Do Dollar Store Bans Work?
Grocery Store Choice and The Nutrition Impact of Dollar Stores

Felix Nguyen*

May 9, 2022

Abstract

In this paper, I study the potential effects of a ban on dollar stores, a recent trend that has been adopted by local governments, purportedly to combat nutrition inequality. Combining Dubois et al. (2014) nutrient demand model with an upper level store choice model, I estimate the household decision to shop at a store, and how many calories to purchase from each store. With this demand system, I perform counterfactual policy experiments to simulate a dollar store ban, and see how that affects calories intake of local residents by income group. The results help shed light on whether a dollar store ban is justified, and contribute to the literature on the sources of nutrition inequality.

Keywords: Calories Demand, Random Coefficients Nested Logit, Nutrition Inequality, Dollar Store Ban.

*PhD Student in Marketing - Emory University - pnguy38@emory.edu

1 Introduction

In the last decade, dollar store has been one of the fastest growing sectors of retail industry in the United States, especially in an era that has witnessed the nadirs of several brick-and-mortar retail giants. Within ten year from 2009 to 2019, the number of dollar stores across the country almost doubled, from approximately 17,500 stores in 2009 to more than 30,000 stores in 2019. These discounted stores appeal to customer segments with tighter budget constraints, in neighborhoods with lower access to affordable grocery. During the "brick-and-mortar retail slump" caused by COVID-19 pandemic, dollar store is the only segment that experienced positive growth. According to CNN and industry research firm Coresight Research, " Three dollar store chains will make up almost half of all the new stores opening up in the United States " in 2021[1], and Dollar General, the largest dollar chain, alone will make up 1 in 3 new stores opening. As American consumers struggle with rising food and fuel prices, the demand for affordable groceries within reasonable travel distance would grow more and more, and it appears the expansion of dollar stores will not cool down in the near future.

Yet, dollar stores are also known for providing mainly highly processed food that may have lower nutritional value, as these are often the choices with the lowest prices. Additionally, dollar stores main positioning is to deliberately target lower income communities in areas often called "food deserts" in popular media. These facts have led to rising concerns from policy makers, public health officials, and urban planners about whether the popularity of dollar stores would perpetuate the nutritional inequality and food-related epidemics (e.g. obesity) in these communities. According to the non-profit consumer advocacy group Center for Science in the Public Interest, "Dollar stores are likely to exacerbate existing diet-related health disparities"[2]. Since 2018, there has been a rising wave of legislations by local government throughout the United States banning

[1] https://www.cnn.com/2021/05/06/business/dollar-store-openings-retail/index.html
[2] https://www.cspinet.org/resource/rise-dollar-stores

dollar stores or forbidding the entry of new dollar stores. According to a report by DeKalb County, Georgia, one of the local governments where new dollar stores are banned, by 2020 there have been 25 cities with similar legislation[3], including large metro areas such as Fort Worth, TX; Tulsa and Oklahoma City, OK; New Orleans, LA; Cleveland, OH; and many more are exploring this type of bans.

There have been also pushbacks from other advocacy groups, with some calling these bans a case of "Class elitism" by policy makers against lower income demographics[4]. Another argument points out that "The idea that dollar stores are invaders ignores the fact that these retailers are expanding in neighborhoods that want them"[5]. Without actual in-depth research on this topic, it is all but impossible to conclude which side has the correct answer. It is possible that while dollar stores indeed offer mostly food items with lower nutrients, they are simply addressing existing demands of lower income consumers. Therefore, without dollar stores these consumers would still purchase similar items elsewhere for higher prices, resulting in net welfare losses from a dollar store ban. The main goal of my study, hence, is addressing this information gap and provide policy makers with a better understanding of the effectiveness and potential backlashes of dollar store ban.

Through this research, I contribute to the growing literature on food consumption habit, food desert, and the sources of nutrition inequality. Prior studies in this area have been mostly ambivalent on whether supply side factors such as accessibility, product assortment and price have substantial effects on consumer's food choices and calories intake. Using Walmart Supercenters entry dates and locations combined with data from the Behavioral Risk Factor Surveillance System, Courtemanche and Carden (2011) find that an additional Walmart Supercenter per 100,000 residents leads to an 0.24 increase in average BMI, and 2.3% increase in obesity rate. On the other hand, merging the same

[3]https://www.dekalbcountyga.gov/sites/default/files/users/user3566/Small Box Discount Retail Store (SBDR) Report.pdf

[4]https://fee.org/articles/let-them-eat-whole-foods-the-appalling-elitism-of-dollar-store-bans/

[5]https://www.city-journal.org/banning-dollar-stores

data with additional surveys, in a follow up study the same authors find that Walmart Supercenters entrance leads to an improvement in food insecurity, especially amongst the lower income groups (Courtemanche et al. 2019). From a public health perspective, Larson et al. (2009) provide a comprehensive review of extant literature on the effect of access to healthy foods and obesity. Overall, the results are mixed, with modest and inconsistent effects, and due to the cross-sectional nature cannot account for endogeneity in demand and stores entry. Similarly, Hut and Oster (2022) find that dietary changes are unusual, with most households remain consistent in their eating habits in spite of changes in financial circumstances or health diagnosis.

