Healthiness of Consumption and Retail Environment

Oleksii Khvastunov

Department of Economics Penn State University

WORKING PAPER

DO NOT CITE WITHOUT AUTHOR'S PERMISSION

Abstract

The determinants of a households diet decision have been studied extensively. This paper contributes to the literature by focusing on how store type, store size, household income, and distance to store affects healthiness of household consumption. I am implementing a micro-founded structural model which includes grocery budget allocation across stores and food types in the vicinity of household location. The micro-founded model allows me to capture the rich heterogeneity in the retail environment that a household is facing. The estimated model permits me to conduct counterfactual experiments that suggest that policies targeting travel costs and product variety would have larger effects on household consumption decisions than policies that provide additional income to households. This result is even more drastic for low income households living far from large stores.

Keywords: Store Choice; Scanner Data; Food Access; Food Deserts;

JEL Classification: I00, I14, I32, P46, Q18, Z18

Disclaimer: Any opinions, findings, recommendations, or conclusions are those of the author 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 should also not be attributed to either Nielsen or Information Resources, Inc. (IRI).

Acknowledgement:I am grateful for the insightful comments provided by Paul Grieco, Lauren Chenarides, Ted Janicke, Kala Krisha, Charles Murry, Mark Roberts, Martin Hackman and Daniel Grodzicki. Please send correspondence to `ouk111@psu.edu` .

1 Introduction

A households diet decision and the factors that determine it have attracted the attention of both health-care and agricultural economics researchers. The reason this interdisciplinary question has been so intensively studied is that it makes possible evaluation of policy interventions result which have a goal to improve consumption patterns of households. Answering this question involves solving two problems: identifying factors that affect households choice of diet and uncovering the relative importance of these factors. In this paper I am concentrating on identifying the relative importance of household income, distance to store, and store size for healthiness of household consumption. Determining the influence of these factors will help to evaluate how effective different government policies can be that try to improve consumption habits. My focus is on how important these factors would be for households with different demographics. I am particularly interested in how influential are: household income, distance to store, and store size for low income households who live far from the store. These are the households which are: - mostly living in food deserts - low income areas with few available stores - eating a less healthy diet, and targeted by government policies.

The difference in the relative impact of factors would result in different policy implications. For example, if distance to the grocery store plays a larger role in household choice than other factors, then policies which focus on improving public transportation or entry of new stores should be chosen. If household income plays a larger role than other determinants, then expanding existing income-assisting programs like SNAP (Supplementary Nutrition Assistance Program) would be a more efficient policy. If store size plays a larger role in healthiness of household consumption than other aspects, then policies which expand product variety in the existing stores should be considered. For example, policy which would put higher requirements on product variety for retailers who are eligible for redeeming SNAP coupons should be implemented.

Numerous studies have related poor diet to bad health outcome which in turn has a direct economic impact in the form of direct medical, productivity, and human capital costs. Productivity costs include employees being absent or unproductive at work. Poor health conditions are also related to effectiveness of schooling and human capital accumulation. As a result, determining the main reasons for a poor diet will help to construct effective policies which would improve health outcomes and reduce the economic impact associated with bad health.

In order to understand the main contributing factors of a household's poor diet, I implemented a structural model of grocery budget allocation across stores and food categories (healthy and unhealthy) in the vicinity of a household's location. The food categories are defined based on The Center for Nutrition Policy and Promotion's (CNPP) classification of products into the ones which are recommended for limited or increased consumption. This structural model will be able to precisely tract all available grocery options

and capture heterogeneity in the choice set. I am able to do that because of a rich TDLinx dataset which has information on practically all stores that sell groceries within the United States. I also utilized an IRI dataset, which provided information on household grocery purchases. Combining the TDLinx dataset with the IRI dataset on household grocery purchases allowed me to see how much households spend in each grocery chain in the vicinity. In turn, this made it possible to estimate a structural model.

The microfounded model allowed me to see how each store's characteristics affects a household's choice of allocating grocery budget across stores and across healthy/unhealthy products within stores. In particular, I am interested in the effect of store format, distance to store, household income, and store size on the share of healthy consumption in household grocery budget. The importance of these factors for household consumption decision was discussed in the literature.

As an example of these effects, the impact of distance was studied in Chen et al. (2010), who established that the presence of a nearby grocery store is linked with lower body mass index (BMI). Laraia et al. (2004) found that as distance to the supermarket increases quality of diet decreases for low-income pregnant woman. Rose and Richards (2004) and Morland et al. (2002) find relationship between density of grocery stores and consumption of fruits and vegetables. Rahkovsky et al. (2015) found that low-income households who are living far from the large supermarkets tend to have a less healthy diet. Other studies, such as (Kyureghian et al. (2013)), discovered no relationship between store density and healthiness of consumption. As a result, the actual effect of distance on household consumption is still debatable (Bitler and Haider (2011)).

