereas δ_{jk}^{c} acts as a global measure of closeness (i.e., even if stores are very dissimilar, the value of this term will be small by non-zero). 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, 2010).

We write our local measure of closeness, δ_{jk}^{d}, as

$$(2) \qquad \delta_{jk}^{d} = \begin{cases} 1 \; if \; \left|z_j^d - z_k^d\right| = 0 \\ 0 \; if \; \left|z_j^d - z_k^d\right| = 1 \end{cases}$$

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

$$(3) \qquad \delta_{jk}^{c} = \frac{1}{1+2\sqrt{\sum_l (z_j^c - z_k^c)^2}}$$

where z_j^d is a discrete attribute of store j and z_j^c is a continuous attribute of store j.

Using the distance measures $\delta_{jk}^{d}$and δ_{jk}^{c}, the cross-price parameter can be written as $\gamma_{jk} \log p_{kt} = \sum_{c=1}^{C}\left(\lambda_j^c \sum_{k\neq j}^{J} \delta_{jk}^{c} \log p_{kt}\right) + \sum_{d=1}^{D}\left(\lambda_j^d \sum_{k\neq j}^{J} \delta_{jk}^{d} \log p_{kt}\right)$. In addition, the constant term a_{ij}, the own-price coefficient γ_{jj}, and the coefficient on the price index β_{ij} may also be written as functions of household i's demographics and store j's attributes, such that $a_{ij} = a_0 +$

$\sum_{l=1}^{L} a_{jl} z_{jl}^{a} + \sum_{d=1}^{D} \varphi_i h_{id}$, $\gamma_{jj} = \gamma_0 + \sum_{m=1}^{M} \gamma_m z_{jm}^{\gamma}$, and $\beta_{ij} = \beta_0 + \sum_{n=1}^{N} \beta_n z_{jn}^{\beta}$, where store j's characteristics are represented by z_{jl}^{a}, z_{jm}^{γ}, and z_{jn}^{β}, and household i's characteristics are represented by h_{id}. By construction, the characteristics $z^{\theta}, \theta \in (a, \gamma, \beta)$, and h are each be represented separately. Therefore, combining these elements, we get the following estimable equation:

(4)
$$w_{ijt} = a_0 + \sum_{l=1}^{L} a_{jl} z_{jl}^{a} + \sum_{d=1}^{D} \varphi_i h_{id} + \left(\gamma_0 + \sum_{m=1}^{M} \gamma_m z_{jm}^{\gamma}\right) \log p_{jt} + \sum_{c=1}^{C} \left(\lambda_j^c \sum_{k \neq j}^{J} \delta_{jk}^{c} \log p_{kt}\right) + \sum_{d=1}^{D} \left(\lambda_j^d \sum_{k \neq j}^{J} \delta_{jk}^{d} \log p_{kt}\right) + \left(\beta_0 + \sum_{n=1}^{N} \beta_n z_{jn}^{\beta}\right) \log\{x_{it}/P_{it}^{L}\} + \varepsilon_{ijt}$$

According to the DM method, stores in a consumer's choice set are defined by the bundle of their attributes, with each store viewed as a unique bundle. Substitutability between stores is modeled as a function of the relative distance between the stores along several attribute-space dimensions (e.g., assortment, service levels, and relative prices), as well as geographical distance.[4] While the DM method has been used previously to model brand choice, this paper is the first to use it to investigate store choice.

The motivation behind using the DM method for estimating consumer store choice is supported by three major benefits. The first major benefit of using the DM method is its ability to accommodate a multi-stage budgeting assumption. Current models of consumer store choice that assume a discrete choice framework do not adopt a multi-stage budgeting approach. Although this method has traditionally been applied to consumer demand for differentiated products to solve dimensionality issues associated with a large number of products (e.g., Hausman, Leonard, and Zona, 1994; Cotterill and Samson, 2002), we apply this framework to consumer store choice. In the same way that households distribute their income into broad groups, such as food and clothing, households can also choose to distribute their income within a specific subcategory. In the context of store choice, if we assume that households have a food-specific budget, the first stage would represent the amount of discretionary income households allocate to food purchases

[4] At this point, we will be omitting distance between stores from our model and will save this for a future iteration.

in general (i.e., food dollars), and the second stage would correspond to the portion of food dollars households choose allocate to each store. Using a multi-stage budgeting approach supports the idea that households simultaneously decide where to shop, as opposed to assuming households act according to a nested decision-making process.

Second, the DM method can adequately manage the use of micro-level data, whereby a wide array of stores and their attributes can be mapped into an attribute-space matrix that describes the closeness of food stores within store-attribute space. Our datasets provide key information about (i) consumer demographics, (ii) fixed costs (e.g., distance measures) and (iii) variable costs (e.g., store characteristics, store services) of shopping, which can be integrated directly into the DM method to represent the true costs of shopping.

While discrete choice methods for modeling store choice can accommodate each of the aforementioned benefits, the DM method has one principal advantage over discrete choice models, such as the computationally intense random-coefficient mixed logit and the Berry, Levinsohn, and Pakes (2004) method. The main advantage of the DM method is not in its ease of use; rather, the main benefit of this technique allows for consumers' preferences to purchase more than one variety and make multiple purchases of the same product are able to be modeled (i.e., multiple discreteness). The analogue to store choice is that the DM method is able to consider multiple shopping trips to the same and different stores within a given time frame. The micro-level data shows that consumers make purchases at multiple stores over the course of a month or even a week, so limiting the model to only consider one shopping trip at a time is a strict assumption. In this way, the DM method is capable of characterizing more realistic shopping behavior.