From a more structural perspective, Dubois et al. (2014) estimate a demand system for food and nutrients three different markets: the US, the UK, and France and then simulate counterfactual choices if households faced prices and nutritional characteristics from other countries. They find that while differences in prices and product assortments are explanatory of some of the variation in food choices, overall inherent differences in preferences and eating habits are needed to explain cross-country differences. Building on this approach, Allcott et al. (2019) estimate a structural model of calories demand, and then show through counterfactual simulation that changing low income households supply factors (prices, stores, and product assortments) to match those of richer neighborhoods reduces nutritional inequality by only less than 10%, while the remaining 90% can only be explained by inelastic differences in demand. This finding was also corroborated by reduced form analyses of stores entry and movers, which rejects the hypothesis that neighborhood environments contribute meaningfully to nutritional inequality.

In this research, I adopt the same approach as Dubois et al. (2014) and Allcott et al. (2019)'s to construct the "lower level" of my demand system, while adding an "upper level" of spatial store choices estimated with a random coefficients nested logit, in order to accurately capture the effects of dollar stores on consumer calories demand through both "accessibility", i.e. shorter travel distance, and prices and available product assortments.

With the estimated demand system, I can then perform counterfactual policy experiment to examine the effect of a dollar store ban on calories consumption of local consumers, as well as its effect on overall consumer welfare. Alternative policies can also be considered through counterfactual simulation, such as enforcing a quota on the number of dollar stores within a market, and calculating the optimal quota. In in all, the study will provide much needed insights into real societal and public health impact of dollar stores, and an answer to whether a ban on dollar store is truly beneficial to their local communities.

2 Context and Data

2.1 The Growth of Dollar Store

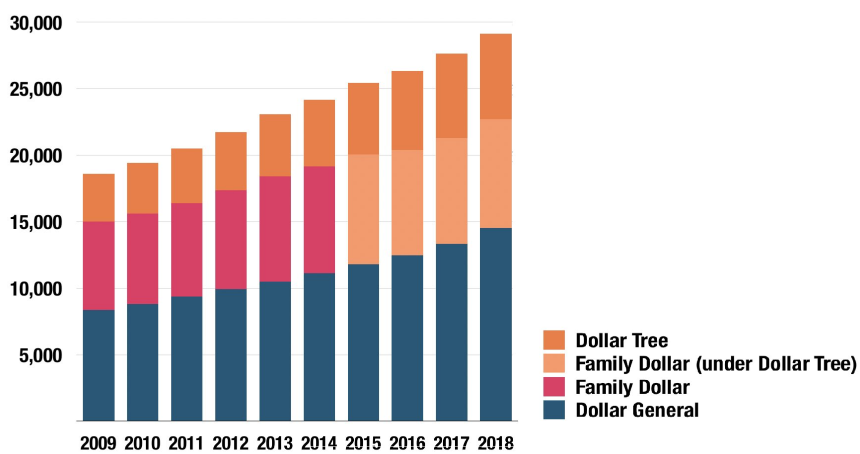

Figure 1: Dollar Store Growth from 2009-2018
Source: Institute for Local Self-Reliance

As discussed in the introduction section, dollar store is amongst the fastest growing section in the United States retail sector, and the only brick-and-mortar sector with substantial growth in the recent decade. From **Figure 1**, we can see that the sector consists of three main store brands: Dollar Tree, Family Dollar (acquired by Dollar Tree in 2014) and Dollar General, and Dollar General is growing faster than their competitors. According to Statista[6], dollar store sector receives in total $95 billion in revenue in 2021, tripled from

[6]https://www.statista.com/topics/1343/dollar-stores-in-the-us/

$30 billion in 2010. This sector currently accounts for more than 10% of total grocery retail revenue in the United States. Additionally, a June 2021 survey by *Consumer Reports* of 2,280 adults report that 9 out 10 Americans shop at dollar store at least a few times a year[7]. According to credit card spending data aggregator *Bloomberg Second Measure*, dollar store is the only segment witnessed growth during COVID-19 pandemic, with over 30% growth in sales in two years for Dollar General[8].

