Welfare Impacts from Store Attribute-Based Policy Interventions in an Urban Setting: An Application to Philadelphia

Lauren Chenarides[*1,2], Edward C. Jaenicke[†2], and Jing Li[‡3]

[1] *Department of Agricultural Economics, Sociology, and Education, Penn State University, University Park, PA 16802*
[2] *Morrison School of Agribusiness, Arizona State University, Mesa, AZ 85212*
[3] *Department of Strategy and International Business, The University of Hong Kong, Hong Kong*

May 2017

DO NOT CITE WITHOUT AUTHORS' PERMISSION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`
[‡] Electronic address: `jingli7@hku.hk`

Abstract

This paper attempts to model both households' underlying preferences for food retailers as well as the retailers' decision on the supply of store attributes. In this paper, we measure the changes to producer and consumer welfare from a policy scenario based on counterfactual changes to retail chains' attributes. To emulate a policy currently being evaluated in Philadelphia and other U.S. metropolitan areas, we estimate the change in equilibrium prices after increasing the assortment of healthy products in small-scale food chains. Using counterfactual analyses, we compare the effectiveness of "attribute-based" policies, such as one based on product assortment, to policies based on adding a new store to a household's choice set.

1 Introduction

Current research has generally documented a number of hardships for households residing in areas with poor food access. Most notably, households living in underserved communities are at a higher risk for diet-related chronic diseases, such as diabetes and heart disease. The SNAP program and related income-assistance programs attempt to address some of the persistent diet-quality and health issues by increasing the disposable income of at-risk households, and thus work via the demand side of the market. A second class of policy mechanisms that looks to indirectly address households' poor food choices that contribute to declined health do so through the means of a supply-side intervention. One prominent policy in this class is a subsidy, often in the form of tax credits, to a larger format supermarket in return for entering the market. In effect, these types of policies provide incentives for firms to enter an area with less favorable demand conditions to spark behavioral changes in consumers who reside in these communities.

Although these supply-side mechanisms may seem intuitive, an emerging line of the literature suggests that supermarket access alone is not necessarily a solution to addressing concerns about diet and nutrition. 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 & Nayga 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 do not explain the differences in diet. Evidence from each of these reports suggests that consumer and strategic retailer behavior play a larger role in appropriately addressing policy-related issues around food access.

Despite the prominence of food access-related issues, very little research exists that examines how consumer behavior would adapt, in an equilibrium setting, to a supply-side intervention, such as the entrance of a new store. In other words, no research exists that investigates or explains the Cummins et al. (2014) result in an equilibrium context. Thus, the primary objective of this paper is to develop an equilibrium model of retailer and consumer behaviors and use it estimate the changes to producer and consumer welfare from policy scenarios based on counterfactual changes to stores' attributes. Using counterfactual analyses, we plan to compare the effectiveness of hypothetical "attribute-based" policies, such as as policies that encourage healthier product offerings, such as fruits and vegetables, at smaller stores. Especially for households residing in underserved communities, the importance of identifying how store attributes are linked to welfare impacts may explain where policy interventions might be most successful (Fitzpatrick & Ver Ploeg 2013, Handbury et al. 2016, USDA 2009).

In this paper, we follow the method of Nevo (2001), Rojas & Peterson (2008), and others and provide a theoretical model that describes both households' underlying preferences for food retail chains as well as the retailers' decisions on the supply of store attributes. Our demand model is constructed from an expenditure-based, censored retailer-choice model that relies on the linear approximated almost ideal demand system ("LA/AIDS") as its starting point. Because we are interested in conducting counterfactual analyses that highlight attribute-based policy scenarios, we incorporate into our framework the Distance Metric ("DM") method of Pinkse et al. (2002), which provides several empirical benefits for estimating retail-level demand. A primary benefit of a retailer-choice model using an AIDS-based demand system is that it allows for multiple store trips. In other words, our motivation for choosing a demand system that is derived from an expenditure function allows us to reflect households' food-at-home budget allocation across multiple retail chains. The multiple-chain shopping patterns of households is widely observed in the consumer-level scanner data, yet it is not easily incorporated into traditional store-choice models based on discrete choice methods without strict independence assumptions. Moreover, because our retailer-choice model incorporates the DM method into the demand specification, household food retailer demand is modeled as a function of attributes, such as relative prices, product assortment measures, and physical geographical distance traveled to a retail chain, thereby revealing consumer preferences on types of retail chains (e.g., supermarket versus convenience) and retailer attributes (e.g., assortment levels), and provides insight into policy prescriptions that attempt to improve food access.

To construct our model of firms' pricing behavior and recover marginal costs, we adapt a structural framework similar to Nevo (2001) and Slade (2004). In general, this method relies on using coefficient estimates from the demand model to compute price-cost margins under alternative specifications of supply conduct. An attractive property of this approach is that price-cost margins can be recovered without observing the actual costs faced by the food retailers. Using store attributes as cost shifters, we can estimate the impact of various store attributes on marginal cost to calculate new marginal costs under different counterfactual scenarios by altering some store attributes. The new marginal costs from the counterfactual policy, together with the demand model, generates a new set of equilibrium prices in the market. Following these steps to solve for new equilibrium prices allows us to accomplish our ultimate goal – to estimate welfare changes resulting from hypothetical policy-based scenarios for households and firms within the Philadelphia-metro area.

In the section that follows, we describe the food retailing landscape and outline specific supply-side policies currently being evaluated in Philadelphia and other U.S. metropolitan areas. After outlining the empirical model, we estimate the equilibrium model and present preliminary results. We use the remainder of the paper to present preliminary welfare estimates and conclude with a discussion of possible improvements to our approach.

2 The Food Retailing Industry and Policy Incentives

A well-established line of research characterizes the nature of the food retailing industry as highly concentrated, as measured by four-firm concentration ratios, and localized, meaning that a firm's market reach is geographically centralized (Richards & Pofahl 2010). Within these localized markets, research that describes the equilibrium structure of the food retailing industry generally classifies food retail chains according to the types of consumers they target. Ellickson (2006) stratifies the competitive behavior among food retailers according to two tiers. The first tier of food retailers are high-quality supermarkets that attract the portion of customers who value a wide assortment of products. He notes that, as the market size grows, existing market leaders will feel pressure to improve the variety of their products, in effect raising endogenous fixed costs and creating barriers for other high-quality firms to enter, which results in a natural oligopoly within the supermarket industry. The remaining portion of the market is serviced by "fringe" stores that target customers who value lower prices and cost savings. Whereas the number of high-quality stores plateaus with increasing market size, these lower-quality, fringe stores choose not to invest in variety-enhancing, endogenous costs and, unlike the high-quality supermarkets, the number of these firms grows as market size increases, yielding a competitive market among the fringe stores. While both high- and low-quality stores exist in the same market, the segmentation between variety-seeking and cost-minimizing consumers impacts the nature of competition among traditional food retailers (Ellickson 2007).

As the landscape of the food retailing industry continues to shift, the level of vertical differentiation extends beyond supermarkets and grocery stores to include other non-traditional food retailers, such as supercenters and club stores. As it becomes profitable for alternative-format stores to sell food items, large supermarkets must now compete with mass merchandisers. On the consumer side, one major trend driven by consumer preferences that supports these shifts in the food retailing landscape is a preference for convenience. This trend can be seen as increasing shares of households' food budgets are being allocated to supercenters. The idea of “one-stop” shopping, where consumers can shop at a single store to make all of their purchases, has been a successful model for food retailers, supporting the idea that firms with the widest selection prevail (Ellickson 2006). In addition, these stores, which generally have robust distribution systems in place, are able to pass these cost savings on to consumers. In effect, these format stores are able to cross over both consumer segments, thereby targeting consumers who value variety as well as those who care only about price.

Given the body of literature that examines the equilibrium structure of the food retailing industry, the question that arises relates to what factors could lead to extreme food landscape outcomes, in particular food deserts. Economic theory suggests that the uneven dispersion of consumer types, coupled with the fixed and variable costs faced by different food retail formats, significantly affects equilibrium outcomes (Ellickson 2007). Therefore, in order to justify entering a market, food retailers must be sure that they can gain a competitive foothold. The level of fixed costs required for food stores to increase their quality level is high, since increasing quality requires additional space and more advanced distribution systems. If high-quality firms want to remain profitable, they must rely on less price-sensitive consumers, so a fewer number of these stores will be located in low-income markets (Bonanno & Lopez 2009). On the other hand, the low-quality stores, which do not invest in the same technologies as supermarkets or large chain retailers that contribute to endogenous fixed costs, are not subject to the same location constraints as their high-quality counterparts.

For households who reside in food deserts, the consequence of this food-retailing equilibrium is an abundance of small-scale, low-quality stores. In markets where access is limited, high-quality retailers do not have an incentive to overcome high fixed costs and therefore choose to locate in markets with more stable demand. As a result, the homogeneity of the food retailing landscape in food deserts appears to compound the hardships faced by these households. For example, the limited selling space, an attribute of fringe stores, is correlated with smaller product assortment of fresh fruits and vegetables (Ver Ploeg et al. 2010). Research examining economizing practices for poor households, a segment of the population that may care about cost savings, has indicated that low-income shoppers value volume discounts and product promotions typically offered at formats like supercenters and mass merchandisers. However, within underserved markets, access to larger food stores often entails higher transportation costs, especially for low-income individuals, and, consequently, these households are often unable to take advantage of the benefits of shopping at larger format stores because even these retailers tend to locate in the suburbs or higher-income areas (Leibtag & Kaufman 2003).

Given the addition of new-format stores to the food retailing landscape and the evidence of non-traditional food stores' effectiveness in targeting both types of consumers, the significance and relevance of quality as a determinant of vertical differentiation may provide deeper insight into understanding and evaluating policy that addresses food access issues. Focusing on the class of supply-side interventions, these types of policies take various forms. At the highest level of financial support, programs, such as the Pennsylvania Fresh Food Financing Initiative, that entice new large-scale supermarkets to enter underserved markets do so by offering tax incentives and lump-sum loans. While the intention of this type of policy is intuitive, placed in context with the endogenous fixed-cost ("EFC") framework of Ellickson (2007), it is not guaranteed that this type of policy will benefit the segment of consumers who need it most. In other words, if this subsidy is given to defray the fixed cost of entry for a high-quality supermarket that, in fact, has differentiated itself to target the variety-seeking consumer, and the demand for a high-quality store does not exist in the market, then economic theory would predict that this particular supply-side policy would not be universally effective. Therefore, result of Cummins et al. (2014) which finds that new supermarket entry in a food desert does not significantly impact consumers' consumption of fruits and vegetables may not seem surprising.[1]

While large-scale, subsidy-based programs may offer other economic benefits, such as

[1] The treatment group in Cummins et al. (2014)'s study received a new 41,000 square-foot supermarket. According to Ellickson (2005), this store may constitute as a high-quality supermarket. On average, the average square-footage of a top six firm is approximately 38,100.

providing new jobs and bolstering community development, if the profitability of the store in the long run cannot be sustained, then this type of program does not appear to be cost-effective. Alternative supply-side policies that have been proposed require less financial public and private financial investment, and at the same time offer comparable positive economic externalities. These types of policies provide incentives to existing stores to stock healthier foods. In Philadelphia, one such program, called the "Healthy Corner Store Initiative," works with food retailers in underserved communities to increase the availability of healthy foods at small-scale grocery and convenience stores.[2] In effect, these policies capitalize on consumers' preferences for convenience and cost savings by targeting the more price-sensitive consumer segment who spend a considerable portion of their food budget at fringe stores. Recalling the EFC framework, this type of policy might be considered a variety-enhancing, but, where Ellickson (2007) distinguishes quality as impacting supermarkets' fixed costs, he notes that small-scale, fringe stores – the stores that qualify for support from the Healthy Corner Store Initiative – are not subject to the same endogenous cost structure. Therefore, enhancing quality (i.e., increasing the variety of products) at a store in the fringe does not impact fixed costs in the same way as it would in a high-quality store.[3] As a result, we might expect that a policy promoting an increase in the assortment at a fringe store would impact variable costs.