The effect of store format on healthiness of household consumption was extensively studied in the literature. Volpe et al. (2013) found that households make less healthy purchases at supercenters than supermarkets. Andreyeva et al. (2008) and Liese et al. (2007) established that convenience stores have smaller availability of healthy products than grocery stores. Yoo et al. (2006) discovered that consumers who frequently shop in convenience stores make less frequent big shopping trips to large grocery stores.

The effect of income on a household's consumption is consistent in the literature. For example, Handbury et al. (2015) and Chen et al. (2016) show that higher income households consume more healthy products than low income households.

Store size also plays an important role in household consumption decisions because it reflects a store's quality and product assortment (Ellickson (2006)). In turn, higher product assortment increases the likelihood of shopping in the store Chenarides et al. (2016). Bonanno and Lopez (2009) found that larger supermarket sizes result in higher chain level demand. Laska et al. (2010) established that healthy food availability increases significantly with the store size.

This paper is different from the aforementioned papers in that it takes all of the previously discussed factors into account and captures the choice set of a household in a very precise way. In addition, this paper

uses geographically rich data which contains information on households from all states but Hawaii. These differences, combined with using a microfounded structural model, allowed me to measure the importance of these factors for households with different demographics. In particular, I studied the importance of these factors for low-income households who are living far from large grocery stores. As a result, I found that, for these households, distance to the store and store size matter more than income. This difference is consistent across areas with different population densities. However, distance to the store - which reflects travel cost - and store size - which reflects healthy food availability - play an even more important role for rural households than for urban households.

The paper is organized as follows: Section 2 describes in detail the data sources that I used, Section 3 shows evidence that distance to store, household income, and store size affect healthiness of household consumption, Section 4 presents the structural model I used to estimate the parameters that govern household grocery budget allocation, Section 5 contains the results of the structural model, Section 6 contains the result of the counterfactual experiment of decreasing distance to all large grocery stores by 10 %, increasing household income by 10 % and increasing large grocery stores size by 10 %. This experiment helps to understand what is relative importance of these factors for households with different demographics. Section 7 provides the conclusion drawn from my research.

2 Data

I used three main data souses in the empirical exercise:

- The Information Resources, Inc. (IRI) household store scanner data, and the Consumer Network Panel (CNP) for the calendar year 2013.
- Neilsen TDLinx Store Characteristics Data for calendar year 2013.
- 2010 U.S. Census.

The IRI Consumer Network Panel contains information on the purchasing behavior of more than 200,000 households. Household demographic information is taken from the National Consumer Panel which is the joint work of IRI and Nielsen. Households in the panel report shopping trips they make. Information from the trips include the chain where purchases were made, format of the chain, UPC (Universal Product Code) of products purchased, product price, and date of purchase. Chain formats include supermarket, supercenter, mass merchandise, pharmacy, grocery, club, convenience and dollar stores. Households in the panel are using a scanner to collect this information and may not participate through the period of entire year in the Panel. Only the households which sufficient participation in the Panel were considered in my examination. To be

more precise, I included households which had shopping trips for at least 8 months and spent at least 40 dollars on groceries each month. This constraint allowed me to filter out households which did not participate in enough month or may not have been accurate reported. Finally, 41,856 households were used for a reduced form exercise and structural model. It is also important to note that these households represent all states but Hawaii.

Another important feature of the IRI Consumer Network Panel data which I need to emphasize is that consumer reports chain where they did shopping. This means that if a household shopped in store A and store B of the same chain, I cannot distinguish purchases in store A from the ones in store B. This feature of the data will be taken into account for the structural model by constructing store level shares for households, aggregating these shares to the chain level, and then matching chain level shares to the ones observed in the data.

Nielsen TDLinx Store Characteristics Data contains information for more than 200,000 food stores operating in the United States. This dataset includes store format, size, and geographical location. Nielsen TDLinx data is matched with the IRI Consumer Network panel using the chain string name. This allows me to see all the stores within the vicinity of a household's location and conclude where the household did its grocery shopping.

Information on census tracts' population, location, and area is taken from the 2010 US Census dataset. This information is used to construct population density within a 5 mile radius from the centroid of the census tract where a household is living. This density is used to parametrize of outside option in the model, which is household consumption in the mass merchandise, pharmacy, grocery,convenience and dollar stores.

Another piece of data that I am using is an UPC classification within the healthy/unhealthy category. This classification is taken from Volpe and Okrent (2013) and is based on The Center for Nutritional Policy and Promotion's (CNPP) categorization of products. CNPP categories are divided into healthy/unhealthy groups based on the Dietary Guidelines for Americans 2010 (developed by USDA and Health and Human Services) and are using the fact if the products in the category recommended for increased or limited consumption.

Table 1 contains the summary statistics of the households' demographics and area characteristics that I am using in the analysis. I excluded households which are living in areas with a population density more than 5,000 people per square mile from the analysis because due to enormous choice set size that households face in these areas. Information about distance to the closest large store is presented because these stores account for the majority of grocery purchases (Table 2 shows that large stores receive, on average, more than 87% of grocery purchases) and play an important role in a household's grocery budget allocation decision. Large are defined as stores which have more than $2 million yearly grocery sales in 2013.

