Does a Nutritious Diet Cost More in Food Deserts?

Linlin Fan, University of Illinois at Urbana-Champaign
Kathy Baylis, University of Illinois at Urbana-Champaign
Craig Gundersen, University of Illinois at Urbana-Champaign
Michele Ver Ploeg, USDA Economic Research Service

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

Does a Nutritious Diet Cost More in Food Deserts?

Linlin Fan*
Ph.D. Student
Agricultural and Consumer Economics
University of Illinois at Urbana-Champaign
402 Mumford Hall
1301 West Gregory Drive
Urbana, IL 61801
Phone: (217) 974-5508
E-mail: lfan3@illinois.edu

Kathy Baylis
Associate Professor
Agricultural and Consumer Economics
University of Illinois at Urban-Champaign

Craig Gundersen
Professor
Agricultural and Consumer Economics
University of Illinois at Urbana-Champaign

Michele Ver Ploeg
Economist
Economic Research Services
U.S. Department of Agriculture

* Corresponding author

Disclaimer: We gratefully acknowledge financial support through Cooperative Agreement No. 58-5000-1-0051 between the University of Illinois and the Economic Research Service (ERS) of the U.S. Department of Agriculture (USDA). We would like to thank Jessie Handbury for providing us with her elasticities of substitution between and within brand-products. We also acknowledge Paula Dutko, Joseph Llobrera, Lisa House and Parke Wilde for their assistance with the FoodAPS Geography Component datasets. The views expressed in this paper are of the authors and should not be attributed to those of the Economic Research Services, the USDA or the IRI.

Does a Nutritious Diet Cost More in Food Deserts?

Abstract

Food deserts and their potential effects on diet and nutrition have received much attention from policymakers. While some research has found a correlation between food deserts and consumer outcomes, it is unclear whether food deserts truly affect consumer choices. In this article, we compare food prices in food deserts, defined as low-income, low-access census tracts, and non-food deserts to observe whether and to what extent consumers face higher prices for a complete diet in food deserts. If a nutritionally complete diet costs significantly more in food deserts, resident consumers may be constrained from consuming healthier foods. We use data on store-level sales from a nationally representative sample and calculate a census-tract level Exact Price Index (EPI) based on a food basket defined by the Thrifty Food Plan (TFP). The EPI addresses potential biases from both product heterogeneity and variety availability. We have three central findings. First, prices for common foods are not significantly different between food deserts and non-food deserts. Second, after controlling for differential access to food variety, we find that the EPI in food deserts is 3% to 5% higher than similar census tracts with more store access and 4% higher than low-access census tracts with higher income. Third, the higher EPI in food deserts is largely driven by the lack of supermarkets nearby.

Keywords: food deserts, food price, price indices, product variety, nutritious diet

JEL codes: Q18, D40, L66, R32, I3

1. Introduction

Limited access to healthy food in the U.S. has been associated with poorer diet quality (Morland, Wing and Roux 2002; Bodor et al. 2008; Zenk et al. 2009; Michimi and Wimberly 2010), and a higher probability of obesity and other dietary related health problems (Larson, Story and Nelson 2009; Carroll-Scott, 2013). In response, several federal, state, and local initiatives have emerged to address the challenge of food deserts, including subsidizing large grocery retailers to move into underserved areas, improving food options in corner stores, encouraging mobile grocery vendors and allowing Supplemental Nutrition Assistance Program (SNAP) participants to use benefits at farmers' markets. Multiple states have also enacted legislation aimed at increasing the number of healthy food retailers or have subsidized local stores to provide fresh fruits and vegetables.

Implicit in these interventions is the idea that food deserts, defined as geographic areas with low-income and low food access, are thought to have higher food prices[1] (MacDonald and Nelson 1991) and less access to those foods required for a healthy diet than non-food deserts (e.g. Chung and Meyers 1999; Block and Kouba 2005; Andreyeva et al. 2008). These assumptions are based on case-study comparisons of food prices that focus on a single community (e.g. Chung and Meyers 1999; Block and Kouba 2005; Andreyeva et al. 2008) or use prices from only one or two store chains (e.g. Hatzenbuehler, Gillespie and O'Neil 2012). Other studies compare prices of specific food items, such as fresh fruits and vegetables (e.g. Hayes 2000; USDA 2009). In this article, we compare the price of a nutritious diet in food deserts to non-food deserts using a representative sample and a wide variety of foods. We specifically

[1] We use prices and price indices interchangeably in this article.

compare food prices in food desert census tracts to those in census tracts of similar income but higher food access, and higher income but similar access to differentiate the effect of access from that of income.

We use weekly barcode level store sales data for a nationally representative geographic sample from Information Resources, Inc. (IRI) in 2012 and build a localized price index for each census tract to be able to relate it to the same geographic scale used to designate food deserts. We define an affordable and nutritious diet following the USDA Thrifty Food Plan (TFP), which is a minimum cost diet based on low-income households' purchasing behavior and nutritional guidelines.