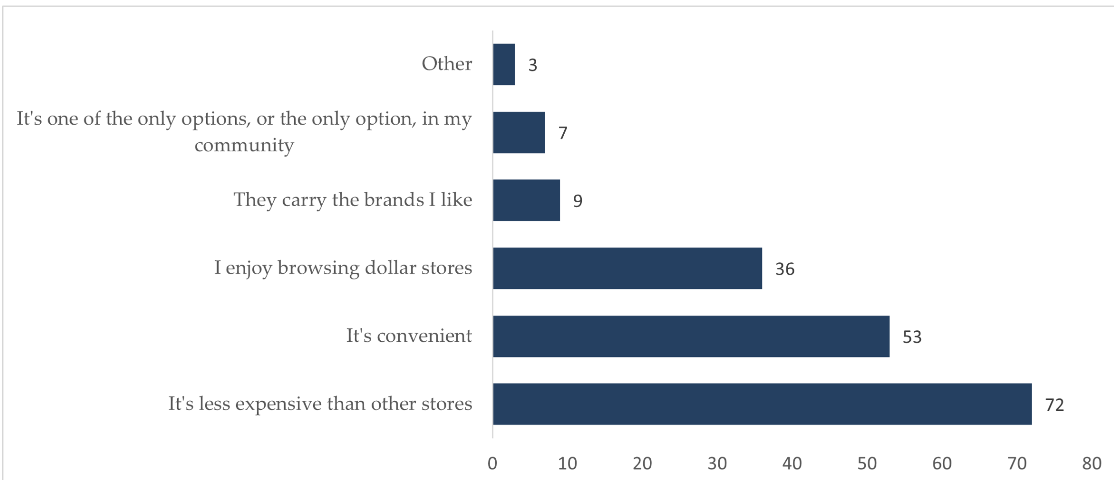

Figure 2: Survey Responses to "Why Shop at Dollar Store?"
Source: Consumer Reports

Overall, the claim that dollar stores target low-income segments and neighborhoods appear to have some validity. According to the same report mentioned above by *Bloomberg Second Measure*, the average dollar store shopper has an average income of just over $40,000. However, roughly 20% of dollar store shoppers have an income over $100,000, and more than 20% of frequent shoppers (more than once a week) have an income over $60,000. From *Consumer Reports* survey, most shoppers shop at dollar stores because they are less expensive (72 %) and more convenient (53 %). Yet, we do not know whether the same factors that influence store choice would influence the level of calories consumption, as average spending at dollar stores is also much lower in comparison to traditional grocery stores.

[7]https://article.images.consumerreports.org/prod/content/dam/surveys/Consumer$_{R}eports_{A}ES_{J}une_{2}$021
[8]https://secondmeasure.com/datapoints/dollar-store-sales-buck-covid-era-retail-trends/

2.2 Data

The main datasets for this research come from NielsenIQ through Kilts Center of Marketing at the University of Chicago. The first source is NielsenIQ's Retail Scanner Data, which contains weekly pricing, volume, and store environment information generated by point-of-sale systems from more than 90 participating retail chains (35,000-50,000 stores) across all US markets. Store location in this data is at a county-level granularity. I supplement this data with Nielsen TDLinx, a dataset of stores exact location, their characteristics, as well as their entry and exit data. Stores in both datasets can be merged using Nielsen unique store IDs. The merged data will serve as the main source of pricing and product characteristics information, store location, and entry and exit data for reduced form analysis.

The next source is the NielsenIQ's Consumer Panel Data, a panel of 40,000-60,000 households from 2004 - 2019, who, through the use of in-home scanners, record all of their purchases (from any outlet) intended for personal, in-home use. These panelists provide information about their households and what products they buy, as well as when and where they make purchases, including from stores that are not in the Scanner data. They are geographically diverse and demographically balanced, and each is assigned a projection factor, which enables purchases to be projected to the entire US population. With this dataset, we can estimate the household level demand for calories and store choices, as well as collecting moments of travel distances.

The data on calories and other nutrition facts are from the Food and Nutrient Database for Dietary Studies and the National Nutrient Database for Standard Reference, public available at the U.S. Department of Agriculture. These nutrition facts can be matched to UPCs in Nielsen datasets using crosswalks developed by the USDA (Carlson et al. 2019). Prior studies (Allcott et al. 2019) have confirmed that these facts closely match information provided by third party providers, as well as information from the manufacturers themselves.