Data and Summary Statistics

Consumer Network Panel and Choice Set

We begin our analysis with the Information Resources, Inc. (IRI) household-based store scanner data, the Consumer Network Panel (CNP), for the year 2012.[5] These data document food purchase transactions at a number of food retail outlets that span the U.S. These households are

[5] 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. Analysis for additional years following 2012 will be extended for future research. For more information, see Sweitzer et al. (2016).

taken from the National Consumer Panel, a joint effort between IRI and Nielsen (Sweitzer et al., 2016). Households that are considered in our analysis come from the static panel, which is comprised of households who provide sufficient purchase data. Each shopping trip is linked to the retail chain at which a household shopped. It is important to note that information on the specific location of these retail outlets is not known; only the retail chain (Sweitzer et al., 2016). For example, if a household shops at Supermarket A in market m and records purchases for that store, we observe only that the household shops at the Retail Chain A, of which Supermarket A is but one location. In our analysis, we refer to "stores" as the retail chain, rather than the unique store location.

To construct the choice set of stores for our analysis, we use the demographic data available from IRI and link it with the transactions data according to the household ID. This demographic information includes various household-specific codes and variables, such as race, ethnicity, household composition, and some location information. For this analysis, we focus on Philadelphia County[6] (FIPS code 42101). According to the USDA's Food Access Research Atlas, 37% 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). Our intention is to expand this analysis to incorporate the entire market area (as defined by IRI to be representative of the region encompassing Philadelphia and surrounding counties).[7]

Once we link households' demographic information with the transaction data and identify the households who reside in Philadelphia County, we use information available from the 2012 TDLinx Store Characteristics data to identify a set of retail-level attribute variables. The information available in TDLinx identifies specific store locations as well as information about store sales, amenities, and square footage. Using information about each TDLinx-store's parent company, we are able to match 26 food retail outlets[8] in Philadelphia County with the retail chains where households report food transactions in the CNP. Our final choice set of retailers in Philadelphia County is comprised of 18 food retail chains. These 18 stores capture 62% 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

[6] We chose Philadelphia so that our results can provide context for the Cummins et al. (2014) result.
[7] Our motivation for expanding to a wider region (MarketID) is so we can use the market-level projection factors (survey weights) to obtain representative estimates of household shopping behavior. This step is currently missing from our analysis.
[8] We specifically only look at the following channels: Grocery, Convenience, Mass Merchandisers, and Dollar Stores.

greater than zero, and (3) monthly store price information for that store is complete (see section "Store Price" for more information). The composition of food retail chains in the choice set is made of up eleven grocery stores, two convenience stores, three dollar stores, and two mass merchandisers.

Household-level store expenditure shares are constructed in two ways. The numerator, x_{ijt}, is the total food expenditure by household i at store j in month t. We use two methods for the denominator, or base expenditure. The first method uses the total expenditure a household spends across all stores, including the expenditure at stores outside of the choice set. Using this first method, the sum of the shares sum to a value less than or equal to one, and allows for an outside option. The second method accounts only for the expenditure at the stores in our choice set, so the sum of the shares equals one.

A closer look at the shopping frequency patterns of households in Philadelphia County during the year 2012 shows the following statistics (n=210 households). First, the average number of trips to any food store made in a month is just under seven, while the average number of unique food retail chains visited within a month is over three. Roughly 84% of households, on average, visited more than one unique store in a given month. Excluding stores outside of our choice set, on average, the primary store receives approximately 60% of the household's monthly food expenditure, followed by a second and third store totaling close to 30%. Subsequent stores account for the remaining 10%. Similar trends are present when we look at a household's monthly expenditure across all stores. On average, the household's primary store receives approximately 54% of the household's food expenditure, while the second and third stores receive 35%. The remainder is spread evenly across other retail outlets.

Table 1 summarizes select household demographic characteristics and expenditure shares for the 210 households in the Philadelphia County. In general, the average household size (HHSIZE) for our sample is just over two, with a median annual income (MEDINC) of $50,000. The sample is older, with the average age of the household head at sixty years old (AGEHEAD) and roughly 55% of the household race is identified as Caucasian. Only ten percent of the households report having achieved a college degree (COLLEGE).

Store Price

Since price enters the demand expenditure share equation, we construct a market basket price that reflects the average weighted price for a certain basket of goods at store j in month t. To generate this basket price, we begin by selecting a non-random sample of products that are observed purchases in the CNP at each store in our choice set during 2012. IRI organizes UPCs according to a specific hierarchy that resembles a grocery store environment (Figure 1). According to this hierarchy, we select four categories[9] with the most transactions across all stores, and construct our market basket price by generating a unit price for each category, weighted by the share of expenditure within each of the four categories over the total sales in store j during month t.[10] First, we create an average price per unit volume for each of the four categories:

$$Average\ Price\ Per\ Unit\ Volume_{jtg} = \frac{sales_{jtg}}{volume_{jtg}}$$

where $sales_{jtg} = \sum_{i=1}^{N} DollarsPaid_{ijt} + Coupons_{ijt} \mid p \in G$ is the total sales[11] for product p purchased in month t at store j given product p is contained in category G, and $volume_{jtc} = \sum_{i=1}^{N} Quantity_{ijt} * tot\_volume_{ijt} \mid p \in G$ is the total volume moved for product p purchased in month t at store j given product p is contained in category G. The $average\ price\ per\ unit\ volume$ can be interpreted as the weighted average price (in cents) per one ounce of a product sold in category G during month t at store j. The ultimate price we assign to each store-month combination (p_{kt}) is calculated as follows:

$$(5) \qquad p_{kt} = \sum_{G=1}^{4} \left[Average\ Price\ Per\ Unit\ Volume_{jtc} * \frac{sales_{jtg}}{\sum_C sales_{jtc}} \right]$$

which is the sum of the $average\ price\ per\ unit\ volume$ weighted by the share of $sales_{jtg}$ over the total monthly sales in store j, $\sum_C sales_{jtc}$, where C includes all categories present in store j

---

[9] $G \in \{Carbonated\ Beverages, Cookies, Salty\ Snacks,\ Milk\}$

[10] Additional selection criteria include that at least one item from each of the four categories was sold at each store j during month t in 2012. As mentioned in the previous section, due to missing price information based on this criteria, some stores were dropped from the list of matched TDLinx-IRI stores.