The emphasis on how increasing assortment might impact the cost structure of a certain class of food retailers, namely convenience stores, in the preceding discussion is important in contextualizing the proposed policy we model in our counterfactual analysis. With the availability of micro-level scanner data, it has become more accessible for researchers to look at different measures of quality, such as directly measuring the assortment (i.e., variety) of products within a given retail chain. Where the research of Ellickson (2006) has previously looked at using average square footage of a retail chain, a measure highly correlated with

[2]A similar policy, called the "Staple Food Ordinance," was implemented in Minneapolis, MN in 2008.

[3]The distinction here is that fringe stores, by definition, do not invest in developing their distribution systems in the same way as a high-quality store would. Store size, on the other hand, is generally assumed to be predetermined by a store's location decision. In other words, the market a store decides to enter puts a constraint on the ability for a store to increase its size and therefore store size remains fixed, regardless of whether that store is a high-quality or fringe store.

product variety, to differentiate between high- and low-quality food retailers, the use of micro-level scanner data coupled with our specification of demand provides the foundation to motivate a more detailed scenario for our policy-based counterfactual analysis.[4] For this scenario, we have chosen to emulate the type of policy that would increase the assortment of healthy foods at a small-scale, non-traditional food retailer. The choice of our demand model, in particular, allows us to estimate the welfare changes specifically related to a change in store attributes – here, product assortment at convenience stores. In addition, where previous research has focused predominantly on the supermarket channel, we provide a contribution to the literature by allowing households to allocate their food expenditures across an array of store formats, including stores outside of the supermarket channel, such as supercenters, mass merchandisers, convenience stores, and dollar stores. Thus, our model seems appropriate to conduct a welfare analysis that seeks to evaluate alternative policy-based recommendations.

3 The Empirical Model

Prior research that focuses on consumers' revealed preferences for store selection, store attributes, or supermarket competition in an equilibrium setting focuses predominately on supermarket choice and competition among the supermarket and grocery channels. Informed by the line of research outlined in section 2, there are three main determinants that motivate the theoretical framework for our equilibrium model. First, in order to evaluate the welfare implications of an attribute-based policy, our demand model must focus on preferences for attributes. In this way, we can infer households' price sensitivity, store-switching behavior, and expenditure sensitivity with regard to specific attributes observable at the retail-level and then use the results from our estimation in our supply side. Second, the model must adequately manage the use of micro-level data. Our datasets provide key information about (i) consumer demographics, (ii) fixed costs (as reflected by distance measures or square

[4] Ellickson (2007) notes that a better metric of quality would be to combine variety with store size. Our equilibrium framework incorporates both of these dimensions.

footage) and (*iii*) variable costs (as reflected by store characteristics) of shopping, which are directly integrated into the demand side, and must also represent variable costs that impact food retailers in the short run. Although not a requirement for welfare analysis, in general, the third reason for our model choice is the ability to reflect realistic shopping behaviors of the households in our study area of the Philadelphia-metro area. Therefore, the model must be able to accommodate multiple shopping trips households make to the same and different retail chains and formats within a short time frame, as well as censoring, in that consumers generally do not choose to shop at all the stores in the choice set.[5] In the sections that follow, we describe the theoretical framework in constructing our equilibrium model, a precursor for our welfare analysis.

3.1 Demand

Following Rojas & Peterson (2008), we adopt the LA/AIDS model of Deaton & Muellbauer (1980) and incorporate the DM method of Pinkse et al. (2002) into this framework. Rojas & Peterson (2008) estimate a brand-choice model, which we transform into a retailer-choice model. Let i denote the household, j denote the retail chain within the choice set, and t denote the month. Therefore, the expenditure share function for a household i shopping at retail chain j in month t would resemble:

$$w_{ijt} = \alpha_{ij} + \sum_{j} \gamma_{jk} \log p_{jt} + \beta_{ij} \log \left\{ x_{it}/P_{it}^{L} \right\} + \epsilon_{ijt} \tag{3.1}$$

where $w_{ijt} = q_{ijt}p_{ijt}/x_{it}$ represents the expenditure share for household i's total food purchases at retail chain j in month t; p_{jt} represents a retail-level price index of retail chain j in month t; and x_{it} represents the total food expenditure by household i in month t. The parameters a_{ij}, γ_{jk}, and β_{ij} are to be estimated and ϵ_{ijt} is an error term. Each share equation (w_{ijt}) represents the share of total food expenditure a household allocates to the respective retail

[5] For the remainder of this paper, "stores" and "retail chains" may be interchanged and can be interpreted synonymously.

chain ($j = 1, ..., J$).

To linearize the price index term $\log P_{it}^L$, Moschini (1995) proposed to approximate this term with a log-linear analog of the Laspeyres index such that:

$$\log P_{it}^L \equiv \sum_{j=1}^{J} w_{ij}^0 \log p_{jt} \tag{3.2}$$

where w_{ijt}^0 is retail chain j's base share for household i, $w_{ij}^0 \equiv T^{-1} \sum_{t=1}^{T} w_{ijt}$, and $t \in (t, ..., T)$ represents the month. The base share of retail chain j for household i represents a yearly average of household i's purchase shares at retail chain j.

The Distance Metric

The cross-price coefficients γ_{jk} are specified as a function of distance measures between any two retail chains j and k (Pinkse et al. 2002). These distances (δ_{jk}) are spatial measures, measured in terms of physical space or attribute space, such that $\gamma_{jk} = g(\delta_{jk})$. This specification indicates that the level of substitutability depends on the "closeness" of attributes between retail chain j and retail chain k. Attributes may be discrete (δ_{jk}^d) or continuous (δ_{jk}^c). Retail chains can be "neighbors" in attribute space if discrete attributes are identical (for example, if both retailers are of the same format) or if continuous attributes are close in level measures (for example, if two retailers sell a similar number assortment of products). The closer the two retailers are in observable characteristics, the more likely they will be considered substitutes for one another (i.e., the closer competitors they are). Conversely, the farther apart in attribute space, the less likely the two retailers will be considered substitutes (Bonanno 2013).

The local measure of closeness, δ_{jk}^d, can be expressed as

$$\delta_{jk}^d = \begin{cases} 1 \quad if \quad |z_j^d - z_k^d| = 0 \\ 0 \quad if \quad |z_j^d - z_k^d| \neq 0 \end{cases} \tag{3.3}$$

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

$$\delta^c_{jk} = \frac{1}{1 + 2\sqrt{\sum(z^c_j - z^c_k)^2}} \tag{3.4}$$

where z^d_j is a discrete attribute of retail chain j and z^c_j is a continuous attribute of retail chain j.[6]

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

$$\begin{aligned} \gamma_{jk} \log p_{kt} = & \sum_{d=1}^{D} \left(\lambda^d_j \sum_{k \neq j}^{J} \delta^d_{jk} \log p_{kt} \right) \\ & + \sum_{c=1}^{C} \left(\lambda^c_j \sum_{k \neq j}^{J} \delta^c_{jk} \log p_{kt} \right) \end{aligned} \tag{3.5}$$

where λ^d_j and λ^c_j are the parameters to be estimated. By specifying our cross-price parameter as a function of distance, our system of $J-1$ equations and $J(J-1)/2$ cross-price parameters is reduced to a single equation and we estimate an elasticity matrix with dimension JxJ.

Additional Parameterizations

In addition, the constant term α_{ij}, the own-price coefficient γ_{jj}, and the coefficient on the expenditure term β_{ij} may be written as functions of household demographics and additional retail-level attributes, such that:

$$\alpha_{ij} = \alpha_0 + \sum_{h=1}^{H} \alpha_h h_{ih} + \sum_{l=1}^{L} \alpha_l z^{\alpha}_{jl} \tag{3.6}$$

$$\gamma_{jj} = \gamma_0 + \sum_{m=1}^{M} \gamma_m z^{\gamma}_{jm} \tag{3.7}$$

[6]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.

$$\beta_{ij} = \beta_0 + \sum_{n=1}^{N} \beta_n z_{jn}^{\beta} \tag{3.8}$$

where household i's characteristics can each be represented by h_{ih} and retail chain j's attributes can be represented by z_{jm}^{α}, z_{jm}^{γ}, and z_{jn}^{β}. Imposing 3.6, 3.7, and 3.8 into equation 3.1, the specification of the DM-LA/AIDS model is as follows:

$$\begin{aligned} w_{ijt} =& \alpha_0 + \sum_{h=1}^{H} \alpha_h h_{ih} + \sum_{l=1}^{L} \alpha_l z_{jl}^{\alpha} \\ &+ \left(\gamma_0 + \sum_{m=1}^{M} \gamma_m z_{jm}^{\gamma}\right) \log p_{jt} \\ &+ \sum_{d=1}^{D} \left(\lambda_j^d \sum_{k \neq j}^{J} \delta_{jk}^d \log p_{kt}\right) + \sum_{c=1}^{C} \left(\lambda_j^c \sum_{k \neq j}^{J} \delta_{jk}^c \log p_{kt}\right) \\ &+ \left(\beta_0 + \sum_{n=1}^{N} \beta_n z_{jn}^{\beta}\right) \log \frac{x_{it}}{P_{it}^L} + \epsilon_{ijt} \end{aligned} \tag{3.9}$$

Equation 3.9 above shows that retailer 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. In the intercept term, α_{ij}, we include a set of household-specific demographic variables (h_{ih}), chain-level variables (z_{jl}), and time fixed effects. These variables are included as additional measures that shift the demand share equation. The own-price parameter, γ_{jj}, is also written as a linear function of retailer attributes (z_{jm}). The signs on the γ's represent how the slope of the demand curve changes with its own price. For example, we might expect that households who shop at supercenters are more price sensitive and therefore we would expect the sign to be negative and significant. In addition, households' overall food-at-home expenditures may play a role in their budget allocations across different retail chains. Therefore, we interact the expenditure term with a geographic distance to determine households' expenditure sensitivity, represented by β, as distance to each retail chain increases.[7]

[7]Further discussion on the set of variables selected is presented in a later section.

Finally, drawing from a broad set of store characteristics, substitutability (and competition) between retail chains is modeled as a function of the relative distance between the retail chains along several attribute-space dimensions, such as attributes that reflect the format of the retail chain, the physical distance between stores, as well as attributes that reflect what is inside the stores (e.g., assortment, service levels, and relative prices). In equation 3.9, the sign and significance of λ_j^d and λ_j^c characterize consumer's 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 retail chain j by switching to another retail chain k with the same or similar attributes. In other words, store switching would not involve much distance in attribute space.