	Mean	St. Dev.	1st Quartile	Median	3rd Quartile
Household and Area Characteristics					
Household Size	2.44	1.29	2	2	3
Household Income	68,289	46,667	30,000	55,000	85,000
Income per Household Member	30,487	24,468	14,167	22,500	42,000
Area Characteristics					
Density in 5 mile radius, people per sq. mile	1,508	1,390	275	1,022	2,567
Distance to the Closest Large Store, miles	2.06	2.16	0.67	1.17	2.57

Table 1: Household Demographics

	Mean	St. Dev.	1st Quartile	Median	3rd Quartile
Grocery Spending					
Grocery Spendings	2,192	1,289	1,249	1,910	2,820
Grocery Spendings per household member	1,048	653	584	925	1,367
Grocery Share of Income, %	5.81	8.03	2.13	3.81	6.62
Share of Healthy Food in Groceries, %	30.21	11.31	22.79	29.65	37.10
Number of stores with shares $> 5\%$	3.14	1.28	2	3	4
Large Stores					
Share of Large Stores in Groceries, %	87.22	16.50	83.52	93.47	98.07
Total Number of Households 41,856					

Table 2: Household Consumption Characteristics

Table 2 present the summary statistics of households' grocery purchases. Grocery purchases are presented for the entire 2013 calendar year in order to reflect average healthiness of consumption and get rid of temporary factors which can effect household decisions, such as store discounts. There is quite a large heterogeneity in the share of healthy food purchases from 22.79% for the first quartile of the distribution to 37.10% for the third quartile. Another interesting point that Table 2 presents is that households spend their grocery budgets in more than one store, which gives rise to the distribution of grocery spending. I fit this grocery budget allocation using a microfounded structural model.

Table 3 presents a summary of household grocery spending across store formats. There are a total 8 formats in the IRI data - grocery, supercenters, club, convenience, mass merchandise, pharmacy, dollar, and other. The latter five formats and small grocery stores account for only a small portion of grocery spending (12.63% on average) and will be combined together for analysis. Large grocery, supercenters, and club stores are classified as large stores (more than $2 million yearly grocery sales) and in addition serve more healthy food than other stores.

Channel	Mean Share, %	Std Share,	Healthy Share, %	Healthy Share std,	Households > 5%
Large Grocery	60.77	29.16	34.41	12.46	94.86
Supercenters	18.90	25.90	30.26	18.35	52.70
Club Stores	7.70	14.43	39.78	23.16	32.01
Other	12.63	16.60	26.75	22.78	56.80

Table 3: Household Spendings Across Store Formats

3 Stylized Facts and Reduce Form Evidence

In this section I will provide stylized facts and reduced form evidence that store size, household income, and distance to large stores affect healthiness of consumption. First, I am going to provide evidence for the pairs of factors which will allow me to present them in table form. After that, I am going to combine all the factors together through regression analysis to show that store size, household income, and distance to the large stores have an impact on a household's consumption decision. The reduced form exercise is able to capture some of the heterogeneity that consumers are facing in terms of retail environment. However, in order to better understand the effect of retail environment on household consumption decisions I will also need to take into account each individual store where a household can shop. This approach, combined with the ability to observe how much household spends in each store, allow me to identify why households shop in particular stores and what governs their decisions within a store. As a result, a microfounded model will provide a better understanding of the relative importance of factors that drive household grocery budget allocation. This model will be presented in Section 5.

Table 4 shows the median share of healthy consumption across households with similar income levels and distances to the closest large store. The table provides evidence that for all income levels a greater distance to a large store is associated with less healthy consumption. In the same manner, for households with similar distances to a large store, higher income is correlated with more healthy consumption.

Table 5 presents the median share of healthy consumption across households with similar income levels and who do shopping in stores of a similar size. This table provides evidence that larger store size is linked with more healthy consumption for all income levels. Average store size is calculated as a weighted sum of store sizes where a household is shopping and weights are based on grocery budget share, which is allocated to the store. Table 4 provides evidence that distance to a large store and household income affect consumption decision while Table 5 shows similar evidence for store size. Now I will combine all these three factors together using regression analysis. In addition, two more factors were added that affect household choice. The first is population density in the area where household lives. More dense areas will be capturing factors such as access to a better variety of stores. The second factor is household size, which capture the fact that

Distance to the store, miles		Annual Income, $1,000 Low (<40)	Medium (40≤ and <70)	High (70≤)	Total
Small (< 1.5)	Healthy share	0.276	0.298	0.324	0.300
	Density	1,921	2,088	2,347	2,130
	Number of HH	7,590	9,013	8,274	24,877
Medium (1.5 ≤ and <2.5)	Healthy share	0.272	0.296	0.318	0.299
	Density	593	682	835	704
	Number of HH	1,735	2,245	2,105	6,085
Large (2.5 ≤)	Healthy share	0.266	0.286	0.306	0.286
	Density	105	128	167	130
	Number of HH	3,427	4,211	3,256	10,894
Total	Healthy share	0.272	0.295	0.318	0.296
	Density	866	946	1,290	1,022
	Number of HH	12,752	15,469	13,635	41,856