[11] Sales are scaled up by 100, so this amount is measured in cents rather than dollars.

but may not be present in another store k $(j \neq k)$ and $G \in C$ (i.e., the sum of weights is less than or equal to one). Because we use a log transformation of price in our estimation equation and some values of store price (p_{kt}) are less than one, we add a constant term of one to store price before taking logs.

Although constructing this market basket price is relatively rudimentary, we argue that for this first stage analysis it is sufficient for the following reasons. First, we want to choose products that are available at each store in a household's choice set. While we could have used the IRI InfoScan panel to construct a market basket price, certain stores in our choice set would be dropped since certain retailers may not participate in the InfoScan panel of retailers. As such, the CNP offers a more diverse set of retailers that represent where households choose to shop. Second, to get the most variation in our prices across stores and across months, we need to find products that offer such variation. [12] Figure 2 shows the weighted average monthly prices across retail chains. Finally, while these four categories represent a class of food items that are typically deemed as unhealthy, we argue that they capture other elements of product characteristics that are important when constructing a representative market basket. For example, this basket of goods includes both private-label and nationally-branded UPCs, healthy and conventional alternatives, as well as perishable and non-perishable food items.

Store Characteristics

Aside from price, additional store characteristics are incorporated into our model by using the 2012 TDLinx Store Characteristics data. As mentioned earlier, we use information about the variable costs of shopping as well as measures of competition between stores, besides price, to motivate our choice of additional store characteristics that might influence a household to favor one store over another (Bonanno and Lopez, 2009; Smith, 2004; Taylor and Villas-Boas, 2006). These characteristics represent the quality, variety of product offerings, and amenities within the set of food retailers in our analysis. Table 2 summarizes a list of store attributes.

Given the store attributes and price, the DM measures, as well as the own-price and expenditure interaction terms are constructed. We choose two continuous DM measures. To capture the breadth of products sold at each retailer, we construct product assortment (ASSORT), which is measured by the number of unique UPCs carried in the store. In addition, we calculate

[12] Except for one of the two mass merchandisers, we were able to calculate store prices for each month in 2012. Instead of dropping the other mass merchandiser, for the missing month, we imputed the store price as the straight average of the preceding and subsequent months.

the annual sales reported at each retailer over the total market-level sales within Philadelphia County (SHSALES) as a proxy for market coverage. The inverse Euclidean distance measures are computed and stored in "weighting" matrices, where the j, k element in each matrix corresponds to the distance measure between two stores' characteristics. The j, j entries are set to zero. Because the assortment measure (ASSORT) is so large, we scale this value down by dividing by its maximum value to obtain a value between zero and one.

Our measures of discrete distance include the presence of a pharmacy (PHARM) and identifiers for channel type other than grocery (CONV, DOLLAR, MASSMERCH). The discrete measures can be interpreted as the substitutability between stores of the same channel and with the same amenities, given changes in price. Discrete weighting matrices are constructed such that each j, k element is equal to one if stores j and k are of the same channel or both have a pharmacy, otherwise the value is zero.[13]

Finally, in addition to channel type (CONV, DOLLAR, MASSMERCH), we use information about the total square footage at each retail chain to construct an average square foot (AVSQFT) measure.[14] These variables appear in our model specification as demand shifters, or as interactions with either own-price (LNP) or expenditure (LNEXP). Zip code and monthly fixed effects are controlled for in the estimation.

Estimation Concerns

While the use of the DM/LA-AIDS model may prove to be a tractable and realistic technique to model store choice, we note three estimation and empirical concerns before discussing our results: (1) censoring, (2) endogeneity of store characteristics and prices, (3) and market coverage.

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 our choice set to 18 stores, a disproportionate number of zero expenditure shares exists, as we observe households shopping at nine stores at most in a given month. In some cases, households do not shop at any store during a month, and therefore we observe a total

[13] For future iterations, including a discrete measure of EDLP or Hi-Lo pricing strategy will be considered.
[14] Average square feet (AVSQFT) is also scaled down by dividing the value by the maximum average square feet of the stores in the choice set, yielding a value between zero and one.

monthly food expenditure of x_{it} equal to zero. The number of households who do not make any purchases at stores in our choice set varies anywhere between 10% and 20% in a given month. While the use of traditional 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 so that we may estimate store expenditure shares using a Tobit model (Li, Jaenicke, Anekwe, 2013; Rojas and Peterson, 2008). Therefore, the ability of the DM method to easily accommodate the censored nature of the data is a major benefit.[15]

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 w_{ijt} is assumed to be equal to the latent share w_{ijt}^* whenever the latent share is greater than zero (Tobin, 1958), such that:

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

where

(6)
$$w_{ijt}^* = a_0 + \textstyle\sum_{l=1}^{L} a_{jl} z_{jl}^a + \sum_{d=1}^{D} \varphi_i h_{id} + \left(\gamma_0 + \sum_{m=1}^{M} \gamma_m z_{jm}^\gamma\right) \log p_{jt}$$
$$+ \sum_{c=1}^{C} \left(\lambda_j^c \sum_{k \neq j}^{J} \delta_{jk}^c \log p_{kt} \right) + \sum_{d=1}^{D} \left(\lambda_j^d \sum_{k \neq j}^{J} \delta_{jk}^d \log p_{kt} \right)$$
$$+ \left(\beta_0 + \sum_{n=1}^{N} \beta_n z_{jn}^\beta \right) \log\{x_{it}/P_{it}^L\} + \varepsilon_{ijt}$$

and $\varepsilon_{ijt} \sim N(0, \sigma^2)$. The empirical results we present in this paper show results using OLS as well as a Tobit model as shown here in equation (6).