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. Even after reducing the choice set and aggregating to the retail-chain level, a disproportionate number of zero expenditure shares exists. While the use of traditional LA/AIDS may not be practical to resolve this dimensionality issue 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).

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:

$$w_{ijt} = \begin{cases} w_{ijt}^* & if \quad w_{ijt}^* > 0 \\ 0 & if \quad w_{ijt}^* = 0 \end{cases} \tag{3.10}$$

where

$$
\begin{aligned}
w_{ijt}^* =& \alpha_0 + \sum_{h=1}^{H} \alpha_h h_{ih} + \sum_{l=1}^{L} \alpha_l z_{jl}^{\alpha} + \Big(\gamma_0 + \sum_{m=1}^{M} \gamma_m z_{jm}^{\gamma}\Big) \log p_{jt} \\
&+ \sum_{d=1}^{D} \Big(\lambda_j^d \sum_{k \neq j}^{J} \delta_{jk}^d \log p_{kt}\Big) + \sum_{c=1}^{C} \Big(\lambda_j^c \sum_{k \neq j}^{J} \delta_{jk}^c \log p_{kt}\Big) \\
&+ \Big(\beta_0 + \sum_{n=1}^{N} \beta_n z_{jn}^{\beta}\Big) \log \frac{x_{it}}{P_{it}^L} + \epsilon_{ijt}
\end{aligned}
\tag{3.11}
$$

and $\epsilon_{ijt} \sim N(0,1)$. The final set of parameters to recover are α_h, α_l, γ_0, γ_m, λ_j^d, λ_j^c, β_0, and β_n. The empirical results presented in this paper present results using a Tobit model as shown in equation 3.11.

3.2 The Firm's Optimal Pricing Strategy

To set up the supply side, suppose there are N firms competing in the food retail market, where each firm represents the parent company that controls a subset N_r of the $j = 1, ..., J$ retail chains. We assume a linear-pricing model where each firm sets prices p_{jt} to maximize profits π_{rt}. To provide further context around the optimization problem, we suppose that the price the firm "chooses" is the single, unit price that maximizes the profit associated with providing a specific "shopping experience" – one that is characterized by the attributes of the retailer. As such, the price is meant to reflect not only the price of food items at the retailer, but also the markup that reflects retailer services and amenities, in addition to the costs associated with offering that assortment of items.

Assuming constant marginal costs and taking p_{kt} as given, each firm r chooses p_{jt} to maximize his profit function:

$$
\pi_{rt} = \sum_{j \in N_r} (p_{jt} - mc_{jt}) q_{jt}(p) - F_{jt} \tag{3.12}
$$

where p_{jt} is the retail-level price index for retail chain j in month t, $q_{jt}(p)$ is the market-level quantity demanded at retail chain j in month t, p is the vector of equilibrium price indexes, and mc_{jt} is the retail chain's short-run marginal cost in month t.[8]

Differentiating with respect to p_{jt} gives the following expression for retail chain j's profit-maximizing condition assuming pure-strategy Nash-Bertrand equilibrium in prices:

$$\frac{\partial \pi_{jt}}{\partial p_{jt}} = q_{jt}(p) + \sum_{k=1}^{J} (p_{jt} - mc_{jt}) \left(\frac{\partial q_{jt}}{\partial p_{kt}} * \frac{\partial p_{kt}}{\partial p_{jt}} \right) = 0 \tag{3.13}$$

where $\frac{\partial q_{jt}}{\partial p_{kt}}$ are the slopes of the demand function and $\frac{\partial p_{kt}}{\partial p_{jt}}$ are the price response functions.

Using the households' budget shares, w_{ijt}, and monthly expenditures, x_{it}, we calculate the aggregate quantity demanded at retail chain j in month t as follows:

$$q_{jt}(p) = \frac{\sum_{i=1}^{I^t} w_{ijt}(p) * x_{it}}{p_{jt}} \tag{3.14}$$

where $w_{ijt} = (p_{ijt} q_{ijt})/x_{it}$ follows from 3.1 and x_{it} is the total expenditure of household i in month t across the retail chains that comprise the choice set.

The first-order conditions in equation 3.13 can be stacked into a system of equations and written in vector notation:

$$(p - mc) = -[\Delta \times \Omega^*]^{-1} q(p) \tag{3.15}$$

where Δ is the retailer's response matrix, which contains the slopes of the demand curve such that each element, Δ_{jk}, is represented by $\frac{\partial q_{jt}}{\partial p_{kt}}$, the first derivative of quantity demanded with respect to price p_{kt}. Differentiating equation 3.14 with respect to p_{jt} gives us the elements of the Δ matrix:[9]

$$\frac{\partial q_j}{\partial p_k} = \begin{cases} \sum_{i=1}^{I} \frac{x_i}{p_j^2} \left[\frac{\partial w_{ij}}{\partial \ln(p_j)} - w_{ij} \right] & j = k \\ \sum_{i=1}^{I} \frac{x_i}{p_j p_k} \left[\frac{\partial w_{ij}}{\partial \ln(p_k)} \right] & j \neq k \end{cases} \tag{3.16}$$

In equation 3.15, Ω^* is the ownership matrix, where the element $\Omega^*_{jk} = 1$ if $j, k \in N_r$, zero otherwise. This notation implies that a firm maximizes profits over both j and k if both

[8]Following from price, marginal cost can also be interpreted as an index.
[9]Suppressing time notation t for brevity.

retail chains belong to their portfolio of stores. Without imposing any *a priori* assumptions on Ω^*, let $\Omega = \Omega^* * \Delta$, where Ω is the element-by-element multiplication of the two matrices. In the case of vertical price competition, specifications of Ω^* may also imply different forms of conduct between retail chains (e.g., single-firm pricing, multi-firm pricing, collusion). In principle, the most likely specification of ownership matrix in our application of retail-level competition is one where we specify a diagonal of ones.[10] However, we consider alternative games, such as joint profit maximization that occurs among stores within the same channel, as well as a case of collusion among all firms in the market.[11]

3.3 Marginal Cost Estimation

Solving for mc in equation 3.15, gives us the following system of equations:

$$mc = p + \Omega^{-1} q(p) \tag{3.17}$$

which can be calculated given our knowledge of retail-level price indexes, the ownership matrix, and estimated demand parameters.

Before implementing the counterfactual and solving for new equilibrium prices, we regress the calculated marginal costs from equation 3.17 and recover the coefficient estimates of store-level attributes (z_j) and other cost shifters (τ_{jt}). Formally, we express this relationship as follows:

$$mc_{jt} = f(z_j, \tau_{jt}) \tag{3.18}$$

The estimated parameters from this equation represent how a change in the level of a retail chain's attributes or other factors that vary across the retailers and time impact

[10] In this analysis, we have aggregated to the retail level, so with the exception of two retail chains, each store j is owned by a different parent company. Therefore, the likelihood that stores are maximizing profit across retail chains is not expected. Refer to section 4.1 for more information.

[11] As Slade (2004) demonstrates, given parameter estimates from a demand equation, a specification of Ω^*, and marginal costs, one can calculate estimates of market power, as measured by the Lerner Index. With this application in mind, we save this as an exercise in Appendix B.

marginal costs. Once we know the marginal effects of each attribute and cost shift, we can calculate a new predicted marginal cost after making changes to an attribute. With these new marginal costs, and by rearranging the terms in equation 3.17, we are able to solve for new equilibrium prices.

3.4 Welfare Calculations

As noted in section 2, the policy-based counterfactual scenario we simulate in this paper is one that emulates a recently-proposed initiative launched in the Philadelphia-metro area, namely the "Healthy Corner Store Initiative." Through this program, small-scale grocery and convenience stores receive targeted assistance from *The Food Trust* via a three-pronged approach. First, the program works with owners to enhance the variety of healthy food options, such as fresh fruits and vegetables, in the store. Second, to support their marketing plan, the program provides shelving and refrigerators to stock and display fresh foods, and provides color-coded bilingual marketing materials posted throughout the aisles to help customers distinguish between healthy products. Finally, the program offers a structured training plan to both the business owners and employees that covers topics on how to properly stock, display, and sell new items at affordable prices, as well as more technical training topics, such as how to use the new equipment and maximize limited space. Each aspect of the program is meant to support participating small-scale businesses sustain profitability given a change in their product offerings. The appeal of this program for customers is predominantly convenience, whereby they have access to healthier food items sold at an affordable price, yet do not have to travel outside of their neighborhood to purchase them.[12]

While this program targets small-scale grocery stores and independent convenience stores, which are largely underrepresented in our data, we do observe two similar-looking retailers that receive comparable levels of market demand, offer a similar level of assortment, and are roughly the same size in terms of square footage. Because our model is one based on

[12]Research that evaluates the effectiveness and profitability of this program for the retailer has not been established; however, anecdotally, participating stores have indicated that business has increased.

attributes and our policy scenario is attribute based, we consider these retailers as proxies for the "fringe"-type store described by Ellickson (2007) that falls in the low-quality tier of food retailers. Specifically, for our policy scenario, we consider a 10% increase in the assortment of products available at two convenience stores, holding square footage constant.

Depending on the type of game that is being played among the food retailers in this market will determine our expectations of new equilibrium prices. For our preliminary results, however, we assume that the retailers in this market all play a Bertrand game. On one hand, we might expect that an increase in assortment at convenience stores will generate new equilibrium prices lower than the original because the increase in the supply of these items will force other stores to lower their prices as well. On the other hand, an increase in assortment could lead to higher equilibrium prices for cost reasons. Therefore, the change in consumer surplus from a counterfactual policy could be positive or negative. For the total producer surplus, although we expect demand to increase at the two stores subject to the counterfactual, this adjustment might not yield a net positive result in producer surplus. We explore possible interpretations in the results section that follows.

Formally, following from section 3.3, we calculate producer and consumer surplus using the new prices after implementing the counterfactual and generating new equilibrium prices, where the aggregate producer surplus is:

$$PS = \sum_{r}^{N_r} \sum_{j \in N_r} \sum_{t}^{T} \left(p_{jt}^{new} - mc_{jt}^{new}\right) \frac{\sum_i^{I^t} x_{ijt}}{p_{jt}^{new}} \tag{3.19}$$

which is calculated using new equilibrium price and new marginal costs, using the estimated parameters from 3.18, and total consumer surplus is:

$$CS = \sum_{t}^{T} \sum_{j}^{J} \sum_{i}^{I} CS_{ijt}$$

where the individual consumer surplus for each household i who shops at retail chain j in month t is:

$$\begin{aligned} CS_{ijt} =& x_{it}\Big(\alpha_0 + \sum_{h=1}^{H} \phi_h h_{ih}\Big)\big(p_{jt}^0 - p_{jt}^{new}\big) + \frac{1}{2}\Bigg[\Big(\gamma_0 + \sum_{m=1}^{M} \gamma_m z_{jm}^{\gamma}\Big) \\ & - \Big(\beta_0 + \sum_{n=1}^{N} \beta_n z_{jn}^{\beta}\Big)\Bigg]\Big[\big(\log p_{jt}^0\big)^2 - \big(\log p_{jt}^{new}\big)^2\Big] \end{aligned} \tag{3.20}$$

4 Data

4.1 Market-Level Demographics and the Food Retail Landscape

We rely on three data sources to construct the panel of households and food retailers for our analysis. These three sources include (*i*) Information Resources, Inc. (IRI)'s household-level transactions dataset, called the Consumer Network Panel (CNP), (*ii*) TDLinx's Store Characteristics database (TDLinx), and (*iii*) USDA-ERS's Food Access Research Atlas (FARA). The CNP documents household-level food purchase transactions at food retail outlets across the U.S. Each household is geographically linked 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 the retail outlet where they shop is not known; only the identity of retail chain is known (Sweitzer et al. 2016). For example, if a household located in census tract m shops at Supermarket A and records purchases for that store, the researcher only observes that this household made a purchase at Retail Chain A, of which Supermarket A is perhaps but one location.[13] We augment the demographic data by including information on food accessibility from the FARA database, which provides information on whether a household's home census tract is considered a food desert. Various measures of food access available through the FARA are considered in our analysis, such as whether the household lives in a low-income census tract ($LowIncomeTract$) or in a census tract where a high percentage of households report not having access to a vehicle and live farther than a half

[13]Although this distinction is an important one, in the sense that our model represents household choice and competition among retail chains as opposed to specific store locations, it does not detract from the value of this analysis and the counterfactual we simulate.

mile to the nearest supermarket.