Table 4: **Median Share of household healthy consumption across distances to the closest supermarket and income levels**

Distance to the store, miles		Annual Income, $1,000 Low (<40)	Medium (40≤ and <70)	High (70≤)	Total
Small (<25)	Healthy share	0.233	0.256	0.279	0.254
	Number of HH	3,633	3,727	2,878	10,238
Medium (25 ≤ and <45)	Healthy share	0.280	0.295	0.315	0.297
	Number of HH	5,219	6,387	5,374	16,980
Large (45 ≤)	Healthy share	0.297	0.314	0.340	0.319
	Number of HH	3,900	5,355	5,383	14,638
Total	Healthy Share	0.272	0.295	0.318	0.296
	Number of HH	12,752	15,469	13,635	41,856

Table 5: **Median Share of household healthy consumption across stores' size and income levels**

households with kids try to eat more healthy.

Table 6 provides regression results of the share of healthy consumption on household and retail environment characteristics. One of the interesting results is that adding regressors largely impacts the effect of distance on household consumption. This result is used to illustrate the point that omitting one of the factors that affects household consumption decisions will change the conclusion on the relative importance of these factors.

Section 4 will present the structural model which takes into account household income, distance to store, store size, and store format. The microfounded model will allow me to see how each individual store's characteristics affects the consumption decision of households with different income levels. These features of the model will permit me to run an exercise of how important distance to store and store size are for households with different income levels. This exercise will be presented in Section 6.

	(1)	(2)	(3)	(4)
Const	30.978	30.460	28.316	23.901
	(0.167)	(0.168)	(0.345)	(0.483)
Distance to the Closest Large Store	-0.415	-0.391	-0.315	-0.208
	(0.069)	(0.068)	(0.071)	(0.082)
Log of Household Income		2.340	2.305	2.470
		(0.135)	(0.135)	(0.140)
Average Store Size			0.592	0.591
			(0.143)	(0.143)
Population Density				0.448
				(0.205)
Household Size				0.453
				(0.085)
R2	0.003	0.029	0.031	0.034
Number of obs.	41,856	41,856	41,856	41,856

Table 6: Healthy Food Share %, Regression Results

4 Structural Model

This section presents my structural model which includes parameters that represent consumers' preferences over healthy and unhealthy food categories. These parameters will also include distance to the store, which will help us to see how retail environment affects consumers' choices.

Consumer i who is living in tract t obtains the following utility from spending one unit of expenditures e on category c in store j that belong to chain s.

$$\begin{aligned} u_{cs_ji(t)e} &= u_{cs_ji(t)e} + \epsilon_{cs_ji(t)e} = \\ &= \beta_{cf(s_j)0} + \beta_{cf(s_j)1} * \log(inc_i) + \\ &+ \alpha_{cf(s_j)0} * dist_{s_jt(i)} + \alpha_{cf(s_j)1} * dist_{s_jt(i)} * \log(den_{t(i)}) + \alpha_{cf(s_j)2} * \mathbf{1}\{i \in T\} dist_{s_jt(i)} \\ &+ \gamma_{cf(s_j)0} \log(size_{s_j}) + \epsilon_{cs_ji(t)e} \end{aligned} \tag{1}$$

where

$f(s_j)$ - format of the store j (which can be grocery, supercenter or club) that belongs to chain s, inc_i - income of the household i, $dist_{s_jt(i)}$ - distance between store s_j and centroid of the census tract $t(i)$ where household i is living, $den_{t(i)}$ - population density in the area where household i is living ,T - set of households which might have different from other's preferences over distance for particular format-category pair, $size_{s_j}$ - size of the store s_j in thousands of square feet.

In order to reduce the number of parameters in the model, restrictions on the parameters will be imposed

such that household utility becomes:

$$\begin{aligned} u_{cs_ji(t)e} &= u_{cs_ji(t)e} + \epsilon_{cs_ji(t)e} = \\ &= \left(\beta_{hf(s_j)0} + \beta_{u0}\mathbf{1}\left\{c = unhealthy\right\}\right) + \left(\beta_{hf(s_j)1} + \beta_{u1}\mathbf{1}\left\{c = unhealthy\right\}\right) * \log(inc_i) + \\ &+ \left(\alpha_{cf(s_j)0} + \alpha_{u0}\mathbf{1}\left\{c = unhealthy\right\}\right) * dist_{s_jt(i)} + \alpha_{c1} * dist_{s_jt(i)} * \log(den_{t(i)}) + \\ &+ \alpha_2 * \mathbf{1}\left\{i \in T\right\} dist_{s_jt(i)} + \gamma_{c0}\log(size_q) + \epsilon_{cs_ji(t)e} \end{aligned} \tag{2}$$

The consumer i baseline utility from spending on category c in store j of chain s is $u_{i(t)s_jc}$, which is a function of distance from store s_j to the centroid of a census tract t where consumer i is living. The baseline utility also depends on store s_j characteristics, including size and format, and consumers' characteristics, including consumers' income. Each unit of expenditure is subject to idiosyncratic shock $\epsilon_{i(t)s_jce}$ which follows a Generalized Extreme Value Distribution (GEV).