The second estimation concern is price endogeneity, or even potential endogeneity of other variables. 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-attributes, such as product availability. Using price as an example, the error term, ε_{ijt}, in equation (4) represents information other than price, and unobservable to the econometrician such as taste preferences, that cannot be quantified by our data but have 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

[15] Other common econometric techniques exist that can offer alternative approaches to the censorship issue, such as Shonkwiler and Yen's (1999) consistent two-step estimation procedure, and can be applied using the DM method.

equation (4) 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. Future versions of this paper will use instruments to account for potential endogeneity. However, now we use the observed price and hope that the weighted average price across four major store categories is not overly correlated with error term. We also use household-level fixed effects to control for other location-specific endogeneity (i.e., recognizing that household's location and store location are endogenous) (Currie et al., 2010; Taylor and Villas-Boas, 2016).

Our third concern relates to the limitations due to the construction of the available data. IRI's CNP covers household transactions at stores that carry UPC-coded products. Therefore, information on purchases at food retail outlets or venues where items do not have UPCs, namely farmers markets and food away from home, are not available. Although our analysis fails to capture the full set of possible food sources across the households' food retail environment, what we do observe are all food retailers that sell UPC-coded items.[16] Among those food retailers, we choose a subset of food retailers that receive at least 62% of all purchases recorded in the CNP transactions data, which we believe to be a sizeable portion of total market transactions.

Empirical Results

We estimate equation (4) via OLS and equation (6) via Tobit and our results are presented in Table 3. We refer to this specification as the full model, where our set of demographics (h_{id}) includes household size, median income, age of the head of the household, as well as binary variables for race, employment, ethnicity, and education. Average square feet is used to shift the intercept (z_j^{α}), while different store characteristics shift the own-price (z_j^{γ}) and expenditure (z_j^{β}) parameters. In addition to the full model, we include sensitivity analyses using three alternative specifications and these results are presented in Tables 4 and 5. Since multicollinearity may be a concern due to the number of demographics, store characteristics, and distance measures we include in our model, we measure the Variance Inflation Factor (VIF) and find that the average VIF is 3.02, suggesting the degree of multicollinearity is trivial (O'Brien, 2007).

[16] 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.

Estimated Coefficients

The results from the OLS regression are presented in Table 3, column (1) and the results of the Tobit estimation in column (2). For both models, signs and significance are consistent, while magnitudes of the estimated coefficients are somewhat larger when we account for censoring in the Tobit model. In the discussion that follows, we focus on the estimated parameters obtained from the Tobit model.

The first general observation we make is the sign and significance of the coefficient on own-price (LNP). Despite the suspected presence of endogeneity in our model, these results are consistent with theory, indicating that as the price of the store increases, the expenditure share decreases. The coefficient on the interaction term between price and square footage (LNPxAVSQFT) is also negative and significant. Adding these two numbers together we get a value of -0.388, implying that, everything else constant, households are more price sensitive to stores with higher square footage. According to this result, we might infer that as price increases households are less loyal to larger stores.

Next, we consider the parameter estimates on the distance metric terms. These coefficients can be interpreted as the households' response to price changes as stores become more competitive and similar in attributes. The estimated coefficients associated with δ_{jk}^{c} are negative and significant, indicating that households respond to price increases by switching away from stores with similar assortments or similar market coverage. The estimated coefficients associated with δ_{jk}^{d} are also mostly negative and significant, suggesting that households are more likely to switch to stores outside of the same channel, given a price increase. However, based on this specification, it appears that the strongest determinant of substitution is the presence of a pharmacy, where we see a positive and significant sign on DM_PHARM. This result suggests that as price increases, households tend to switch to stores where a pharmacy is also present.

We observe the coefficient on expenditure (LNEXP) is positive, implying that as a household's budget increases, the share of expenditure at that store also increases. Coefficients on the interactions with LNEXP are negative and significant, so adding any of the channel interactions to the coefficient on LNEXP shows a dampening effect, suggesting that households are less likely to remain shopping within the convenience, mass merchandiser, or dollar store channels given expenditure increases. The intercept shifters (h_{id} and z_{j}^{a}) are all non-significant.

It should be made clear that the results of our analysis are preliminary, so at this stage, placing these results within the context of existing literation is premature. Despite estimating alternative specifications whereby certain variables from the full model are excluded or replaced by other store characteristics, the coefficient estimates on the intercept shifters remain non-significant. To assess what role these variables, and others, play in understanding consumer store choice requires future evaluation; therefore, we will continue to investigate these results and consider alternative specifications as we expand our analysis.[17]

Conclusion

As the food retailing environment continues to evolve, our research aims to expound on the relationship between consumer behavior, shopping decisions, and food access. Once this relationship has been established, we can begin to answer a number of straightforward, yet currently unanswered questions, such as: Are people of certain demographics attracted to particular store attributes? How do shoppers substitute across store attributes? Do store attributes and consumer demographics play different roles in food deserts, as compared to other areas? Despite the importance of this topic and questions such as these, little or no research exists that documents how the distribution of consumer types and geographic patterns are associated with store choice. Our paper will inform subsequent research that uses policy-informed scenarios to simulate changes in the food retail landscape to investigate welfare changes for consumers and food retailers.

Methodologically, 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 described to create a more flexible model of store selection within different markets across the U.S. The use of the DM method fits directly within this conceptual framework. Food retailers decide their marketing strategies and how they compete in order to attract potential customers. 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 markups), as well as geographical distance. Via the DM method, households are able choose the store(s) that possess the most

[17] Elasticity estimates are not available at this time.

desirable set of attributes and substitute between stores that are relatively closer in proximity across the set of characteristics.

By combining several rich data sets – Nielsen's TDLinx store-attribute data, the IRI Consumer Network scanner data, and the IRI Store scanner data – our analysis supports an ongoing effort to examine 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 our methods and results generate significant interest and discussion with applications in marketing, health, and food policy.