For our study period, we focus on the year 2012. We have chosen this year for our analysis for two reasons. First is that, prior to 2012, necessary household demographic information is not available through IRI, so 2012 is the first year in which demographic information can be linked with the purchase data Sweitzer et al. (2016). Second, the FARA database is updated only when new Census data is collected. At the time of data collection, the most current year was based off of the 2010 Census. Because our analysis focuses on the impact of a store attribute-base policy on welfare, our primary objective is to match datasets that are collected for the same year, or remain as close between years in order to capture the most cohesive snapshot of the food environment as possible and minimize the differences across years when data is infrequently collected.[14]

Our model of store choice and retailer competition is applied to the Philadelphia metro-area.[15] 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, or an envelop of the set of census tracts that fall within these four counties, thereby creating a self-contained, isolated market area.[16] This region identifies our study area, which includes the households and stores inside that area. Using TDLinx, 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.[17] The final choice set of retailers in the Philadelphia

[14]The extent to which the differences between years of data collection has not been evaluated. Analysis for additional years will be extended for future research. Extending this analysis to other years, especially 2011, is important as it will allow us to include another dataset, namely the EmpowerIT data, which provides information on wholesale prices. As we discuss later, this additional data may help us begin to address some of the empirical limitations with the supply side. The tradeoff for using the 2011 EmpowerIT data is making the assumption that household demographics are the same for 2012 and 2011.

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

[16]For more discussion on how we selected this region, please refer to chapter 3.

[17]Only the following channels are considered as viable food retail outlets: Grocery, Convenience, Mass Merchandisers, Supercenters, and Dollar Stores. Of the 26 stores that match, 22 are in the top 27 revenue-generating food retail chains, excluding club stores, reported in the CNP.

study area is comprised of 21 food retail chains.

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 retail chain j in month t. The denominator (x_{it}), or base expenditure, represents the total expenditure a household spends across all retail chains in the choice set and excludes any food purchases that households made at retail chains outside of the choice set.[18] This information is used to we recover the retail-level quantity demanded, $q_{jt}(p)$, for each retail chain j in month t according to equation 3.14.

A closer look at the shopping frequency patterns of households in the Philadelphia study area during the year 2012 in table 1 shows the following statistics (n=267 households): The average number of trips to any food store made in a month is seven, while the average number of unique food retail chains visited within a month is three. 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, 66% of the household's monthly food expenditure, followed by a second and third store totaling close to 20%. Subsequent stores account for the remaining 14%.

According to table 2, the four retail chains that receive the largest portion of households' expenditures are Grocery 1, 4, 6, and 9. Combined, they received 74% of the total market share during our study period. We match these figures with TDLinx, and, for the same time period, these chains account for roughly 17% of annual sales volume for the Philadelphia metro-area. This discrepancy could indicate that the market is less concentrated than is suggested by strictly looking at the IRI data. In addition, while Grocery 5 appears to be a less popular retail chain based on the shopping behaviors of CNP households, according to TDLinx, it ranks higher by annual sales volume for 2012.

After reviewing table 2, several insights can be gleaned about the competitive nature of the food retailing environment within the Philadelphia-metro area. If it is the case that the

[18] Using this method, the sum of the shares equals one, and does not allow for an outside option.

market is indeed less concentrated, then economic theory might predict that, all else equal, a policy that promotes the addition of a new store in the Philadelphia-metro area may be viable in the long run. In other words, if the local food retail market behaves competitively, then the profitability of a supermarket that receives an initial investment for overcoming entry costs will be proportional to the the size of the market. As Bonanno (2012) notes, a firm's decision to enter is conditional on the tradeoff between short-run profitability and fixed costs. So, if we observe that a market is competitive, then the entry of a new supermarket would result in a proportional decrease in the variable profits split equally among the food retailers (Ellickson 2007, Bonanno 2012, Bresnahan & Reiss 1991). In theory, a supply-side policy that targets fixed costs seems appropriate, as long as entry costs can be recovered by short-run profitability; however, this condition may not hold if vertical differentiation plays a role in firm-level strategies (i.e., not purely competitive), whereby firms separate themselves into different tiers as described by Ellickson (2007)'s EFC framework. In effect, a policy recommendation that acknowledges the non-uniformity of variable profit distributions among the competitors in the market after the policy has been realized is an important consideration and perhaps a better-suited policy for the Philadelphia-metro area.

4.2 Retailer Attributes

4.2.1 Retail-Level Price Index

Unlike previous uses of the DM method where the demand for the brand or product has a specific unit price, the concept of a unit price in the demand for a retailer is not as explicit. In other words, there is no single observed price that reflects the unit cost for a shopping experience at each separate retail chain. Therefore, we specify p_{jt} in equation 3.11 by creating a retailer-specific price index:

$$p_{jt} = \sqrt{\frac{\sum_g p^g_{jt} q^g_0}{\sum_g p^g_0 q^g_0} \frac{\sum_g p^g_{jt} q^g_{jt}}{\sum_g p^g_0 q^g_{jt}}} \tag{4.1}$$

where p_{jt}^g and q_{jt}^g are the price and purchase quantity for UPCs in category g at retail chain j during month t, respectively, and p_0^g and q_0^g are the market-level average price and quantity, or base price and base quantity, for UPCs in category g, respectively, where g denotes the product "category," which we define in the same way as IRI to capture similar food items sold. This price index is modeled after the Fisher Ideal price index and in meant to capture the relative prices for categories of products sold across all retailers in the choice set.[19] The price index appears in the model as *LNP*.

4.2.2 Retailer Type and Location

In addition to constructing a retail-level price index, additional characteristics are incorporated into the model by using both the CNP and TDLinx Store Characteristics data for 2012. We categorize retailer attributes according to two distinct classes. The first class are those attributes that motivate a household to choose the retail chain based on format type and location. These include the channel (namely, supermarket, dollar store, convenience store, supercenter, or mass merchandiser) and geographical distance measures (distance between stores and distance between household location 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. If two retail chains are of the same channel type, then the "distance" between these two retailers is zero, and this measure would receive a value of one because retail chains within the same channel are considered closer competitors. We also interact *Supercenter* with own-price to generate *LNP*x*Supercenter*. This variable represents the price sensitivity of households who shop specifically at supercenters.

We use the coordinates of each unique store from TDLinx to calculate the average physical distance between retail chain j and all other retail chains in the choice set ($SDIST$), such

[19] Additional price index measures are currently being considered; however, these measures are not presented in this paper.

that $j \neq k$. *SDIST* enters the model as a continuous distance measure (*DM_SDIST*). We use the inverse of *SDIST* so between-chain distance is translated as physical "closeness." The way to interpret this distance is to consider the geographical space in which any two given retail chain compete. In other words, the geographical closeness between the two chains measured by *SDIST* represents a measure of relative geographic spatial competition. It is expected that if two retailers, j and k, are close in physical space, a household will substitute to one of the stores within retail chain k that is closer in proximity to their original store within retail chain j, given a change 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 between stores located closer in physical space.

Likewise, the average distance between a household's census tract and retail chain is also calculated. *HHDIST* is measured by taking the sum of the miles between each census tract centroid and each unique store coordinate in TDLinx, and then dividing by the number of stores within that retail chain to create aggregate census tract-retailer distance pairs.[20] 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. *HHDIST* is interacted with the expenditure term and enters into the model as *LNEXP*x*HHDIST*. There are two notes to make about calculating this distance. First is that the CNP does not indicate the unique store location where a household shops, only the retail chain. In addition, we do not have the addresses of the households, only their home census tract, so we assume that the residence of the household is at the census-tract centroid. Second, recent evidence suggests that households bypass the store that is located closest to their home to do their food shopping (Ver Ploeg et al. 2015). Therefore, we calculate an average household to store distance, without making the assumption that households shop at the store closest to their home census tract.[21] More recent applications of the DM method have not incorporated a geographic

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

[21] Additional geographical distance measures are currently being considered; however, these measures are not presented in this paper.

distance closeness measure, as these applications focus on the brand and product space. Because we focus on retailer choice and competition, a contribution of our paper is that our application incorporates this additional dimension of geographical distance that was first used in Pinkse et al. (2002)'s model of spatial price competition.

4.2.3 "Inside the Store" Attributes

The second class of attributes reflects those attributes "inside the store" that might influence a household's retailer choice decision, including factors, aside from price, that might influence how much of their budget they allocate to each retail chain. These characteristics include product assortment levels, uniqueness 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).

To capture the diversity of attributes that reflects observable characteristics inside the retailer, 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, or retailer-choice models more generally. *Breadth* reflects the number of unique UPCs carried in retail chain 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, 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. The relative assortment across retail chains may be an important determinant of how households choose to substitute between stores if the overall price index of the retail chain increases. The more similar two stores are in their levels of product assortment, then the value of the distance measure will be closer to one to represent the two stores as being closer competitors. The level of assortment captured by *Breadth* is the variable of interest for our policy scenario.[22]

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*). 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. We suspect that share of unique products at each retail chain impact the household's price sensitivity; therefore, this measure enters as an interaction term with own-price (LNPx$ShUniqueUPC$).

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. Similar to the *PctMaxAssort* calculation, we normalize the average square footage by dividing by the square footage of the retail chain with the largest square footage, so *AvSqft* takes a value between zero and one. This measure enters the model as an interaction term with own price (LNPx$AVSQFT$) and captures the price sensitivity of households when they shop at a store with larger overall square footage. In this model specification, we use square footage as a proxy for other store services and amenities that are not reported in the data.

4.3 Fixed and Variable Costs

From the side of the consumer, we are able to measure important factors that impact the cost to consumers, such as travel distance (a fixed cost) and in-store attributes (a variable cost). On the supply side, however, although we are able to capture some cost shifters, given the data sources available for our analysis, it becomes exceedingly difficult to measure

[22] We note here that this is not the ideal measure to implement the policy-based scenario similar to the program outlined in the Healthy Corner Store Initiative, since we have not separated the healthy fruit and vegetable UPCs from all others. This information is available and we intend to make this update.

factors that impact costs that vary across the retail chain and time. In general, the lack of cost information available to the researcher looking to model retailer competition is a long-standing challenge, and one we face in this paper.