Consumer can also allocate his grocery expenditures on outside good, which includes small stores that mostly sell unhealthy food.

$$u_{0i(t)e} = u_{0i(t)} + \epsilon_{0i(t)e} = \gamma_0 + \gamma_1\log(den_t) + \epsilon_{0i(t)e} \tag{3}$$

As a result, for each unit of expenditures e the probability that it will be allocated to category c in store s_j will be:

$$p_{cs_ji(t)} = \frac{e^{\frac{\bar{u}_{cs_ji(t)}}{\lambda_{N(cs_j)}}}\left(\sum\limits_{\{k,q\}\in SC_{N(cs_ji(t))}} e^{\frac{\bar{u}_{kqi(t)}}{\lambda_{N(cs_j)}}}\right)^{\lambda_{N(cs_j)}-1}}{e^{\bar{u}_{0i(t)}} + \sum\limits_{n\in N(t)}\sum\limits_{\{k,q\}\in SC_{N(n)}}\left(e^{\frac{\bar{u}_{kqi(t)}}{\lambda_n}}\right)^{\lambda_n}} \tag{4}$$

where $\lambda_{N(cs_j)}$ - parameter of the nest where consumption of category c in store s_j belongs to, $SC_{N(cs_ji(t))}$ - set of store-category pairs in the neighborhood of tract t where category c in the store s_j belongs to, $N(t)$ - nests in the neighborhood of tract t.

Organizing format-category pairs into nests helps to take into account that household have taste socks which are correlated within nest and that household will more likely to substitute to products within same nest.

We observe household spending on the chain level, so I aggregated the predicted shares to the chain level as well. The probability of spending the unit of expenditures in category c at chain s will be:

$$p_{csi(t)} = \sum_{s_j\in S_{st}} p_{cs_ji(t)} \tag{5}$$

where S_{st} - set of stores in the neighborhood of tract t that belong to chain s.

Denote Q_i - grocery budget of consumer i and Q_{csi} - actual expenditures of consumer i on category c at the chain s. If the unit of expenditure is one dollar, then the likelihood of the actual distribution of expenditures given parameters $\theta = (\alpha, \beta, \gamma, \theta)$ will be:

$$L_i\left(Q_{csi}|\theta\right) = \frac{Q_i!}{\prod\limits_{s\in S_t}\prod\limits_{c\in C} Q_{csi}!}\prod_{s\in S_t}\prod_{c\in C}\left(p_{csi(t)}(\theta)\right)^{Q_{csi}} \tag{6}$$

where S_t - is the set of chains in the neighborhood of tract t.

When size of unit of expenditures approaches 0 the likelihood of actual distribution of grocery budget will become dependent on expenditure shares but not dollar amount of expenditures. As a result, if I denote $q_{csi} = Q_{csi}/Q_i$ - expenditure shares of consumer i on category c at the chain s, then the likelihood of the actual distribution of expenditures of household i across all chain-categories in the vicinity becomes:

$$L_i\left(q_{csi}|\theta\right) = \prod_{s\in S_t}\prod_{c\in C}\left(p_{csi(t)}(\theta)\right)^{q_{csi}} \tag{7}$$

As a result, the likelihood function becomes:

$$L\left(q_{csi}|\theta\right) = \prod_{i\in I}\prod_{s\in S_t}\prod_{c\in C}\left(p_{csi(t)}(\theta)\right)^{q_{csi}} \tag{8}$$

The log-likelihood function is:

$$\text{Log L}\left(q_{csi}|\theta\right) = \sum_{i\in I}\sum_{s\in S_t}\sum_{c\in C}\left(q_{csi}\ \log\left(p_{csi(t)}(\theta)\right)\right) \tag{9}$$

Where parameters of the structural model are obtained by maximizing the log-likelihood function with respect to parameters $\theta = (\alpha, \beta, \gamma, \theta)$.

5 Structural Model Results

This section contains the results of the structural model estimation. Table **??** contains results for four specifications. I will use the model results of specification (4) for my counterfactual experiment in Section 6.

One of the main differences between the given specifications is the nesting structure that I used. Nesting of the household consumption alternatives plays an important role in the identification of the parameters of the model. It takes into account that some of the households' options have a higher degree of substitution than other's. This degree of substitution is driven by consumer preferences over format and food categories

and takes into account the similarity of product offerings across retailers.

Table 7 contains model estimation results for four specifications.

Column 1 contains the results of a structural model where nesting of household choices is performed on the format level. Three nests are present in this specification grocery, supercenter, and club. Healthy and unhealthy consumption within the same format belongs to the same nest. The final results suggest that households have heterogeneous preferences over store formats. Parameter that governs format slope difference between healthy and unhealthy food categories is -0.036 this indicates that higher income households consume more healthy foods than low income. Consumers exhibit different inclination to travel to stores of different formats. The highest willingness (disutility of distance equal to -0.062) to travel to supercenters which is a sign of one stop shopping behavior when consumers buy not only groceries at supercenters. Consumers are less willing to travel to stores in more dense areas, but this difference is the practically the same for healthy and unhealthy food categories. The effect of store size is the same for healthy and unhealthy products but this result is driven by the nesting structure of error terms.