Table 1. Household Summary Statistics

Variable	Description	Mean	Std. Dev.
Purchase			
SH1	Household expenditure share for each store, $j \in J$	0.322	(0.297)
SH2	Household expenditure share for each store, including $j \notin J$	0.434	(0.363)
Demographics			
Continuous			
HHSIZE	Number of individuals in the household	2.40	(1.50)
MEDINC	Household median annual income ($00s)	$512.59	(322.33)
AGEHEAD	Max age of household head	60.50	(12.07)
Discrete		*Frequency of one*	
WHITE	Race is white for the household	0.552	
DINKS	Double income household, no children	0.143	
HISP	Hispanic origin for the household	0.043	
COLLEGE	Maximum educational attainment for the household is college	0.100	

Table 2. Store Summary Statistics

Variable	Description	Mean	Std. Dev.
Continuous			
STPRICE	Average price (per unit), in cents	0.272	(11.894)
ASSORT	Number of unique UPCs sold in store *j*	2169.833	(2785.613)
AVSQFT	Average square footage of store *j*	29.355	(23.852)
SHSALES	Share of annual volume over total volume in FIPS 42101 of store *j*	0.0463	(0.040)
Discrete		*Frequency of one*	
CONV	Store format is categorized as conventional convenience store	0.111	
DOLLAR	Store format is categorized as dollar store	0.111	
MASSMERCH	Store format is categorized as mass merchandiser or supercenter	0.111	
PHARM	Presence of a pharmacy at retailer	0.389	

Figure 1. Product Information

UPC	$\subset$ Product ($p \in P$)	$\subset$ Category ($c \in C$)	$\subset$ Aisle ($a \in A$)	$\subset$ Department ($d \in D$)
e.g., Sprite Lemon Line Soda, Regular 144 oz.	e.g., Regular Soft Drinks	e.g., Carbonated Beverages	e.g., Aisle-Carbonated Soft Drinks	e.g., Dept-Beverages

Figure 2. Price Variation

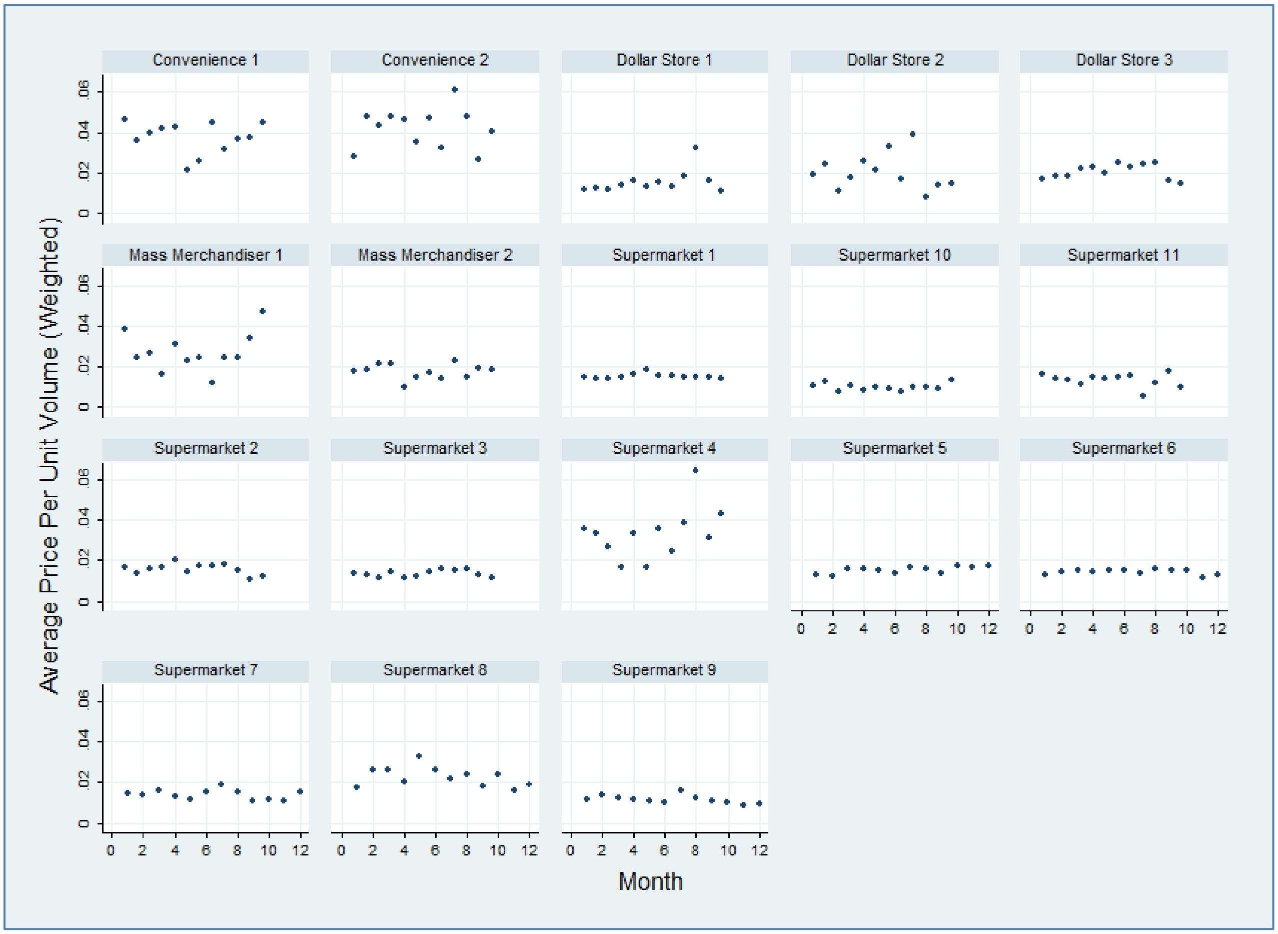

Table 3. Estimation Results (Preliminary)