Food retailers, similar to consumers, face two types of costs: fixed costs and variable costs. As described in section 2, fixed costs are the major investments necessary for a food retailer to enter a market. These costs include the "brick and mortar" components of operating a store, such as the investments in building and storage facilities as well as the size of the selling space. In addition, a fixed cost may also be the investment a food retailer makes to enhance their distribution and logistic systems. Variable costs, on the other hand, are those investments that change with respect to the level of output produced by the firm. Common across all retailers, variable investments include costs associated with day-to-day operations, such as labor and energy. The number of employees or the level of energy usage may be correlated with the size of the store or the number of departments a store has, which are generally considered fixed costs. An increase in product assortment, for example, may play a different role on the cost structure of a small-scale retail store than a large-format supermarket. For a large-scale store, an increase in assortment may result in that retailer upgrading an entire logistics system, whereas for a small-scale grocery store, expanding a product line may require that store to hire additional employees or to offer short-term training.

To incorporate fixed and variable costs into our supply-side model, we rely predominantly on TDLinx. TDLinx provides useful information on the average square footage ($AvgSqft$) of a store, which we include as a cost shifter in our estimation of marginal costs.[23] We consider average square footage to be a store attribute on our demand side that impacts how a household's budget is allocated across retail chains, so we add $AvgSqFt$ as a shifter to control for size, which may be correlated with missing information on variable costs. TDLinx offers some information on the number of employees in the supermarket channel;

[23]The distance from a retail chain to its distribution center or measures of road conditions may be used as cost shifters. The data identifies the parent company of each retail chain, so we may be able to augment the current dataset with information about miles to the nearest distribution center. We plan to explore this option.

however, this information is not available for the supercenter, mass merchandiser, convenience stores, and dollar stores, and therefore we are unable to control for number of employees as a cost shifter when we estimate marginal cost. We acknowledge that we might be missing key information to control for factors that impact marginal costs, so we suspect that our estimation of marginal costs suffers from omitted variable bias. With these considerations in mind, we proceed with our analysis and present preliminary results below.

5 Preliminary Results

In this section, we begin with a brief summary of results from the demand estimation. Then, we turn our attention to the supply-side results. These preliminary results include summary statistics on the marginal cost calculations from equation 3.17, parameter estimates from equation 3.18, and estimates of new equilibrium prices under the policy-based counterfactual scenario and three different specifications of the ownership matrix.

5.1 Demand

We estimate equation 3.11 via a Tobit model and the results are presented in table 4. The set of demographic variables includes the following: household size, age of the head of the household, median income, as well as household-level fixed effects for SNAP eligibility, race identifiers (white, black, Asian), ethnicity (Hispanic), 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-ERS's Food Access Research Atlas.[24]

The first set of results involves the own price and attributes that are interacted with own price. Table 4 shows that the estimated coefficient on own-price ($LNP = -0.0356$) is negative

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

and significant, indicating that as the price of the retailer increases, the expenditure share decreases as expected. The coefficient on the interaction term between price and supercenter is also negative and significant ($LNP\text{x}Supercenter = -0.0238$). 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 ($LNP\text{x}AVSQFT = 0.0.0843$), implying that, all else constant, households are less price sensitive to retailers 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 retailers 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 ($LNP\text{x}ShUniqueUPC = 0.0.0845$). A positive and significant sign on this variable suggests that households are less price-sensitive to a retailer 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 retail chain to purchase a specific set of items with very limited substitutability, then a positive sign supports the notion that retailer 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 retail chain 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 retailers become more competitive or similar in attributes. A positive coefficient, for example, implies that households facing a price increase would tend to switch to retailers with similar attributes. The estimated coefficient associated with retailer closeness in terms of physical proximity is positive and significant ($DM\_SDIST = 0.1180$), supporting the notion that if the price of one retail chain increases, households will substitute to another retail chain that is close in physical distance, rather than travel to a retailer farther away. By suggesting that a retailer's location relative to other retailers 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 retail chain within the same channel ($DM\_Channel = 0.0039$). Both these results suggest that retail chains strongly compete for shoppers using location and channel type, and these relationships are important for investigating impacts to consumer welfare.

The substitutability metric that measures the expanse of product offerings compared to all other retailers in the choice set is negative and significant ($DM\_PctMaxAssort = -0.0231$), implying that households will switch to a retail chain 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 retail chains that do not carry the same set of products. Second, the switching result is also consistent with the notion that households choose retailers that serve for primary and secondary shopping purposes which are measured by differences in terms of product assortment. Faced with a price increase, consumers may switch to their secondary retailer instead of seeking out a retailer that is similar in product assortment to their primary retailer. The ability to understand this type of substitution behavior appears to be a direct potential benefit of the DM-LA/AIDS model whereby we can observe households' multiple shopping trips.

The final set of coefficients in table 4 involves the expenditure term and interaction between expenditure and household distance. The estimated parameter for the expenditure term is positive and significant ($LNEXP = 0.0292$), implying that as a household's budget increases, so does its share of expenditure at that retail chain. When expenditure is interacted with average household distance to a retail chain, the estimated coefficient is negative and significant ($LNEXP$x$HHDIST = -0.0042$). Adding this result to the coefficient on $LNEXP$ suggests a dampening effect, meaning the positive effect that the food budget has on retailer shares decreases as the distance to the retail chain 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 retailers 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 retailer. 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 5. 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 (−5.19 for mass merchandisers versus −2.25 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 relatively lowest elasticities (−1.21 and −2.27, respectively). This finding indicates that demand for both convenience stores and supercenters are comparable to the retail chains within the grocery channel (−2.20), 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 retailer choice, regardless of the price implications. Finally, specialty, or high-end, supermarkets have higher elasticities when compared with other supermarkets.

5.2 Supply

Marginal Costs

At this point, the results presented in the sections that follow are preliminary. Using the estimated parameters from the demand equation, we calculate equation 3.17 and our results are presented in the third column of tables 7-9. Table 7 represents the scenario of a single-firm profit maximizer. In this scenario, each firm r owns only one retail chain j and chooses the "price" that maximizes his profit for supplying the "shopping experience" at that retail chain. Tables 8 and 9 present the scenarios of multi-firm pricing and full collusion, respectively. A multi-firm profit maximizer would mean that there is one manager, or firm, that maximizes profit across all retail chains within the same channel, and full collusion would mean that one firm jointly maximizes profit across all retailers. Because each of our retail chains is owned by separate firms, the most theoretically consistent specification of the ownership matrix under a Bertrand game is a matrix of a diagonal of ones, or the single-firm profit maximizer in table 7.

Focusing on the single-firm pricing scenario, based on our calculations of marginal costs, we see that many of our marginal costs are negative. In equilibrium, under this scenario, a firm should set its prices according to the following:

$$\frac{p_{jt} - mc_{jt}}{p_{jt}} = -\frac{1}{\eta_{jj}}$$

where η_{jj} is the own-price elasticity from table 5. Therefore, we are not confident in our results presented in table 7, since our results are not consistent with the inverse elasticity rule (Hausman et al. 1994).[25] Regardless, we proceed with our analysis and use these calculations to estimate the impact of various attributes on our calculation of marginal cost. We estimate the following equation via OLS:

$$\begin{aligned} mc_{jt} =& a_{0jt} + a_{1j} AvgSqft_j + a_{2jt} UniqueUPC_{jt} + a_{3jt} Breadth_{jt} \\ &+ a_{4j} SC_j + a_{5j} MM_j + a_{6j} CONV_j + a_{7j} DS_j + \sum_t^T d_t Month_t + \varepsilon_{jt} \end{aligned} \tag{5.1}$$

where $AvgSqFt$ represents the average square footage of stores within retail chain j, $UniqueUPC$

[25] Refer to Appendix B.

is the total number of UPCs sold only within retail chain j in month t, $Breadth$ is the measure of assortment at retail chain j in month t, and dummy variables for each format are represented by SC, MM, $CONV$, and DS, with the Grocery channel and December as the reference case. Table 6 presents the results under three scenarios of firm conduct: column (1) is single-firm pricing, column (2) is multi-firm pricing (within the same channel), and column (3) is full collusion. Depending on our specification of the Ω^* matrix, i.e., the model of firm conduct we impose, the sign of the parameter estimates on the attributes changes across the three scenarios. In addition, the results show a lack of significance for all the attributes, where each of the parameter coefficients are not statistically different than zero.

There are two issues that could be driving these results. First, the demand specification and elasticity estimates are directly linked to the calculation of marginal cost that we derive from equation 3.17, so we must consider how the specification of the demand model, including which attributes are included as own-price shifters and cross-price shifters, might be impacting these results. As a first step, we will consider adding channel-specific fixed effects and other covariates to equation 3.11 and then reestimate the demand side. Second, the aggregation of household-level expenditure shares that generates a measure of market-level demand for each retail chain, as represented in equation 3.14, may also be impacting our results. Further exploration is needed to fully understand the relationship between w_{ijt}, p_{jt}, and $q(p)$.

Moreover, due to the limited number of cost shifters available from our data sources, we are not confident that the specification of marginal cost regression in equation 5.1 is the correct specification to fit the data. For example, $UniqueUPC$ is highly correlated with $Breadth$ and therefore these two variables should not both be included in the model. Our ultimate goal is to measure welfare impact after changes in attributes, so the proper specification of this equation is a critical step to implement our policy scenario. Because the retailers in our study area likely face the same or similar variable costs, such as energy costs, further consideration will be given on what additional cost shifters that vary by retail chain, such as distance to each retail chain's nearest distribution center, we could include in our estimation of marginal cost.

After recovering the parameter estimates $\hat{a}_{jt}$ from our estimation of marginal costs on attributes, we change the level of assortment at convenience stores by increasing assortment at each retail chain by 10% to represent our policy-based counterfactual scenario. Column (2) in table 7 presents the new assortment levels (*Breadth*) under the policy scenario. We recalculate marginal costs using equation 5.1 and our parameter estimates $\hat{a}_{jt}$. The new marginal costs after changing assortment levels for convenience stores are presented in column (4). After increasing the assortment at convenience stores, the marginal cost increases by 9% for Convenience Store 1 and decreases by 9% for Convenience Store 2. Tables 8 and 9 present the results under the multi-firm and collusive scenarios, respectively.

Counterfactual Price

Now that we have new equilibrium marginal costs, we are able to solve for the equilibrium price under the counterfactual scenario and we present these results in table 7 columns (6)-(7). The difference between column (6) and (7) is that we remove any months where the new equilibrium price was an extreme outlier.[26] While we admit that these are preliminary results, we have the necessary components to calculate changes in consumer surplus, producer surplus, and total welfare. The figures that follow tables 7-9 plot the new equilibrium prices under each of the models of firm conduct, where the points represent the original equilibrium prices from column (5) and the lines represent the new equilibrium prices after the counterfactual from column (7).

Because the marginal costs results and new equilibrium prices generated from the counterfactual are both preliminary, we do not present welfare calculations that rely on the counterfactual prices. Instead, until we can address the potential problems as described throughout the results section above, we propose several steps to systematically uncover the

[26]An outlier here means that the optimization package could not find a local minimum, so the equilibrium price results for select months are extremely large. This issue of convergence is likely due to a large number of very small expenditures made by households to some retail chains in these months, therefore causing issues when we invert this matrix.

underlying issues in the following section.