Column 2 contains the results of a structural model where nesting of household choices is performed on the format-category level. Five nests are present in this specification grocery-healthy, grocery-unhealthy, supercenter-healthy, supercenter-unhealthy, and club. The reason behind combining healthy and unhealthy consumption in club stores into one nest is that customers of the club stores need to pay membership fee to access club stores which makes shock common across healthy and unhealthy categories within club format. Another reason which suggest to combine healthy and unhealthy consumption in the club format into one nest is that there are small number of observations in the club nest which raises identification issues should we divide club format into two nests. The change of nesting structure largely affected the parameter that governs format slope difference between healthy and unhealthy food categories. This parameter becomes -0.097 compared to -0.037 in the Column 1 specification and indicates that Column 2 model managed to better capture the difference in preferences over food categories across income levels.

Column 3 contains results of a structural model where nesting is the same as in Column 2 and additional variable which is indicator for households with less than 3 members and no kids is introduced. This indicator is intended to capture the fact that different type of household may have different disutility of distance for the healthy food which may be related to amount of available time. Introducing this variable amplified the effect of distance on household preferences over consumed food categories: distance difference for unhealthy food parameter changed from 0.009 to 0.015 from Column 2 to Column 3.

Column 4 contains results of a structural model where in addition to nesting in Columns 2 and 3 grocery format is divided into large and small grocery stores which finally makes 7 nests - large grocery-healthy, large grocery-unhealthy, small grocery-healthy, small grocery-unhealthy, supercenter-healthy, supercenter-

	(1) 3 Nests	(2) 5 Nests	(3) 5 Nests	(4) 7 Nests
Format Effects Healthy Food Intercepts				
Grocery	−1.412	−2.062	−2.062	−2.695
	(0.159)	(0.185)	(0.185)	(0.221)
Supercenter	−2.059	−2.737	−2.739	−3.048
	(0.162)	(0.184)	(0.184)	(0.228)
Club	−3.459	−3.976	−3.978	−4.275
	(0.201)	(0.229)	(0.229)	(0.268)
Intercept Difference	0.302	0.755	0.757	0.740
	(0.056)	(0.122)	(0.122)	(0.126)
Format Effects Healthy Food Slopes				
Grocery	0.103	0.144	0.147	0.127
	(0.040)	(0.057)	(0.057)	(0.056)
Supercenter	−0.065	−0.020	−0.016	−0.019
	(0.045)	(0.061)	(0.061)	(0.061)
Club	0.533	0.569	0.572	0.567
	(0.069)	(0.077)	(0.077)	(0.077)
Slopes Difference	−0.036	−0.097	−0.101	−0.102
	(0.029)	(0.067)	(0.067)	(0.067)
Healthy Food Format Distance				
Grocery	−0.106	−0.113	−0.119	−0.144
	(0.008)	(0.014)	(0.018)	(0.019)
Supercenter	−0.062	−0.066	−0.072	−0.098
	(0.010)	(0.016)	(0.019)	(0.020)
Club	−0.089	−0.096	−0.102	−0.116
	(0.016)	(0.019)	(0.022)	(0.023)
Distance Difference	0.002	0.009	0.015	0.016
Unhealthy Food	(0.009)	(0.019)	(0.022)	(0.024)
Distance Difference			0.010	0.011
Healthy Across Households			(0.016)	(0.017)
Distance*log(Density)	−0.149	−0.147	−0.147	−0.190
Healthy Food	(0.018)	(0.033)	(0.033)	(0.035)
Distance*log(Density)	−0.153	−0.155	−0.155	−0.202
Unhealthy Food	(0.014)	(0.020)	(0.020)	(0.023)
Size Healthy	0.644	0.692	0.693	0.864
	(0.045)	(0.086)	(0.087)	(0.106)
Size Unhealthy	0.646	0.643	0.643	0.768
	(0.035)	(0.051)	(0.051)	(0.060)
Outside Option				
log(Density)	0.839	0.856	0.857	1.220
	(0.214)	(0.214)	(0.214)	(0.214)
$\log(\text{Density})^2$	−0.368	−0.373	−0.373	−0.540
	(0.157)	(0.157)	(0.157)	(0.156)

Standard errors in parentheses.

Table 7: Estimation Results

unhealthy and club. Introducing of additional nests amplified the effect of store size on the choice of healthy vs unhealthy food.

Figure 1 presents households' preference over format-categories across income levels. There a substantial heterogeneity in the consumers' preferences over formats-categories. Households shift their grocery budget to healthy food in grocery and club stores from supercenters as their income increases.