Dependent variable:	(1)		(2)	
Expenditure share (w_{ijt})	OLS-Full		Tobit-Full	
Demographics				
HHsize	-0.000269	(0.0007)	0.00690	(0.0051)
MedInc	-0.000000137	(0.0000)	-0.0000390	(0.0000)
AgeHead	-0.0000126	(0.0001)	-0.000322	(0.0006)
White	-0.000895	(0.0025)	0.00187	(0.0193)
Dinks	0.000159	(0.0029)	-0.0212	(0.0224)
College	-0.000384	(0.0032)	-0.0139	(0.0261)
Store Shifters				
AVSQFT	-0.0469***	(0.0118)	0.171	(0.1006)
Own-Price				
LNP	-0.0139***	(0.0035)	-0.152***	(0.0314)
LNPxAVSQFT	-0.0216*	(0.0103)	-0.236*	(0.0929)
Distance Matrices				
DM_assort	-0.00914***	(0.0002)	-0.0270***	(0.0010)
DM_shsales	-0.00290***	(0.0005)	-0.0571***	(0.0041)
DM_pharm	0.00345***	(0.0002)	0.00775***	(0.0014)
DM_conv	-0.000136	(0.0001)	-0.00482*	(0.0021)
DM_massmerch	-0.00176***	(0.0001)	-0.00564*	(0.0024)
DM_dollar	-0.000646***	(0.0001)	-0.0118***	(0.0017)
Expenditure				
LNEXP	0.0115***	(0.0004)	0.230***	(0.0098)
LNEXPxCONV	-0.0110***	(0.0011)	-0.122***	(0.0226)
LNEXPxMM	-0.00799***	(0.0011)	-0.0585*	(0.0257)
LNEXPxDOL	-0.0104***	(0.0011)	-0.141***	(0.0186)
_cons	0.877***	(0.0395)	5.471***	(0.4114)
sigma			0.696***	(0.0082)
N	45360		45360	
adj. R^2	0.170		0.204	
df_m	73		73	
df_r	45286		45287	

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Zip code and monthly fixed effects included in all models.

Table 4. OLS Estimation – Sensitivity Analysis

Dependent variable: Expenditure share (w_{ijt})	(1) OLS-Full		(1a) OLS-No EXP Interactions		(1b) OLS-No LNP Interactions		(1c) OLS-No Z Shifters	
Demographics								
HHsize	-0.000269	(0.0007)	-0.000269	(0.0007)	-0.000269	(0.0007)	-0.000269	(0.0007)
MedInc	-0.000000137	(0.0000)	-0.000000137	(0.0000)	-0.000000137	(0.0000)	-0.000000137	(0.0000)
AgeHead	-0.0000126	(0.0001)	-0.0000126	(0.0001)	-0.0000126	(0.0001)	-0.0000126	(0.0001)
White	-0.000895	(0.0025)	-0.000895	(0.0025)	-0.000895	(0.0025)	-0.000895	(0.0025)
Dinks	0.000159	(0.0029)	0.000159	(0.0029)	0.000159	(0.0029)	0.000159	(0.0029)
College	-0.000384	(0.0032)	-0.000384	(0.0032)	-0.000384	(0.0032)	-0.000384	(0.0032)
Store Shifters								
AVSQFT	-0.0469***	(0.0118)	-0.0469***	(0.0118)	-0.0642***	(0.0085)	--	--
Own-Price								
LNP	-0.0139***	(0.0035)	-0.0110**	(0.0035)	-0.0193***	(0.0023)	-0.00572*	(0.0028)
LNPxAVSQFT	-0.0216*	(0.0103)	-0.0220*	(0.0103)	--	--	-0.0502***	(0.0074)
Distance Matrices								
DM_assort	-0.00914***	(0.0002)	-0.00913***	(0.0002)	-0.00913***	(0.0002)	-0.00909***	(0.0002)
DM_shsales	-0.00290***	(0.0005)	-0.00291***	(0.0005)	-0.00314***	(0.0005)	-0.00252***	(0.0005)
DM_pharm	0.00345***	(0.0002)	0.00343***	(0.0002)	0.00351***	(0.0002)	0.00360***	(0.0002)
DM_conv	-0.000136	(0.0001)	0.000741***	(0.0001)	-0.000166	(0.0001)	-0.000161	(0.0001)
DM_massmerch	-0.00176***	(0.0001)	-0.00113***	(0.0001)	-0.00171***	(0.0001)	-0.00167***	(0.0001)
DM_dollar	-0.000646***	(0.0001)	0.000159***	(0.0000)	-0.000656***	(0.0001)	-0.000665***	(0.0001)
Expenditure								
LNEXP	0.0115***	(0.0004)	0.00826***	(0.0004)	0.0115***	(0.0004)	0.0115***	(0.0004)
LNEXPxCONV	-0.0110***	(0.0011)	--	--	-0.0110***	(0.0011)	-0.0110***	(0.0011)
LNEXPxMM	-0.00799***	(0.0011)	--	--	-0.00802***	(0.0011)	-0.00799***	(0.0011)
LNEXPxDOL	-0.0104***	(0.0011)	--	--	-0.0104***	(0.0011)	-0.0104***	(0.0011)
_cons	0.877***	(0.0395)	0.719***	(0.0378)	0.897***	(0.0384)	0.821***	(0.0368)
N	45360		45360		45360		45360	
adj. R^2	0.170		0.166		0.169		0.169	
df_m	73		70		72		72	
df_r	45286		45289		45287		45287	

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Zip code and monthly fixed effects included in all models.