5.3 Welfare Estimates

FORTHCOMING

6 Next Steps and Concluding Remarks

This paper provides a theoretical model that describes both households' underlying preferences for food retailers as well as the retailers' decisions on the supply of store attributes. Because we are interested in conducting counterfactual analyses that highlight attribute-based policy scenarios, we rely on an expenditure-based, censored store-choice model: the DM-LA/AIDS, which provides several empirical benefits for estimating retail-level demand, including the ability to account for more realistic shopping patterns of households, such as shopping at multiple stores within the same time period. Our results show that households value convenience, which is reflected in households' lack of price-sensitivity in larger stores and households' valuation of choosing stores that are closer in physical distance – both to their home census tract and stores closer in proximate distance.

To construct our model of food retailer competition and to estimate new equilibrium prices under a policy-based counterfactual scenario, we adapt a structural framework similar to Nevo (2001) and Slade (2004), which uses coefficient estimates from the demand model to compute price-cost margins under alternative specifications of firm conduct. Using store attributes as cost shifters, we estimate the impact of various attributes on marginal cost to calculate new marginal costs under a counterfactual scenario that increases the assortment of products at convenience stores by 10% to emulate a policy similar to the Healthy Food Financing Initiative. The new marginal costs from the counterfactual policy, together with the demand model, generate a new set of equilibrium prices in the market. Solving for new equilibrium prices allows us to estimate welfare changes resulting from a hypothetical policy-based scenarios for households and firms within the Philadelphia-metro area.

Given the preliminary results from the supply-side estimation, there are several steps that we plan to explore in order to generate reliable welfare estimates. Further analysis is underway.

Table 1: Monthly Household Shopping Patterns from Consumer Network Panel

	Chains in Choice Set (j=21)	**All Retail Chains (J=113)**
Shopping Frequency		
# Trips to any food retail chain	7	10
# Trips to unique food retail chains	3	5
One-way distance to store (mi)	7.5	n/a
Food Expenditure		
Mean monthly expenditure by HH	\$325	\$391
Portion of budget received by		
Primary store	66% (\$215)	53% (\$207)
Secondary store	20% (\$65)	20% (\$78)
Others	14% (\$45)	27% (\$106)

Table 2: Retail-Level Summary Statistics – Market Shares and Household Expenditures

	"Market Share"		**Expenditure Share (w_{ijt})**				**Censored**
	IRI	**TDLinx**	**Unconditional**	**Conditional**			**Obs.**
Retail Chain	*(%)*	*(%)*	*Mean*	*Mean*	*Min*	*Max*	*(%)*
Grocery 1	17.3	4.7	12.1	42.9	38.1	49.1	70.5
Grocery 2	1.7	0.4	2.2	27.8	20.0	36.6	92.0
Grocery 3	0.8	0.2	0.6	31.9	23.3	48.7	98.2
Grocery 4	13.4	1.0	9.7	54.4	51.4	61.9	82.1
Grocery 5	2.1	2.5	3.1	30.0	21.1	37.6	89.8
Grocery 6	10.3	4.8	8.9	40.7	37.4	45.5	78.1
Grocery 7	0.8	0.4	0.6	18.0	13.1	24.1	96.8
Grocery 8	3.8	1.6	3.2	47.2	38.8	53.5	93.2
Grocery 9	33.1	7.9	27.5	62.1	57.1	66.4	55.7
Grocery 10	2.9	1.2	2.9	39.2	35.0	50.8	92.5
Grocery 11	0.7	0.7	1.5	29.0	16.0	38.4	95.0
Grocery 12	1.9	0.6	1.6	31.2	22.3	45.7	94.7
Supercenter	3.0	0.8	4.0	32.9	23.8	43.5	88.0
Mass Merchandiser 1	3.4	3.1	3.4	22.8	17.7	29.4	84.9
Mass Merchandiser 2	0.2	1.2	0.5	13.8	4.0	30.0	96.6
Mass Merchandiser 3	2.5	4.1	3.6	24.8	14.5	33.0	85.5
Convenience 1	0.1	2.0	0.1	6.7	0.5	17.9	97.9
Convenience 2	0.5	6.0	0.9	12.2	5.4	25.4	92.5
Dollar 1	0.4	1.0	0.8	13.5	7.9	24.9	93.7
Dollar 2	0.2	0.5	0.4	16.8	4.1	28.2	97.4
Dollar 3	0.9	1.0	1.7	12.2	7.8	17.5	86.2
CR_4	74.1	19.9					

Table 3: Select Retailer Attributes

Retail Chain	**Square Ft.**	**Breadth**	**UniqueUPC**	**HHDist**	**Gas**	**Pharm**	**General Merch.**	**Employees**	**Registers**
Grocery 1	34.50	2,387.42	945.92	7.43	Y	Y	Y	1306	153
Grocery 2	14.200	501.33	249.42	7.13				74	22
Grocery 3	40.00	1,054.92	427.75	11.04				96	16
Grocery 4	50.50	4,853.58	2,074.58	8.42	Y	Y	Y	404	33
Grocery 5	19.06	284.08	108.50	7.11			Y	596	104
Grocery 6	42.38	1,593.17	598.92	7.06		Y	Y	1,545	192
Grocery 7	20.00	140.75	39.25	6.66				47	12
Grocery 8	30.50	547.67	195.58	7.15		Y	Y	484	65
Grocery 9	43.86	3,294.75	1,365.33	7.20		Y	Y	391	47
Grocery 10	20.33	258.42	140.25	7.02				351	36
Grocery 11	12.00	138.00	83.92	7.13				50	13
Grocery 12	35.00	221.17	80.08	9.36				92	21
Supercenter	41.00	1,481.75	677.25	6.97		Y	Y	1,185	59
Mass Merchandiser 1	123.75	733.33	294.42	7.71		Y	Y	n/a	n/a
Mass Merchandiser 2	61.83	59.58	24.08	7.13		Y		n/a	n/a
Mass Merchandiser 3	116.00	728.00	331.58	7.10		Y	Y	n/a	n/a
Convenience 1	20.33	26.75	11.58	6.99	Y			n/a	n/a
Convenience 2	3.00	103.75	43.50	7.44	Y			n/a	n/a
Dollar 1	76.39	98.92	55.58	6.54			Y	n/a	n/a
Dollar 2	79.38	67.58	32.33	6.62				n/a	n/a
Dollar 3	12.06	278.08	155.58	7.48			Y	n/a	n/a

Table 4: Estimated Parameters for Preferred Model Specification

Variables	**Coefficient**		**Standard Errors**
Demographics (α)			
$HHSize$	0.0157	***	(0.0038)
$AgeHead$	-0.0004		(0.0005)
$MedInc$	-0.0000		(0.0000)
$SNAPelig$	-0.0018		(0.0147)
$White$	0.1380	***	(0.0306)
$Black$	0.1280	***	(0.0306)
$Asian$	-0.0164		(0.0499)
$Hispanic$	0.0828	***	(0.0246)
$ChildUnder18$	-0.0198		(0.0189)
$FemaleHead$	-0.0035		(0.0109)
$SeniorCit$	0.0617	***	(0.0142)
$College$	-0.0235		(0.0154)
$LowIncomeTract$	-0.0278	***	(0.0096)
Own-Price Shifters (γ)			
LNP	-0.0356	**	(0.0145)
LNPx$Supercenter$	-0.0238	***	(0.0052)
LNPx$AVSQFT$	0.0843	***	(0.0043)
LNPx$ShUniqueUPC$	0.0845	***	(0.0117)
Distance Metric (λ)			
DM_SDIST	0.1180	***	(0.0051)
$DM\_Channel$	0.0039	***	(0.0003)
$DM\_PctMaxAssort$	-0.0231	***	(0.0005)
Expenditure Shifters (β)			
$LNEXP$	0.0292	**	(0.0119)
$LNEXP$x$HHDIST$	-0.0042	***	(0.0014)
$Constant$	-1.1880	***	(0.1050)
σ	0.6900	***	(0.00585)
VIF	2.42		
No. of Obs.	67,284	(7,649 uncensored obs.)	
No. of Households	267		

*, **, *** represent 10%, 5%, and 1% significance levels, respectively.
S.e.'s are bootstrapped.

Table 5: Own-Price Elasticities

	(1)			(2)		
Retail Chain	**Conditional** (η_{jj}^{C})			**Unconditional** (η_{jj}^{U})		
Grocery 1	−1.14	***	(0.087)	−0.95	***	(0.033)
Grocery 2	−2.43	**	(1.003)	−0.93	***	(0.051)
Grocery 3	−5.20	**	(2.145)	−0.91	***	(0.044)
Grocery 4	−1.14	***	(0.055)	−0.93	***	(0.026)
Grocery 5	−1.51	**	(0.659)	−0.96	***	(0.046)
Grocery 6	−1.31	***	(.1588)	−0.93	***	(0.035)
Grocery 7	−2.99		(2.451)	−0.97	***	(0.078)
Grocery 8	−1.72	***	(.5129)	−0.96	***	(0.030)
Grocery 9	−1.06	***	(.0279)	−0.95	***	(0.023)
Grocery 10	−1.98	***	(.5997)	−0.94	***	(0.036)
Grocery 11	−3.27	***	(1.322)	−0.92	***	(0.049)
Grocery 12	−2.61	*	(1.401)	−0.95	***	(0.045)
Supercenter	−1.21	**	(.4573)	−0.98	***	(0.048)
Mass Merchandiser 1	−4.28	***	(.5821)	−0.63	***	(0.065)
Mass Merchandiser 2	−7.88	***	(2.332)	−0.70	***	(0.103)
Mass Merchandiser 3	−3.40	***	(0.410)	−0.65	***	(0.060)
Convenience 1	−3.72		(2.733)	−0.79	***	(0.211)
Convenience 2	−0.81		(2.136)	−1.01	***	(0.114)
Dollar Store 1	−3.28	*	(1.708)	−0.86	***	(0.105)
Dollar Store 2	−3.56		(2.670)	−0.92	***	(0.084)
Dollar Store 3	−3.14	*	(1.626)	−0.85	***	(0.116)
Min	−7.88					
Max	−0.81					
Mean	−2.75					
St. Dev.	1.65					
Mean Own-Price Elasticity by Channel Type						
Grocery	−2.20					
Supercenter	−1.21					
Mass Merchandiser	−5.19					
Convenience	−2.27					
Dollar Store	−3.33					

Standard errors in parentheses.

(1) The conditional elasticity measures the response of changes in retail chain expenditure share given an increase in price for households whose expected expenditure share at that retail chain is greater than zero.

(2) The unconditional elasticity measures the overall response of changes in retail chain expenditure share given an increase in price.