We construct a localized TFP Exact Price Index (EPI) following a well-established approach developed by Feenstra (1994), Broda and Weinstein (2010) and applied by Handbury and Weinstein (2014) (see Feenstra 2010 for a review of its use).[2] Our localized TFP EPI is composed of both a Conventional Exact Price Index (CEPI) that accounts for the prices of food available in the census tract and a Variety Adjustment (VA) that addresses the problem that some foods are unavailable in some locations, or variety bias. Assuming nested Constant Elasticity of Substitution (CES) preferences, the price index measures consumers' minimum cost to achieve the same level of utility in a census tract. The VA uses both estimated elasticities of substitution and national expenditure shares of each barcode to capture the impact of variety on prices. We control for product heterogeneity by using barcode-level prices rather than average costs for broad food categories. Handbury and Weinstein (2014) show that after controlling for product

[2] Broda and Weinstein (2010) construct the nation-level EPI for all consumer goods including non-food items such medicine, electronics and appliances in the U.S. in each year. Handbury and Weinstein (2014) devise the city-level EPI for all food. In this article, we focus on a census-tract level EPI for all food.

heterogeneity and variety availability across cities, contrary to previous findings, larger cities have lower food prices than smaller cities.

After constructing the price indices, we regress the CEPI, VA and EPI against an indicator for food deserts along with a low-income census tracts indicator variable and a low-access census tracts indicator variable, neighborhood socio-demographic variables, and county fixed effects. We can then calculate the price differences between food deserts and all three types of non-food deserts using combinations of estimated coefficients for the three indicator variables. Our regression is in the spirit of a hedonic regression where the food prices of a neighborhood are a function of the characteristics of the neighborhood. We restrict our analysis to urban census tracts to avoid comparing food deserts across different definitions in urban and rural areas.

Our article makes several contributions to the literature on food deserts. First, much earlier literature on food deserts does not control for product heterogeneity or variety bias. Concern about product heterogeneity arises when one uses average costs for broad food categories instead of the specific price for a food item to calculate a price index, implying one cannot disentangle whether stores sell identical products at higher prices in one region or whether the higher prices are driven by different quality varieties of these products. Variety bias occurs when not all foods are available in all locations. Insofar as products are not perfect substitutes, consumers, in general, prefer variety, and this loss in utility should be accounted for in the relative price index. We control for product heterogeneity and product variety in the construction of our food price index, and apply that index to food deserts. Second, unlike previous work, we use prices of a complete nutritious diet to capture how much it costs to purchase a full set of groceries, not just one or two food groups. Third, we use a representative geographic sample across urban areas and a comprehensive list of stores within the United States, which allows us to speak broadly to the price effect of living in a food desert.

Our central findings are as follows. First, little difference is found in prices of those foods available both in food deserts and non-food deserts. Second, the variety-adjusted price index EPI in food deserts is 3% to 5% higher than low-income high-access (LIHA) census tracts, 4% higher than high-income low-access (HILA) census tracts and 5% to 8% higher than high-income high-access (HIHA) census tracts. But most TFP food groups are available in both food deserts and non-food deserts. Therefore, the differences in EPI are driven by differences in food availability within a food group between food deserts and non-food deserts, not across food groups. Third, we find that the central reason why the EPI is higher in food deserts is due to the lack of nearby supermarkets, rather than other food outlets.

2. Methods

We begin with the intuition behind the potential gains from variety (from Feenstra 2010) in figure 1. Suppose a consumer gains utility from the consumption of two goods (q_{1c} and q_{2c}). If the consumer has access to both goods, then to achieve the level of utility, *AD*, the consumer would choose to consume at point *C* and only spend the amount of money denoted by *EF*. However, if q_{2c} is not available in the local market, to achieve the same utility level of *AD*, the consumer can only choose point *A* as the consumption bundle and needs to spend more money indicated by the minimum cost line *AB*. How much the cost will increase depends on the per-unit utility of the missing good and the substitutability of the available good compared to the missing one. The increase in the cost needed to achieve the same level of utility when you do not have access to all varieties of goods is a form of gains from variety, which was introduced by Feenstra

(1994). The variety-adjusted price index (EPI) is the relative minimum cost in a census tract of obtaining the TFP for consumers who purchase food within their census tract and neighboring contiguous census tracts.[3] We define the major shopping areas of consumers as their own and contiguous census tracts.