Lastly, empirical distribution of transportation methods can be obtained through the National Household Travel Survey conducted by the Federal Highway Administration. This data include daily non-commercial travel by all modes, including characteristics of the people traveling, their household, and their vehicles. As the data also provide information on the purpose of each trip, we can use it to construct both the empirical distribution of transportation modes and the distribution of shopping trip distance by market. Additionally, I can also use data from SafeGraph, a geolocation data provider that provide daily foot traffic data to over 6,000,000 points of interest in the United States to construct trip distance distribution (as SafeGraph provides home census tracts of visitors) as well as an alternative source of store visit share data.

3 Model

3.1 Store Level Calories Demand

First, we can start with a general model of utility from food consumption, following the setup of Dubois et al. (2014) and Allcott et al. (2019). The household's food preferences are assumed to satisfied a utility function with the follow characteristics: (1) constant elasticity of substitution over calories from each product k in a product group j; (2) Cobb-Douglas preferences over J product groups; and (3) linear preferences over characteristics. It is worth noting that the first assumption does not negate product differentiation, we are just assuming that the basic calorie intake itself does not vary in utility within a group. We can write down the utility equation as follow:

$$U(x, \boldsymbol{a}, \boldsymbol{y}; \Theta, \Psi, \boldsymbol{M}) = \sum_{j=1}^{J} \mu_j \ln \left(\sum_{k=1}^{K_j} \psi_{kj} y_{kj}^{\theta_j} \right) + \sum_{c=1}^{C} \beta_c \left(\sum_{j=1}^{J} \sum_{k=1}^{K_j} a_{kjc} y_{kj} \right) + \lambda x \qquad (1)$$

In equation (1), the first summation represents total utility of calorie intakes from different goods in different product groups. The term μ_j captures the satiation rate over calories consumed in group j (i.e. one may prefer calories from meat over vegetable), while θ_j determines satiation over consuming product k in group j. The term ψ_{kj} allows for product differentiation effect, accounting for unobserved differences in product quality. The term y_{kj} stands for the amount of calories from product k in group j. The second summation represents the total utility gained from consuming characteristics in set $c \in C$, where the characteristics could be fattiness, sweetness, freshness, health implications, shelf life etc. β_c is the (per calorie) marginal utility of consuming the corresponding characteristic, while a_{kjc} is indicator of characteristic c in product k. Finally, x is an outside numeraire good, and λ is the corresponding marginal utility of money.

Consumer utility maximization problem takes the form:

$$\max_{\boldsymbol{y},x} \; U(x, \boldsymbol{a}, \boldsymbol{y}; \Theta, \Psi, \boldsymbol{M})$$
$$\text{s.t.} \; \sum_{j=1}^{J} \sum_{k=1}^{K_i} p_{kj} y_{kj} + x \leq \boldsymbol{I}$$

In which p_{kj} is the per calorie price of product k of group j, and $\boldsymbol{I}$ being the budget constraint. We can then take the first order conditions of the above constraint optimization problem, and arrive at the following equilibrium condition for each product group j:

$$\sum_{k=1}^{K_j} p_{kj} y^*_{kj} = \sum_{c=1}^{C} \frac{\beta_c}{\lambda} \sum_{k=1}^{K_j} a_{kjc} y^*_{kj} + \frac{\mu_j \theta_j}{\lambda} \tag{2}$$

To simplify notation, we can then denote $\mu_j\theta_j = \tilde{\delta}_j$, which is the inherent preference of consumer for product group j. Additionally, let $\tilde{p_j}$ be the product group average price per calorie across all products $k \in K_j$, and similarly $\tilde{a}_{jc}$ be the product group average characteristic c per calorie across all products $k \in K_j$, $\tilde{a}_{jc}$. Furthermore, let $Y_j = \sum_{k=1}^{K_j} p_{kj} y^*_{kj}$. Adding household indicator i, market indicator m, time indicator t, and store indicator s, we can rewrite (2) as:

$$\lambda \tilde{p}_{jtm|s} Y_{ijtm|s} = \sum_{c}^{C} \beta_c \tilde{a}_{jtmc|s} Y_{ijtm|s} + \tilde{\delta}_{ijtms} \tag{3}$$

Taking log of both side of (3), we arrive at the regression equation for household i calories demand (conditional on store chain$s$ in market m) from product group j at time t, conditional on the mean price and characteristics of products in j:

$$\ln(Y_{ijtm|s}) = -\ln\left(\lambda \tilde{p}_{jtm|s} - \sum_{c=2}^{C} \beta_c \tilde{a}_{jtmc|s} - \xi\right) + \ln \tilde{\delta}_{ijtms}$$