Table 6: Estimated Parameters for Marginal Cost Estimation

	(1)		(2)		(3)	
Variables	**Coef.**	**Std. Error**	**Coef.**	**Std. Error**	**Coef.**	**Std. Error**
Constant	756.788	(1298.50)	-671.829	(3039.40)	-788.366	(2002.00)
AvgSqFt	0.939	(2.85)	-0.199	(6.66)	-0.871	(4.39)
UniqueUPC	1491.164	(943.35)	-717.764	(2208.10)	-1260.327	(1454.50)
Breadth	-670.382	(402.17)	339.571	(941.36)	563.153	(620.06)
SC	-1739.891	(1643.20)	1641.722	(3846.30)	1356.024	(2533.50)
MM	-1803.239	(2389.90)	1926.029	(5594.20)	1828.096	(3684.80)
CONV	-967.094	(1251.50)	371.805	(2929.30)	343.733	(1929.50)
DS	-1077.820	(1048.80)	2245.146	(2455.00)	1644.301	(1617.10)
Monthly FE	X		X		X	

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

(1) Single-firm pricing; (2) Multi-firm pricing; (3) Joint profit-maximization

SC="Supercenter"; MM="Mass Merchandiser"; C="Convenience"; DS="Dollar Store"

Reference case is the Grocery channel in December.

Table 7: Counterfactual Scenario – Single-Firm Profit Maximizers

	Policy Scenario		**Equilibrium Outcomes**				
	Breadth (Mean)		**Marginal Cost (Mean)**		**Price Index (Mean)**		
Retail Chain	*Before CF*	*After CF*	*Before CF*	*After CF*	*Before CF*	*After CF*	*After CF**
Grocery 1	2,387.42	2,387.42	-17.04	-415.25	117.74	750.11	92.06
Grocery 2	501.33	501.33	-129.76	1,652.08	44.09	23.38	23.38
Grocery 3	1,054.92	1,054.92	128.55	842.54	88.97	423.36	39.47
Grocery 4	4,853.58	4,853.58	-40.29	32.56	104.93	122.55	71.50
Grocery 5	284.08	284.08	-71.87	1,052.83	67.20	42.61	42.61
Grocery 6	1,593.17	1,593.17	-39.36	-191.02	119.46	509.77	115.93
Grocery 7	140.75	140.75	67.01	989.92	84.87	89.82	69.04
Grocery 8	547.67	547.67	-78.20	691.83	116.88	75.86	75.86
Grocery 9	3,294.75	3,294.75	-11.47	1,921.36	113.90	35.04	35.04
Grocery 10	258.42	258.42	8,242,47	1,710.31	170.69	75.92	63.62
Grocery 11	138.00	138.00	95.21	1,599.26	63.48	24.03	24.03
Grocery 12	221.17	221.17	864.84	1,200.60	104.52	59.99	59.99
Supercenter	1,481.75	1,481.75	-28.92	-28.92	88.27	3,774.23	87.70
Mass Merchandiser 1	733.33	733.33	163.17	-6.44	90.50	314.29	64.22
Mass Merchandiser 2	59.58	59.58	100.07	-102.42	95.34	447.11	94.10
Mass Merchandiser 3	728.00	728.00	138.62	510.73	95.55	138.58	74.78
Convenience 1	26.75	**29.43**	188.18	**205.75**	188.00	**351.66**	**147.45**
Convenience 2	103.75	**114.13**	192.13	**174.56**	139.83	**164.69**	**111.37**
Dollar 1	98.92	98.92	131.08	319.99	93.59	313.30	72.97
Dollar 2	67.58	67.58	133.84	186.15	80.63	69.01	69.01
Dollar 3	278.08	278.08	892.79	651.56	65.04	948.28	29.70

*We drop the months for which the new equilibrium price index is an outlier and present the results in this column.

Figure 1: New Equilibrium Price Index Under Single-Firm Profit Maximization

Table 8: Counterfactual Scenario – Multi-Firm Profit Maximizers

	Policy Scenario		**Equilibrium Outcomes**				
	Breadth (Mean)		**Marginal Cost (Mean)**		**Price Index (Mean)**		
Retail Chain	*Before CF*	*After CF*	*Before CF*	*After CF*	*Before CF*	*After CF*	*After CF**
Grocery 1	2,387.42	2,387.42	-89.50	-510.73	117.74	59.58	59.58
Grocery 2	501.33	501.33	869.45	-1,875.75	44.09	40.31	40.31
Grocery 3	1,054.92	1,054.92	-3,132.68	-1,327.24	88.97	66.79	66.79
Grocery 4	4,853.58	4,853.58	-469.01	-269.31	104.93	35.69	35.69
Grocery 5	284.08	284.08	-954.69	-1,611.68	67.20	56.73	56.73
Grocery 6	1,593.17	1,593.17	-339.44	-732.81	119.46	76.04	76.04
Grocery 7	140.75	140.75	957.66	-1,603.22	84.87	84.66	84.66
Grocery 8	547.67	547.67	-2,063.87	-1,364.42	116.88	88.51	88.51
Grocery 9	3,294.75	3,294.75	-313.82	-1,447.42	113.90	80.23	80.23
Grocery 10	258.42	258.42	-13,424.03	-1,929.26	170.69	141.39	141.39
Grocery 11	138.00	138.00	715.11	-1,917.25	63.48	60.34	60.34
Grocery 12	221.17	221.17	2,991.38	-1,653.05	104.52	118.79	95.32
Supercenter	1,481.75	1,481.75	-28.92	-28.92	88.27	55.58	55.58
Mass Merchandiser 1	733.33	733.33	165.15	297.32	90.50	90.55	90.55
Mass Merchandiser 2	59.58	59.58	91.40	72.89	95.34	79.81	79.81
Mass Merchandiser 3	728.00	728.00	141.51	27.84	95.55	117.69	107.72
Convenience 1	26.75	**29.43**	-3,566.86	**-1,384.23**	188.00	**167.15**	**156.83**
Convenience 2	103.75	**114.13**	828.87	**-1,353.76**	139.83	**110.82**	**110.82**
Dollar 1	98.92	98.92	84.81	407.21	93.59	65.04	65.04
Dollar 2	67.58	67.58	119.75	467.10	80.63	55.12	55.12
Dollar 3	278.08	278.08	958.82	289.07	65.04	42.51	42.51

*We drop the months for which the new equilibrium price index is an outlier and present the results in this column.

Figure 2: New Equilibrium Price Index Under Multiple-Firm Profit Maximization

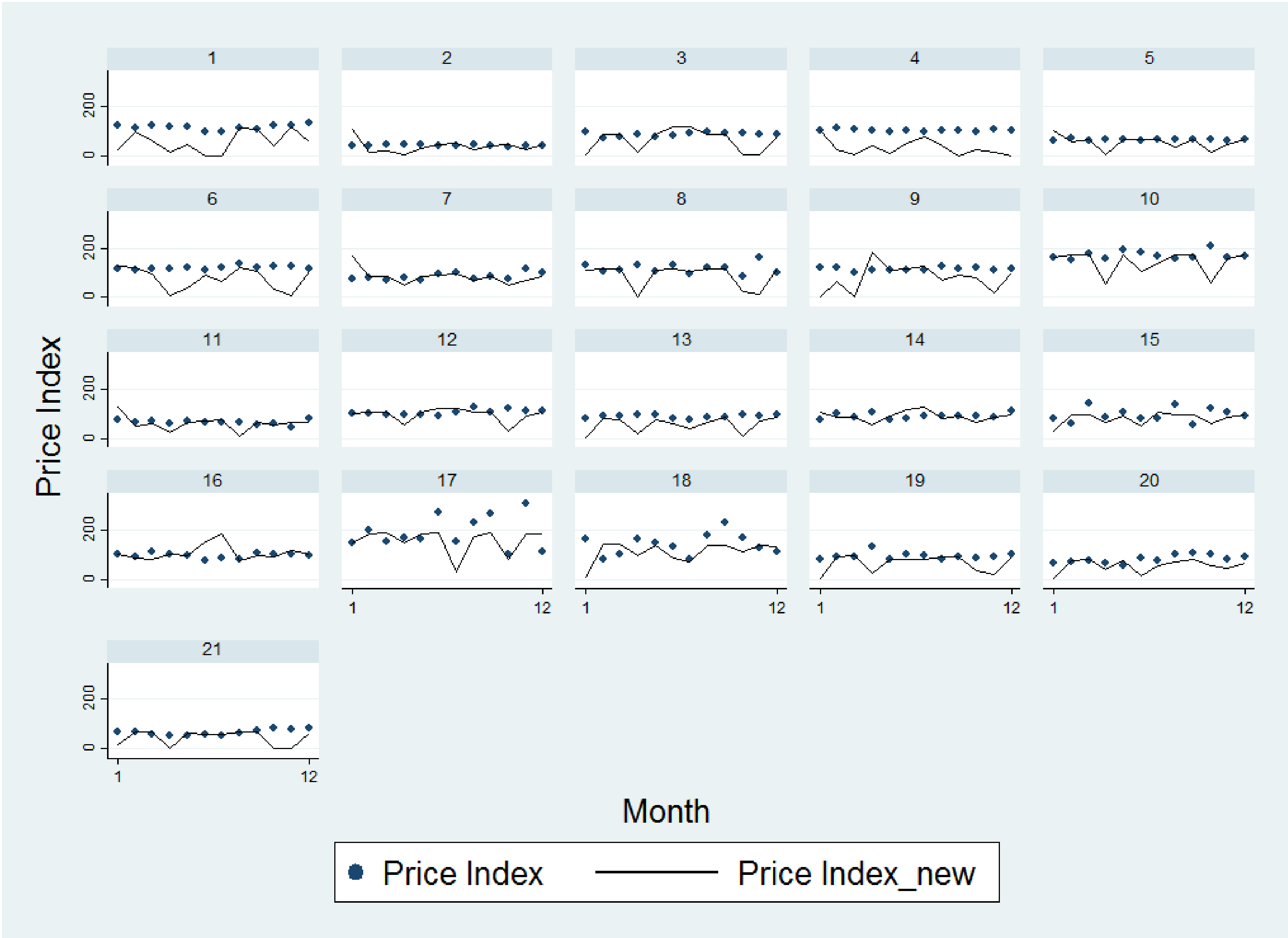

Table 9: Counterfactual Scenario – Joint Profit Maximizers

	Policy Scenario		**Equilibrium Outcomes**				
	Breadth (Mean)		**Marginal Cost (Mean)**		**Price Index (Mean)**		
Retail Chain	*Before CF*	*After CF*	*Before CF*	*After CF*	*Before CF*	*After CF*	*After CF**
Grocery 1	2,387.42	2,387.42	-127.36	68.61	117.74	43.57	43.57
Grocery 2	501.33	501.33	537.33	-1,598.01	44.09	40.39	40.39
Grocery 3	1,054.92	1,054.92	-709.58	952.70	88.97	58.79	58.79
Grocery 4	4,853.58	4,853.58	-358.83	-407.29	104.93	67.77	67.77
Grocery 5	284.08	284.08	-283.81	-1,087.75	67.20	63.75	63.75
Grocery 6	1,593.17	1,593.17	-87.05	-99.55	119.46	57.70	57.70
Grocery 7	140.75	140.75	287.74	-1,030.36	84.87	88.91	88.91
Grocery 8	547.67	547.67	-859.45	-800.52	116.88	99.27	87.12
Grocery 9	3,294.75	3,294.75	-123.22	-1,948.15	113.90	119.06	102.74
Grocery 10	258.42	258.42	-10,231.51	-1,643.55	170.69	181.33	169.66
Grocery 11	138.00	138.00	741.88	-1,538.14	63.48	41.14	41.14
Grocery 12	221.17	221.17	711.67	-1,222.72	104.52	75.58	75.58
Supercenter	1,481.75	1,481.75	-346.17	-346.17	88.27	61.15	61.15
Mass Merchandiser 1	733.33	733.33	-261.24	15.67	90.50	119.11	96.82
Mass Merchandiser 2	59.58	59.58	59.75	167.58	95.34	127.23	89.48
Mass Merchandiser 2	728.00	728.00	-30.38	-415.31	95.55	107.86	107.86
Convenience 1	26.75	**29.43**	829.49	**-823.50**	188.00	**164.67**	**164.67**
Convenience 2	103.75	**114.13**	-2,453.54	**-800.55**	139.83	**112.78**	**112.78**
Dollar 1	98.92	98.92	496.45	280.12	93.59	64.10	64.10
Dollar 2	67.58	67.58	548.99	394.10	80.63	57.94	57.94
Dollar 3	278.08	278.08	-380.91	-9.69	65.04	135.03	95.19

*We drop the months for which the new equilibrium price index is an outlier and present the results in this column.