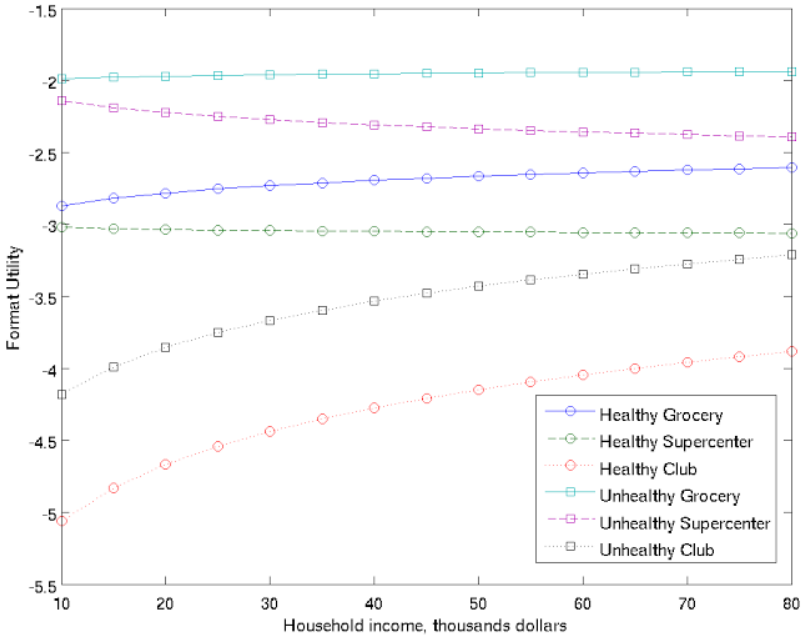

Figure 1: Format Utilities Across Income Levels

6 Counterfactual

In order to address the question of relative importance of household income versus retail environment I will conduct counterfactual experiment using structural model. I will increase household income by 10% and see what would be the effect of this increase on the share of healthy consumption across households. The results will be compared to another experiments when distance to all the stores will be decreased by 10 % and store size of all stores in the area will increase by 10 %. Table 8 and Table 9 contain median increase in the share of healthy consumption for rural and urban households after conducting these counterfactual experiments. I am especially interested in the results of these counterfactual experiments for low income households living far from large stores. The results show that for these households a larger increase in the share of healthy consumption appears after an increase in the store sizes and a decrease in the traveled distance. These result is even more pronounced for rural households. This results suggest that SNAP (Supplementary Nutrition Assistance Program) may not be that efficient in improving the diet of households who are living in food deserts. The problem should be addressed with policies that target product offerings (since store size reflects the healthiness of products available) and cost of travel.

Distance to the store, miles	Percent change	Annual Income, $1,000 Low (<40)	Medium (40≤ and <70)	High (70≤)	Total
Small (< 2)	Distance	0.15	0.18	0.21	0.18
	Income	0.20	0.21	0.22	0.21
	Size	0.36	0.37	0.35	0.36
	Number of HH	2,870	3,180	2,404	8,454
Medium (2 ≤ and <4)	Distance	0.31	0.36	0.37	0.35
	Income	0.19	0.20	0.21	0.20
	Size	0.40	0.40	0.38	0.39
	Number of HH	1,713	2,130	1,795	5,638
Large (4 ≤)	Distance	0.63	0.68	0.69	0.68
	Income	0.16	0.18	0.19	0.18
	Size	0.50	0.50	0.49	0.50
	Number of HH	2,184	2,612	1,873	6,669
Total	Distance	0.32	0.37	0.38	0.35
	Income	0.19	0.20	0.21	0.20
	Size	0.41	0.41	0.39	0.40
	Number of HH	6,767	7,922	13,635	20,761

Table 8: **Median increase of share of household healthy consumption across distances to the closest supermarket and income levels in rural households**

Distance to the store, miles	Percent change	Annual Income, $1,000 Low (<40)	Medium (40≤ and <70)	High (70≤)	Total
Small (< 0.5)	Distance	0.38	0.39	0.41	0.39
	Income	0.21	0.22	0.24	0.21
	Size	0.42	0.43	0.44	0.43
	Number of HH	1,362	1,515	1,343	4,220
Medium (0.5 ≤ and <1)	Distance	0.38	0.41	0.43	0.41
	Income	0.21	0.22	0.24	0.22
	Size	0.41	0.41	0.43	0.42
	Number of HH	2,272	3,477	3,434	9,638
Large (1 ≤)	Distance	0.42	0.45	0.48	0.46
	Income	0.20	0.22	0.24	0.22
	Size	0.44	0.45	0.45	0.44
	Number of HH	1,896	2,555	2,786	7,237
Total	Distance	0.42	0.42	0.45	0.42
	Income	0.21	0.22	0.24	0.22
	Size	0.42	0.43	0.43	0.43
	Number of HH	5,985	7,547	7,563	21,095

Table 9: **Median increase of share of household healthy consumption across distances to the closest supermarket and income levels in urban households**

7 Conclusion

This paper analyzed the relative importance of store' format, household income, store size, and distance to store to the healthiness of a household's consumption. The results suggest that for low income households living far from the large store, distance to the store and store size plays a more important role than household income. These results provide insight for the policymakers about the efficiency of various government policies that try to improve a household's diet.