In which, ξ presents an unobserved product characteristic. We can also model household heterogeneous, time varying preference for j as a combination of their demographic characteristics and product group, time, market, and store brand fixed effects.

$$\ln \tilde{\delta}_{ijtms} = \Gamma \boldsymbol{X}_i + \eta_j + \eta_t + \eta_m + \eta_s + \varepsilon_{ijtms}$$

Combining both, we have the final regression equation:

$$\ln(Y_{ijtm|s}) = -\ln\left(\lambda \tilde{p}_{jtm|s} - \sum_{c=2}^{C} \beta_c \tilde{a}_{jtmc|s} - \xi\right) + \Gamma \boldsymbol{X}_i + \eta_j + \eta_t + \eta_m + \eta_s + \varepsilon_{ijtms} \tag{4}$$

We can see that the set of coefficients $\{\lambda, \boldsymbol{\beta}, \xi\}$ present household's sensitivity to product-level supply side factors such as prices and characteristics. If $\{\boldsymbol{\beta}, \xi\}$ are large and λ is small, this scales down the importance of price and thus lower prices offer by dollar stores may not significantly alter the calories intake if other characteristics are similar to what normal grocery stores offer. Alternatively it is possible that dollar stores exclusively offer more products with calories-inducing characteristics. In order to account for the variation in consumer budget constraint, I separately estimate the vector $\{\lambda, \boldsymbol{\beta}, \xi\}$ for each household income group as categorized by Nielsen.

3.2 Spatial Store Choices

For the upper level of my demand system, I model the consumer choice of which store to shop from using a nested logit model. Before laying down the model, we have to address the problem of defining the stores choice set of a consumer. Here, there are two main approaches I can take. The first, and simpler approach, is to rely on existing geographical or statistical subdivisions to define a market. Here, the market could be at Designated Market Area-level of multiple counties, at county-level, at ZIP Codes level, or at census tract level. Obviously, using the Designated Market Area to construct the store choice sets will be unrealistic, as these areas are often larger than what consumers normally would travel for grocery shopping. According to the USDA, on average American households travel only 3.36 miles away for a grocery shopping trip[9]. This problem persists even if I use counties as markets, as some of the larger counties can be up to 100 miles across (San Bernardino County, California for example is more than 20,000 sq. miles). On the other hand, a census block or a ZIP Code may be too small to use as a reasonable market, and some may not have any grocery stores within them at all. Alternatively, following Aguirregabiria et al. (2016) I can iteratively assign stores into markets following a Voronoi tessellation algorithm. The computational complexity of this approach, however, increases exponentially with the number of stores. Therefore, here I employ a simple heuristic for market construction: Any county that have an area smaller than that of a circle 5 miles in radius (25π sq. miles in area) will be consider a market. Any county above that threshold, I will split into parts that have roughly equal area while also trying to have equal population, and the number of subdivisions will be the lowest one so that each subdivision has an area less than 25π sq. miles.

Once we have the market definition, we can then write down the utility function of a

[9]https://www.ers.usda.gov/webdocs/publications/43953/eib138_erratasummary.pdf

household i choosing to shop at a store of chain s, in market m at time t:

$$u_{ismt} = \sum_{r}^{R} \gamma_{ir} \mathbb{1}\{r_i = r\} d_{istm} + \eta_i \boldsymbol{Z}_{stm} + \sum_{j=1}^{J} \phi_{ij} V_{jstm} + \gamma_s + \gamma_t + \gamma_m + \zeta_{ist} \tag{5}$$

Here, d_{istm} is the distance between the home address of the household and closest store of chain s, r_i the mode of transportation of the household, and γ_{ir} is the corresponding cost of traveling using that mode of transportation. This captures the "accessibility" effect of the dollar stores, as the sheer number of these stores make the distance to travel to one shorter than other types of grocery stores. Here, to simplify the model as number of stores can be extremely large in urban markets, I follow Briesch et al. (2009) and model the choice between store brands, not store, and assume that the consumer will go to the closest store of the chain of their choice. Vector $\boldsymbol{Z}_{stm}$ represents the characteristics of store chain s in market m at time t, including the number of stores, the number of employees, time since first entrance, total floor area and so on. η_i is the corresponding vector of coefficients. Next, $V_{jstm} = \sum_{c=2}^{C} \beta_c \tilde{a}_{jtmc|s} + \xi - \lambda \tilde{p}_{jtm|s}$ (with the coefficients estimated in the lower level demand) is the per calorie expected utility of product group j at store chain s, serving as the proxy for consumer knowledge and expectation about product offerings at the store. Finally, $\gamma_s, \gamma_t, \gamma_m$ are the sets of store chain, time, and market fixed effects to account for brand reputation effect, time invariant unobserved factors of the market, as well as time varying shocks to demand, and ζ_{istm} is heterogeneous difference in utility.