Figure 3: New Equilibrium Price Index Under Joint-Firm Profit Maximization

References

Bonanno, A. (2012), 'Food deserts: Demand, supply, and economic theory', *Choices Magazine* **27**(3).

Bonanno, A. (2013), 'Functional foods as differentiated products: The Italian yogurt market', *European Review of Agricultural Economics* **40**(1), 45–71.

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

Bresnahan, T. F. & Reiss, P. C. (1991), 'Entry and Competition in Concentrated Markets', *Journal of Political Economy* **99**(5), 977.

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.

Ellickson, P. B. (2005), Does sutton apply to supermarkets?

Ellickson, P. B. (2006), 'Quality competition in retailing: A structural analysis', *International Journal of Industrial Organization* **24**, 521–540.

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

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. (2016), Is the focus on food deserts fruitless ? Retail access and food purchases across the socioeconomic spectrum.

Hausman, J., Leonard, G. & Zona, J. (1994), 'Competitive analysis with differenciated products', *Annales d'Economie et de Statistique* (34), 159–180.
URL: *http://www.jstor.org/stable/20075951*

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.

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.

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.

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

Nevo, A. (2001), 'Measuring Market Power in the Ready-to-Eat Cereal Industry', *Econometrica* **69**(2), 307–342.

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

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

Richards, T. J. & Pofahl, G. (2010), 'Pricing Power by Supermarket Retailers: A Ghost Machine?', *Choices* **25**(2).

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.

Slade, M. E. (2004), 'Market Power and Joint Dominance in U.K. Brewing', *The Journal of Industrial Economics* **LII**(1), 133–163.

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

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., 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. et al. (2010), 'Food environment, food store access, consumer behavior, and diet', *Choices* **25**(3), 1–5.

Villas-Boas, S. & Hellerstein, R. (2006), 'Identification of supply models of retailer and manufacturer oligopoly pricing', *Economics Letters* **90**(1), 132–140.

A The Firm's Optimal Pricing Strategy

The profit maximization problem for each firm r in month t is:

$$\max_{p_{jt}} \ \pi_{rt} = (p_{jt} - mc_{jt})q_{jt}(p) - F_{jt} \tag{A.1}$$

Differentiating equation A.1 with respect to p_{jt} gives the following:

$$\begin{aligned}\frac{\partial \pi_{jt}}{\partial p_{jt}} =& q_{jt}(p) + p_{jt}\left(\frac{\partial q_{jt}}{\partial p_{1t}} * \frac{\partial p_{1t}}{\partial p_{jt}}\right) + p_{jt}\left(\frac{\partial q_{jt}}{\partial p_{2t}} * \frac{\partial p_{2t}}{\partial p_{jt}}\right) \\ & + \ldots + p_{jt}\left(\frac{\partial q_{jt}}{\partial p_{kt}} * \frac{\partial p_{kt}}{\partial p_{jt}}\right) + \ldots + p_{jt}\left(\frac{\partial q_{jt}}{\partial p_{nt}} * \frac{\partial p_{nt}}{\partial p_{jt}}\right) \\ & + \frac{\partial C_{jt}(.)}{\partial q_{jt}}\left(\frac{\partial q_{jt}}{\partial p_{1t}} * \frac{\partial p_{1t}}{\partial p_{jt}}\right) + \frac{\partial C_{jt}(.)}{\partial q_{jt}}\left(\frac{\partial q_{jt}}{\partial p_{2t}} * \frac{\partial p_{2t}}{\partial p_{jt}}\right) \\ & + \ldots + \frac{\partial C_{jt}(.)}{\partial q_{jt}}\left(\frac{\partial q_{jt}}{\partial p_{kt}} * \frac{\partial p_{kt}}{\partial p_{jt}}\right) + \ldots + \frac{\partial C_{jt}(.)}{\partial q_{jt}}\left(\frac{\partial q_{jt}}{\partial p_{nt}} * \frac{\partial p_{nt}}{\partial p_{jt}}\right)\end{aligned} \tag{A.2}$$

Therefore, the profit maximizing condition for retail chain j in month t is

$$q_{jt}(p) + (p_{jt} - mc_{jt})\sum_{k=1}^{J}\left(\frac{\partial q_{jt}}{\partial p_{kt}} * \frac{\partial p_{kt}}{\partial p_{jt}}\right) = 0 \tag{A.3}$$

Define Δ as the retailer's response matrix, which contains the slopes of the demand curve such that each element of Δ is represented by the first derivatives of quantity demanded with respect to all prices, $\frac{\partial q_j}{\partial p_k}$:[27]

$$\Delta = \begin{bmatrix} \frac{\partial q_1}{\partial p_1} & \frac{\partial q_1}{\partial p_2} & \frac{\partial q_1}{\partial p_3} & \cdots & \frac{\partial q_1}{\partial p_{21}} \\ \frac{\partial q_2}{\partial p_1} & \frac{\partial q_2}{\partial p_2} & \frac{\partial q_2}{\partial p_3} & \cdots & \frac{\partial q_2}{\partial p_{21}} \\ \frac{\partial q_3}{\partial p_1} & \frac{\partial q_3}{\partial p_2} & \frac{\partial q_3}{\partial p_3} & \cdots & \frac{\partial q_3}{\partial p_{21}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial q_{21}}{\partial p_1} & \frac{\partial q_{21}}{\partial p_2} & \frac{\partial q_{21}}{\partial p_3} & \cdots & \frac{\partial q_{21}}{\partial p_{21}} \end{bmatrix}$$

[27] Suppressing time notation t for brevity.

The elements of the Δ matrix, $\frac{\partial q_j}{\partial p_k}$, are calculated as follows:

$$\frac{\partial q_j}{\partial p_k} = \begin{cases} \sum_{i=1}^{I} \frac{x_i}{p_j^2} \left[\frac{\partial w_{ij}}{\partial \ln(p_j)} - w_{ij} \right] & , \quad j = k \\ \sum_{i=1}^{I} \frac{x_i}{p_j p_k} \left[\frac{\partial w_{ij}}{\partial \ln(p_k)} \right] & , \quad j \neq k \end{cases} \tag{A.4}$$

where

$$\begin{aligned} q_j(p) &= \frac{\sum_{i=1}^{I} w_{ij}(p) x_i}{p_j} \\ \frac{\partial w_{ij}}{\partial \ln(p_j)} &= \gamma_0 + \sum_{m=1}^{M} \gamma_m z_{jm}^{\gamma} - (\beta_0 + \sum_{n=1}^{N} \beta_n z_{jn}^{\beta}) w_{ij} \\ \frac{\partial w_{ij}}{\partial \ln(p_k)} &= \sum_{d=1}^{D} \lambda_j^d \sum_{k} \delta_{jk}^d + \sum_{c=1}^{C} \lambda_j^c \sum_{k} \delta_{jk}^c - (\beta_0 + \sum_{n=1}^{N} \beta_n z_{jn}^{\beta}) w_{ik} \end{aligned} \tag{A.5}$$

Let Ω^* represent the ownership matrix, where each element, Ω^*_{jk}, represents the retailer's price response function, $\frac{\partial p_k}{\partial p_j}$:

$$\Omega^* = \begin{bmatrix} \frac{\partial p_1}{\partial p_1} & \frac{\partial p_2}{\partial p_1} & \frac{\partial p_3}{\partial p_1} & \cdots & \frac{\partial p_{21}}{\partial p_1} \\ \frac{\partial p_1}{\partial p_2} & \frac{\partial p_2}{\partial p_2} & \frac{\partial p_3}{\partial p_2} & \cdots & \frac{\partial p_{21}}{\partial p_2} \\ \frac{\partial p_1}{\partial p_3} & \frac{\partial p_2}{\partial p_3} & \frac{\partial p_3}{\partial p_3} & \cdots & \frac{\partial p_{21}}{\partial p_3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial p_1}{\partial p_{21}} & \frac{\partial p_2}{\partial p_{21}} & \frac{\partial p_3}{\partial p_{21}} & \cdots & \frac{\partial p_{21}}{\partial p_{21}} \end{bmatrix}$$

Without imposing any *a priori* assumptions on Ω^*, let $\Omega = \Omega^* * \Delta$, where Ω is the element by element multiplication of the two matrices. Then, we can stack A.3 and rewrite in matrix notation:

$$q(p) + (p - mc)\left[\Omega^* \Delta\right] = 0 \tag{A.6}$$

Solving for price-cost margins and marginal costs:

$$\begin{aligned} p - mc &= -\Omega^{-1} q(p) \\ mc &= p + \Omega^{-1} q(p) \end{aligned} \tag{A.7}$$

B Single-Firm Price Maximizers

For simplicity, suppose that each firm r owns only one retail chain, and that retail chain produces only one good – that is, the "shopping experience." Under the Nash-Bertrand assumption, each retail chain chooses price p_{jt} to maximize his profit over a single good. Following from section 3.2, we write the ownership matrix as:

$$\Omega^*_{jk} = \begin{cases} 1 & \text{if} \quad j = k \\ 0 & \text{if} \quad j \neq k \end{cases}$$

This assumption greatly simplifies equation 3.13 so that it can be written:

$$q_{jt}(p) + (p_{jt} - mc_{jt})\frac{\partial q_{jt}}{\partial p_{jt}} = 0 \tag{B.1}$$

Solving for price-cost margins yields:

$$(p_{jt} - mc_{jt}) = -q_{jt}(p)\frac{\partial p_{jt}}{\partial q_{jt}} \tag{B.2}$$

Dividing both sides of equation B.2 by p_{jt} gives us the equilibrium condition for single-product firms:

$$\frac{p_{jt} - mc_{jt}}{p_{jt}} = -\frac{1}{\eta_{jj}} \tag{B.3}$$

where η_{jj} are the estimated demand elasticity results. According to Slade (2004) and Villas-Boas & Hellerstein (2006) this condition provides information on the level of market power exercised in a market. The lefthand side of B.3 is the Lerner Index, a common measure of market power exhibited by retail chain j in month t. If the market in which the retailers operate behaves competitively, we can infer deviations from price-taking behavior by using the estimated demand elasticity and imposing the assumption of single-firm ownership. While this condition only holds in the case of a Nash-Bertrand assumption, given additional information on marginal costs one would be able to evaluate a crude measure of market power.