Table 5. Tobit Estimation – Sensitivity Analysis

Dependent variable:	(2)		(2a)		(2b)		(2c)	
Expenditure share (w_{ijt})	Tobit-Full		Tobit-No EXP Interactions		Tobit-No LNP Interactions		Tobit-No Z Shifters	
Demographics								
HHsize	0.00690	(0.0051)	0.00675	(0.0051)	0.00687	(0.0051)	0.00683	(0.0051)
MedInc	-0.0000390	(0.0000)	-0.0000382	(0.0000)	-0.0000386	(0.0000)	-0.0000390	(0.0000)
AgeHead	-0.000322	(0.0006)	-0.000296	(0.0006)	-0.000321	(0.0006)	-0.000328	(0.0006)
White	0.00187	(0.0193)	0.00356	(0.0194)	0.00169	(0.0193)	0.00175	(0.0193)
Dinks	-0.0212	(0.0224)	-0.0212	(0.0224)	-0.0211	(0.0224)	-0.0211	(0.0224)
College	-0.0139	(0.0261)	-0.0138	(0.0261)	-0.0141	(0.0261)	-0.0139	(0.0261)
Store Shifters								
AVSQFT	0.171	(0.1006)	0.186	(0.1006)	-0.0232	(0.0652)	--	--
Own-Price								
LNP	-0.152***	(0.0314)	-0.111***	(0.0309)	-0.212***	(0.0209)	-0.185***	(0.0248)
LNPxAVSQFT	-0.236*	(0.0929)	-0.251**	(0.0926)	--	--	-0.116	(0.0602)
Distance Matrices								
DM_assort	-0.0270***	(0.0010)	-0.0268***	(0.0010)	-0.0270***	(0.0010)	-0.0271***	(0.0010)
DM_shsales	-0.0571***	(0.0041)	-0.0574***	(0.0041)	-0.0594***	(0.0040)	-0.0581***	(0.0041)
DM_pharm	0.00775***	(0.0014)	0.00758***	(0.0014)	0.00838***	(0.0014)	0.00731***	(0.0013)
DM_conv	-0.00482*	(0.0021)	0.00651***	(0.0005)	-0.00517*	(0.0021)	-0.00484*	(0.0021)
DM_massmerch	-0.00564*	(0.0024)	-0.000291	(0.0006)	-0.00585*	(0.0023)	-0.00616**	(0.0023)
DM_dollar	-0.0118***	(0.0017)	0.000698*	(0.0003)	-0.0120***	(0.0017)	-0.0118***	(0.0017)
Expenditure								
LNEXP	0.230***	(0.0098)	0.203***	(0.0081)	0.231***	(0.0098)	0.230***	(0.0098)
LNEXPxCONV	-0.122***	(0.0226)	--	--	-0.123***	(0.0226)	-0.123***	(0.0226)
LNEXPxMM	-0.0585*	(0.0257)	--	--	-0.0653**	(0.0249)	-0.0612*	(0.0254)
LNEXPxDOL	-0.141***	(0.0186)	--	--	-0.141***	(0.0187)	-0.141***	(0.0186)
_cons	5.471***	(0.4114)	3.398***	(0.3007)	5.726***	(0.3974)	5.666***	(0.3948)
sigma	0.696***	(0.0082)	0.697***	(0.0082)	0.696***	(0.0082)	0.696***	(0.0082)
N	45360		45360		45360		45360	
pseudo R^2	0.204		0.202		0.204		0.204	
df_m	73		70		72		72	
df_r	45287		45290		45288		45288	

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Zip code and monthly fixed effects included in all models.

References

Alwitt, L.F., and T.D. Donley. 1997. "Retail Stores in Poor Urban Neighborhoods." *Journal of Consumer Affairs* 31 (1) (June): 139–164.

Bell, D.R., T.H. Ho, and C.S. Tang. 1998. "Determining Where to Shop: Fixed and Variable Costs of Shopping." *Journal of Marketing Research* 35: 352–369.

Bell, D.R., and J.M. Lattin. 1998. "Shopping Behavior and Consumer Preference for Store Price Format: Why 'Large Basket' Shoppers Prefer EDLP." *Marketing Science* 17 (1): 66–88.

Berry, S.T. 1992. "Estimation of a Model of Entry in the Airline Industry." *Econometrica* 60 (4): 889–917.

Bonanno, A. 2012. "Food Deserts: Demand, Supply, and Economic Theory." *Choices*.

Bonanno, A., L. Chenarides, and S.J. Goetz. 2012. "Limited Food Access as an Equilibrium Outcome: An Empirical Analysis". *2012 AAEA/EAAE Food Environment Symposium, May 30-31, Boston, MA.* (No. 123196). Agricultural and Applied Economics Association.

Bonanno, A., and R.A. Lopez. 2009. "Competition Effects of Supermarket Services." *American Journal of Agricultural Economics* 91 (3) (August): 555–568.

Bresnahan, T.F., and P.C. Reiss. 1991. "Entry and Competition in Concentrated Markets." *Journal of Political Economy* 99 (5): 977–1009.

Briesch, R.A., P.K. Chintagunta, and E.J. Fox. 2004. "Assortment, Price, and Convenience: Modeling the Determinants of Store Choice." Working Paper.

Briesch, R.A., P.K. Chintagunta, and E.J. Fox. 2009. "How Does Assortment Affect Grocery Store Choice?" *Journal of Marketing Research* 46 (April): 1–49.

Chenarides, L., E.C. Jaenicke, and R.J. Volpe. 2015. "Patterns of Product Assortment and Price-Cost Margins across the Food Retailing Landscape." *143rd Joint EAAE/AAEA Seminar, March 25-27, 2015, Naples, Italy,* (No. 202710). European Association of Agricultural Economists.

Clarke, I. 2000. "Retail Power, Competition, and Local Consumer Choice in the UK Grocery Sector." *European Journal of Marketing* 38 (4): 975–1002.

Cotterill, R.W., and P.O. Samson. 2002. "Estimating a Brand-Level Demand System for American Cheese Products to Evaluate Unilateral and Coordinated Market Power Strategies." *American Journal of Agricultural Economics* 84 (3): 817–823.

Cummins, S., E. Flint, and S.A. Matthews. 2014. “New Neighborhood Grocery Store Increased Awareness of Food Access but Did Not Alter Dietary Habits or Obesity.” *Health Affairs* 33 (2) (February): 283–91.