With the utility function in hand, I model the store choice problem using a Nested Logit framework, with dollar store brands in one nest, supermarkets in another nest, supercenters in another, and so on. This allows for more flexible substitution patterns and correlation between choices in the same nest, which is more realistic. This assumes the random utility component can be decomposed to:

$$\zeta_{istm} = \zeta_{igtm} + (1 - \rho)\tilde{\zeta}_{istm}$$

Where $\rho \in [0, 1)$ is a nesting parameter, and ζ_{igt} is the group level heterogeneous preference, $\tilde{\zeta}_{ist}$ is the store brand level heterogeneous preference, which is i.i.d. Type I extreme value. Let:

$$\tilde{u}_{ismt} = \sum_{r}^{R} \gamma_{ir} \mathbb{1}\{r_i = r\} d_{istm} + \eta_i \boldsymbol{Z}_{stm} + \sum_{j=1}^{J} \phi_{ij} V_{jstm} + \gamma_s + \gamma_t + \gamma_m$$

We can then model the store choice as:

$$\pi_{istm}(\boldsymbol{V}_{stm}, \boldsymbol{\theta}_i, \boldsymbol{Z}_{stm}) = \frac{\exp((\tilde{u}_{ismt})/(1-\rho))}{\exp(I_{ig}/(1-\rho))} \frac{\exp I_{ig}}{\exp I_i} \tag{6}$$

Where I_{ig} is the group specific "inclusive values" and I_i is the overall inclusive value, defined as:

$$I_{ig} = (1-\rho) \sum_{s=1}^{S_g} \exp\left(\frac{\tilde{u}_{ismt}}{(1-\rho)}\right)$$
$$I_g = 1 + \sum_{g=1}^{G} \exp I_{ig}$$

With individual data, this nested logit model can be jointly estimated directly with the lower level demand equation using Generalized Method of Moments. With aggregated data, however, we have to rely on simulation to account for heterogeneity, estimating a Random Coefficients Nested Logit model (Grigolon and Verboven 2014). Group $\{\boldsymbol{Z}_{stm}, \boldsymbol{V}_{stm}\} = \boldsymbol{X}_{stm}$ and the corresponding coefficients to a vector $\boldsymbol{\theta}_i$ to simplify notation, I assume that $\boldsymbol{\theta}_i \sim \mathcal{N}(\bar{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$. Let:

$$\delta_{stm} = \boldsymbol{X}_{stm}\bar{\boldsymbol{\theta}} + \gamma_s + \gamma_t + \gamma_m + \varepsilon_{stm}$$

Where $\bar{\boldsymbol{\theta}}$ are the mean coefficients. Let $\boldsymbol{\nu}_i \sim \mathcal{N}(0, 1)$ be the random parts, $\boldsymbol{\Sigma}$ be the variance covariance matrix which can be Cholesky decomposed into $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T$. We can then write

$\tilde{u}_{istm}$ as:

$$\tilde{u}_{istm} = \delta_{stm} + \boldsymbol{X}_{stm}\boldsymbol{L}\boldsymbol{\nu}_i + \sum_{r}^{R} \gamma_{ir} \mathbb{1}\{r_i = r\} d_{istm} \tag{7}$$

The transportation cost components d_{istm} and $\mathbb{1}\{r_i = r\}$ are also drawn from empirical distributions obtained from American Travel Survey and Safegraph Data. Plugging (7) into (6) and aggregating over the random distributions, we get the equation for visit share of a chain s in market m at time t:

$$\mathcal{S}_{stm}(\delta_{stm}, \bar{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) = \sum_{r}^{R} \int_{\nu} \int_{d} \pi_{istm}(\delta_{stm}, \boldsymbol{\Sigma}, \bar{\boldsymbol{\theta}}, \boldsymbol{\nu}_i) dF(\boldsymbol{\nu}_i) dF(d_{istm}) P(r_i = r) \tag{8}$$

3.3 Instrument

As with most demand estimation exercises, endogeneity between prices (here are the average prices per calorie of each product group) and household's idiosyncratic preference for the product group, or from simultaneity bias, is an important issue that we have to address. As we can using average prices of the whole category, the traditional "BLP instruments" of competitors' product characteristics (Berry et al. 1995) are not feasible. Additionally, as the defined "market" in this study is relatively small, Hausman (1997) leave-one-out instruments that exploit variation in prices across time in other markets are also likely invalid, as demand shocks can easily affect multiple markets at a time.