References

Andreyeva, T., D. M. Blumenthal, M. B. Schwartz, M. W. Long, and K. D. Brownell (2008). Availability and prices of foods across stores and neighborhoods: the case of new haven, connecticut. *Health Affairs 27*(5), 1381–1388.

Bitler, M. and S. J. Haider (2011). An economic view of food deserts in the united states. *Journal of Policy Analysis and Management 30*(1), 153–176.

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

Chen, D., E. C. Jaenicke, and R. J. Volpe (2016). Food environments and obesity: Household diet expenditure versus food deserts. *American journal of public health 106*(5), 881–888.

Chen, S., R. J. Florax, S. Snyder, and C. C. Miller (2010). Obesity and access to chain grocers. *Economic geography 86*(4), 431–452.

Chenarides, L., E. Jaenicke, et al. (2016). Store choice and consumer behavior in food deserts: An empirical application of the distance metric method. In *2016 Annual Meeting, July 31-August 2, 2016, Boston, Massachusetts*, Number 235925. Agricultural and Applied Economics Association.

Ellickson, P. B. (2006). Quality competition in retailing: A structural analysis. *International Journal of Industrial Organization 24*(3), 521–540.

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

Kyureghian, G., R. M. Nayga, and S. Bhattacharya (2013). The effect of food store access and income on household purchases of fruits and vegetables: A mixed effects analysis. *Applied Economic Perspectives and Policy 35*(1), 69–88.

Laraia, B. A., A. M. Siega-Riz, J. S. Kaufman, and S. J. Jones (2004). Proximity of supermarkets is positively associated with diet quality index for pregnancy. *Preventive medicine 39*(5), 869–875.

Laska, M. N., K. E. Borradaile, J. Tester, G. D. Foster, and J. Gittelsohn (2010). Healthy food availability in small urban food stores: a comparison of four us cities. *Public health nutrition 13*(07), 1031–1035.

Liese, A. D., K. E. Weis, D. Pluto, E. Smith, and A. Lawson (2007). Food store types, availability, and cost of foods in a rural environment. *Journal of the American Dietetic Association 107*(11), 1916–1923.

Morland, K., S. Wing, A. D. Roux, and C. Poole (2002). Neighborhood characteristics associated with the location of food stores and food service places. *American journal of preventive medicine 22*(1), 23–29.

Rahkovsky, I., S. Snyder, et al. (2015). Food choices and store proximity. *ERR-195. US Department of Agriculture, Economic Research Service*.

Rose, D. and R. Richards (2004). Food store access and household fruit and vegetable use among participants in the us food stamp program. *Public health nutrition 7*(08), 1081–1088.

Volpe, R. and A. Okrent (2013). *Assessing the healthfulness of consumers' grocery purchases*. BiblioGov.

Volpe, R., A. Okrent, and E. Leibtag (2013). The effect of supercenter-format stores on the healthfulness of consumers grocery purchases. *American Journal of Agricultural Economics*, aas132.

Yoo, S., T. Baranowski, M. Missaghian, J. Baranowski, K. Cullen, J. O. Fisher, K. Watson, I. F. Zakeri, and T. Nicklas (2006). Food-purchasing patterns for home: a grocery store-intercept survey. *Public Health Nutrition 9*(03), 384–393.

A Additional Figures and Tables

State	Frequency	Percent	Cumulative	State	Frequency	Percent	Cumulative
AL	782	1.87	1.87	MT	168	0.40	53.72
AR	518	1.24	3.11	NC	1,490	3.56	57.28
AZ	816	1.95	5.06	ND	123	0.29	57.58
CA	2,604	6.22	11.28	NE	291	0.70	58.27
CO	779	1.86	13.14	NH	331	0.79	59.06
CT	316	0.75	13.89	NJ	740	1.77	60.83
DC	1	0.00	13.90	NM	224	0.54	61.37
DE	206	0.49	14.39	NV	311	0.74	62.11
FL	3,214	7.68	22.07	NY	1,776	4.24	66.35
GA	1,175	2.81	24.87	OH	2,196	5.25	71.60
IA	752	1.80	26.67	OK	406	0.97	72.57
ID	262	0.63	27.30	OR	621	1.48	74.05
IL	1,686	4.03	31.32	PA	1,949	4.66	78.71
IN	1,239	2.96	34.28	RI	83	0.20	78.91
KS	498	1.19	35.47	SC	643	1.54	80.44
KY	813	1.94	37.42	SD	138	0.33	80.77
LA	540	1.29	38.71	TN	1,048	2.50	83.28
MA	584	1.40	40.10	TX	3,227	7.71	90.99
MD	746	1.78	41.88	UT	322	0.77	91.76
ME	329	0.79	42.67	VA	1,031	2.46	94.22
MI	1,940	4.63	47.31	VT	108	0.26	94.48
MN	1,067	2.55	49.85	WI	953	2.28	99.26
MO	1,085	2.59	52.45	WV	227	0.54	99.80
MS	367	0.88	53.32	WY	82	0.20	100.00
				Total	41,856	100,00	100,00

Table 10: **Households by State**