Currie, J., S. DellaVigna, E. Moretti, and V. Pathania. (2010). “The effect of fast food restaurants on obesity and weight gain.” *American Economic Journal: Economic Policy* (August): 32-63.

Deaton, A., and J. Muellbauer. 1980. “An Almost Ideal Demand System.” *The American Economic Review* 70 (3): 312–326.

Dunkley, B., A. Helling, and D.S. Sawicki. 2004. “Accessibility Versus Scale: Examining the Tradeoffs in Grocery Stores.” *Journal of Planning Education and Research* 23 (4) (June): 387–401.

Dutko, P., and M. Ver Ploeg. 2013. “Different measures of food access inform different solutions.” *Amber Waves*, (1).

Ellickson, P.B. 2004. “Supermarkets as a Natural Oligopoly.” *Economic Inquiry*. Duke University.

Ellickson, P.B. 2006. “Quality Competition in Retailing: A Structural Analysis.” *International Journal of Industrial Organization* 24 (May): 521–540.

Ellickson, P.B. 2007. “Does Sutton Apply to Supermarkets?” *The RAND Journal of Economics* 38 (1): 43–59.

Ellickson, P.B., and P.L.E. Grieco. 2013. “Wal-Mart and the Geography of Grocery Retailing.” *Journal of Urban Economics* 75 (May): 1–14.

Handbury, J., I. Rahkovsky, and M. Schnell. 2015. “What drives nutritional disparities? Retail access and food purchases across the socioeconomic spectrum” *National Bureau of Economic Research*. (No. w21126).

Hausman, J., G. Leonard, and J.D. Zona. 1994. “Competitive Analysis with Differentiated Products.” *Annales d’Economie et de Statistique* (34): 159–180.

Hoch, S.J., X. Drèze, and M.E. Purk. 1994. “EDLP, Hi-Lo, and Margin Arithmetic.” *Journal of Marketing* 58 (4): 16–27.

Holmes, T. J. 2011. “The Diffusion of Wal-Mart and Economies of Density.” *Econometrica*, 79 (1): 253-302.

Kaufman, P.R., C.R. Handy, E.W. Mclaughlin, and G.M. Green. 2000. *Understanding the*

Dynamics of Produce Markets: Consumption and Consolidation Grow. (No. 758). DIANE Publishing.

Kyureghian, G., R.M. Nayga, and S. Bhattacharya. 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) (February): 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) (April): 1193–203.

Leibtag, E.S., and P.R. Kaufman. 2003. "Exploring Food Purchase Behavior of Low-Income Households: How Do They Economize?" *Current Issues*. No. 747-07.

Lewbel, A., and K. Pendakur (2009). "Tricks with Hicks: the EASI demand system." *American Economic Review* 99(3): 827-863.

Li, J., E.C. Jaenicke, and T. Anekwe. 2013. "Health-Related Product Attributes and Purchasing Behavior in the Ready-to-Eat Cereal Market: An Application with Household-Level, Censored Data." *AAEA 2013 Annual Meeting*. Washington, DC.

McFadden, D. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior."

Moore, L.V., and A.V. Diez Roux. 2006. "Associations of Neighborhood Characteristics with the Location and Type of Food Stores." *American Journal of Public Health* 96 (2) (February): 325–31.

Moschini, G. (1995). "Units of measurement and the stone index in demand system estimation." *American Journal of Agricultural Economics* 77(1): 63-68.

O'Brien, R. M. (2007). "A caution regarding rules of thumb for variance inflation factors." *Quality & Quantity* 41(5): 673-690.

Pinkse, J., M.E. Slade, and C. Brett. 2002. "Spatial Price Competition: A Semiparametric Approach." *Econometrica* 70 (3): 1111–1153.

Pofahl, G.M., and T.J. Richards. 2009. "Valuation of New Products in Attribute Space." *American Journal of Agricultural Economics* 91 (2) (May): 402–415.

Leszczyc, P.T.P., A. Sinha, and A. Sahgal. 2004. "The effect of multi-purpose shopping on pricing and location strategy for grocery stores." *Journal of Retailing* 80 (2): 85-99.

Rojas, C. 2008. "Price Competition in U.S. Brewing." *The Journal of Industrial Economics* 25 (1): 1–31.

Rojas, C., and E.B. Peterson. 2008. "Demand for Differentiated Products: Price and Advertising

Evidence from the U.S. Beer Market." *International Journal of Industrial Organization* 26 (1) (January): 288–307.

Shaked, A., and J. Sutton. 1987. "Product Differentiation and Industrial Structure." *The Journal of Industrial Economics* 36 (2): 131–146.

Shonkwiler, J.S., and S.T. Yen. 1999. "Two-Step Estimation of a Censored System of Equations." *American Journal of Agricultural Economics* 81 (November): 972–982.

Sweitzer, M., D. Levin, A. Okrent, M.K. Muth, D. Brown, K. Capogrossi, S.A. Karns, P. Siegel, and C. Zhen. 2016. "Understanding IRI Household-Based and Store-Based Scanner Data."

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." *Review of Economic Studies* 71 (1) (January): 235–263.

Sutton, J. 1992. "Sutton's Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration." *The RAND Journal of Economics* 23 (1): 137–152.

Taylor, R., and S.B. Villas-Boas. 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.

Ver Ploeg, M., and P. Dutko. 2013. USDA Economic Research Service-Documentation.

Ver Ploeg, M. (Ed.). 2010. *Access to affordable and nutritious food: measuring and understanding food deserts and their consequences: report to Congress*. DIANE Publishing.

Ver Ploeg, M., V. Breneman, P. Dutko, R. Williams, S. Snyder, C. Dicken, and P. Kaufman. 2012. "Access to Affordable and Nutritious Food: Updated Estimates of Distance to Supermarkets Using 2010 Data." US Department of Agriculture, Economic Research Service.