Therefore, in this research I follow Allcott et al. (2019) and construct the instruments for category prices using comparative advantages of chain s in supplying goods in a product category. Let $\Delta \ln(p_{krt,-m}) = \ln(p_{krt,-m}) - \ln(p_{kt,-m})$ be the difference in average log price of product k at retailer r at time t in markets other than m and the average log price of k across all retailers nationwide, N^{store}_{rmt} be the number of stores of r in m at t, Sales$_{jrt}$ be the average sales of j at a store of r in time t, we have:

$$IV_{jtms} = \frac{N_{rmt}^{store}\text{Sales}_{jrt} \sum_{k=1}^{K_j} \Delta \ln(p_{krt,-m})}{N_{rmt}^{store}\text{Sales}_{jrt}}$$

The identifying variation in this instrument comes from the interaction between a retail chain pricing advantages $\sum_{k=1}^{K_j} \Delta \ln(p_{krt,-m})$ and their differing presence across geographic markets $N_{rmt}^{store}\text{Sales}_{jrt}$. The rationale of this instrument comes from the fact that since different products are produced in different parts of the country, transportation costs vary across retail brands in different regions. Additionally, the increasing ubiquity of private label products in retail (Dubé et al. 2018) means retail chains will have differing cost advantages in certain product categories that are produced by themselves or through economies of scales via contracting private label manufacturers. For example, Dollar General has around 40 private label brands with the top selling brand brings in more than $1 billion in sales annually, often at half the prices of branded products.

3.4 Estimation

With the instruments in hand, we can move on to the estimation step. As briefly discussed above, if the upper level store choice model is estimated with individual level data, we can simply jointly estimate both the upper level and the lower level model with Generalized Method of Moments estimator using the following sets of moment conditions:

$$\begin{aligned}
E(\varepsilon_{ijtms}\boldsymbol{IV}_{tms}) &= 0 \\
E((\eta_j + \varepsilon_{ijtms})\tilde{a}_{jtmc|s}) &= 0 \\
E(\varepsilon_{ijtms}\boldsymbol{\eta}) &= 0 \\
E(\zeta_{istm}\boldsymbol{IV}_{tms}) &= 0 \\
E(\zeta_{istm}\boldsymbol{\gamma}) &= 0 \\
E(\zeta_{istm}\varepsilon_{ijtms}) &= 0
\end{aligned}$$

With $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$ being the stacked vectors of fixed effects in the lower and upper level demands equation respectively. With Random Coefficients Nested Logit model, we can followed a modified version of Berry et al. (1995)'s Nested Fixed Points algorithm, with the above GMM problem being the outer loop, with the inner loop a "dampened" version of the original BLP contraction mapping:

$$\delta_{tm}^{k+1} = \delta_{tm}^{k} + \ln(\mathcal{S}_{tm}) - (1-\rho)\ln(\mathcal{S}_{tm}(\delta_{tm}^{k}))$$

Additionally, given the large number of possible markets of this problem, it is probably worth exploring reformulating the problem as a mathematical program with equilibrium constraints following Dubé et al. (2012), which the authors claim can reduce the computational complexity significantly for many-market problems.

4 Counterfactual and Other Extensions

Once I obtain the parameters of the demand system, it is straightforward to perform counterfactual analyses using the estimated model. The main counterfactual experiment of interest is simulation of a local dollar store ban by removing dollar stores from the store choice sets. Here, I assume that this ban will not affect prices at other stores in the same markets, as recent literature (e.g. DellaVigna and Gentzkow 2019, Hitsch et al. 2019, Stroebel and Vavra 2019) have documented that "most US food, drugstore, and mass merchandise chains charge nearly-uniform prices across stores, despite wide variation in consumer demographics and competition". Therefore, it is plausible that the removal of dollar stores from one local market may not affect pricing at other local grocery outlets. After removing dollar stores from the choice sets, we can simply use the estimated model to see how average aggregated calories consumption overall all stores changes for households in the affected markets, by re-estimating store shares using the upper level demand, and then calculate the sum of product of store shares and store-specific expected calories consumption. We can perform this calculation separately for each income group. With this, we can evaluate whether a dollar store ban truly reduces calories intake of local residents, or the calories intake is simply a function of "sticky" eating habits and other food preferences.

Even though I assume that there would not be significant supply side responses to the removal of dollar stores, this assumption can be relaxed by incorporating supply side model. A simple supply side set up with constant marginal cost of selling one additional calorie of product category j is:

$$\pi_{st}(\mathbf{p}_{st}) = \left(\sum_{j \in \mathbf{J}_s} (\tilde{p}_{jts} - mc_{jts}) Y_{jts} \right) \mathcal{S}_{st} M$$

The first order condition for a product category j is then:

$$Y_{jts}\mathcal{S}_{st} + \sum_{k \in \mathbf{J}_s} (\tilde{p}_{kts} - mc_{kts}) Y_{kts} \frac{\partial \mathcal{S}_{st}}{\partial \tilde{p}_{jts}} = 0$$

We can then solve this system of equations for the marginal cost and then simulate the price changes so that the first order condition hold when $\mathcal{S}$ changes due to the dollar store ban.

Finally, another extension that may be worth considering is to incorporate endogenous entry and exit decisions of the stores, and their location choices. This can account for concerns about dollar stores actively targeting lower income neighborhoods, and also help with simulating the effect of "new dollar store moratorium" type of legislation, which does not outright ban dollar stores but prohibit openings of new stores in the local market.

5 Conclusion

To summarize, in this research I estimate a structural demand model to address the question of whether a ban on dollar stores would be effective in reducing nutritional inequality and help local residents eat healthier. This research helps settling an important policy question, and also contribute to the nascent stream of literature on sources of nutrition inequality, “food desert”, and the factors that affect healthy eating habits. Through a two-level store choice and calories consumption demand system, I decompose household’s preferences for price, product characteristics, and convenience. Using counterfactual simulations, I can see how a ban on dollar stores affect total calories consumption, in both scenarios when there are no significant competitor price responses and with competitors’ responses. The approach here can also be extended to include endogenous entry and exit of retail stores, and can serve as a basic framework to analyze future dollar store related policies.

References

Aguirregabiria, V., Ershov, D., and Suzuki, J. (2016). Estimating the effect of deregulation in the ontario wine retail market. Report, University of Toronto.[1653, 1674].

Allcott, H., Diamond, R., Dubé, J.-P., Handbury, J., Rahkovsky, I., and Schnell, M. (2019). Food deserts and the causes of nutritional inequality. The Quarterly Journal of Economics, 134(4):1793–1844.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. Econometrica: Journal of the Econometric Society, pages 841–890.

Briesch, R. A., Chintagunta, P. K., and Fox, E. J. (2009). How does assortment affect grocery store choice? Journal of Marketing research, 46(2):176–189.

Carlson, A. C., Page, E. T., Zimmerman, T. P., Tornow, C. E., and Hermansen, S. (2019). Linking usda nutrition databases to iri household-based and store-based scanner data. Technical report.

Courtemanche, C. and Carden, A. (2011). Supersizing supercenters? the impact of walmart supercenters on body mass index and obesity. Journal of Urban Economics, 69(2):165–181.

Courtemanche, C., Carden, A., Zhou, X., and Ndirangu, M. (2019). Do walmart supercenters improve food security? Applied Economic Perspectives and Policy, 41(2):177–198.

DellaVigna, S. and Gentzkow, M. (2019). Uniform pricing in us retail chains. The Quarterly Journal of Economics, 134(4):2011–2084.

Dubé, J.-P., Fox, J. T., and Su, C.-L. (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. Econometrica, 80(5):2231–2267.

Dubé, J.-P., Hitsch, G. J., and Rossi, P. E. (2018). Income and wealth effects on private-label demand: Evidence from the great recession. Marketing Science, 37(1):22–53.

Dubois, P., Griffith, R., and Nevo, A. (2014). Do prices and attributes explain international differences in food purchases? American Economic Review, 104(3):832–67.

Grigolon, L. and Verboven, F. (2014). Nested Logit or Random Coefficients Logit? A Comparison of Alternative Discrete Choice Models of Product Differentiation. The Review of Economics and Statistics, 96(5):916–935.

Hausman, J. A. (1997). Valuation of new goods under perfect and imperfect competition. The economics of new goods, 58:209–37.

Hitsch, G. J., Hortacsu, A., and Lin, X. (2019). Prices and promotions in us retail markets: Evidence from big data. Technical report, National Bureau of Economic Research.

Hut, S. and Oster, E. (2022). Changes in household diet: Determinants and predictability. Journal of Public Economics, 208:104620.

Larson, N. I., Story, M. T., and Nelson, M. C. (2009). Neighborhood environments: disparities in access to healthy foods in the us. American journal of preventive medicine, 36(1):74–81.

Stroebel, J. and Vavra, J. (2019). House prices, local demand, and retail prices. Journal of Political Economy, 127(3):1391–1436.