There are two components of the EPI. One is the unadjusted price index, the CEPI. It measures prices of food that are available in a consumer's census tract and contiguous census tracts and is given by the weighted product of the price indexes of each food group. Specifically, the weight W_g is calculated as the recommended pounds of consumption for the food group divided by the total recommended consumption for all TFP food groups for a male aged 19 to 50[4]. The CEPI for food group g in census tract c is given by

$$CEPI_{gc} = \prod_{u \in U_{gc}} \left(\frac{V_{uc} / Q_{uc}}{\sum_c V_{uc} / \sum_c Q_{uc}} \right)^{W_{uc}} \qquad (1)$$

where V_{uc} and Q_{uc} are local expenditure and quantity spent on Universal Product Code (UPC or barcode) u across all stores in census tract c and its contiguous census tracts respectively, U_{gc} is the set of all UPCs of food group g available in the census tract c and its contiguous census tracts. The variable W_{uc} is the log-ideal CES Sato (1976) and Vartia (1976) weights that give

[3] We define neighboring contiguous census tracts as those that share any boundary points with the census tract of interest. Based on calculations from USDA National Food Acquisition and Purchase Survey (FoodAPS), households' average distance to the primary food store is 1.94 miles which is within the radius of the contiguous census tracts (2.24 miles).

[4] The weight for each food group we use here is different from the Sato-Vartia weights used by Handbury and Weinstein (2014). The TFP weights are based on recommended pounds of consumption for each TFP food category assigned by the USDA to ensure that the male aged 19 to 50 have adequate nutrition in a week. The Sato-Vartia weights are ideal log-change index weights that focus on the consistency between price, quantity and expenditure indices but may overlook TFP food categories' different importance in nutrition. Thus we use TFP weights to construct the EPI of a nutritious diet.

more weight to UPCs that have a larger local market share (a detailed definition of W_{uc} is provided in appendix A).

The other component of the EPI, Variety Adjustment (VA) or a measure of variety availability is given by the weighted product of variety index of each food group multiplied by food group availability index $S_c^{\frac{1}{1-\sigma}}$. The weight for each food group (W_g) is the same as in equation (1) and variety index of each food group is given by

$$VA_{gc} = \left(S_{gc}\right)^{\frac{1}{1-\sigma_g^a}} \prod_{b \in B_{gc}} \left(S_{bc}\right)^{\frac{W_{bc}}{1-\sigma_g^w}} \qquad (2)$$

Specifically, W_{bc} is the log-ideal CES Sato (1976) and Vartia (1976) weight for brand-product *b* in census tract *c* defined in appendix A and B_{gc} is the set of all brand-products belong to food group g in census tract c and its contiguous tracts. Food is split into three tiers within the nested framework. All food items (UPCs) are, first, categorized into the 29 food groups in the TFP and second, into different brand-products. Thus the variables σ, σ_g^a, σ_g^w are the elasticity of substitution between food groups, across brand-products within food group *g*, and within a brand-product of food group *g* respectively. The elasticities σ_g^a and σ_g^w are assumed to be constant for each food group *g*.

We use national expenditure shares to capture the importance of the availability of different UPCs, brand-products and food groups in the variety index. The variable S_{bc} is the national expenditure share spent on the UPCs within a brand-product that are available in census tract *c* and its contiguous census tracts.[5] Analogously, S_{gc} is the national expenditure share spent on the brand-products within a food group that are available in census tract *c* and its contiguous census tracts. The variable S_c is the national expenditure share on the food groups that are available in census tract *c* and its contiguous census tracts. The detailed equations used to calculate national expenditure shares S_{bc}, S_{gc} and S_c are provided in appendix A.

The variety-adjusted price index, EPI_c in census tract c is the product of $CEPI_c$ and VA_c:

$$EPI_c = \prod_{g \in G} S_c^{\frac{1}{1-\sigma}} \left[CEPI_{gc} VA_{gc} \right] = \prod_{g \in G} \left[CEPI_{gc} \right]^{W_{gc}} \prod_{g \in G} \left[VA_{gc} \right]^{W_{gc}} S_c^{\frac{1}{1-\sigma}} = CEPI_c VA_c \quad (3)$$

The $CEPI_c$ can be thought of as the correct way to measure the price level of the census tract if all TFP UPCs are available in the census tract or its contiguous census tracts. Since some census tracts together with their contiguous census tracts do not have all UPCs, brands, or TFP categories, we need to adjust the price index by the Variety Adjustment (VA_c). The variety adjustment consists of three availability indices. The UPC availability index of a census tract is given by $\prod_{b \in B_{gc}} (S_{bc})^{\frac{W_{bc}}{1-\sigma_g^w}}$ where variable S_{bc} provides a utility-adjusted count of missing UPCs in census tract *c* and its contiguous census tracts and the exponent weights the counts by how substitutable they are (σ_g^w) and how important they are in consumers' demand locally (W_{bc}).

The UPC availability index implies that if the census tract misses a UPC with a large national expenditure share (S_{bc}), then the missing UPC is important in utility, and the VA and EPI will be higher. If the missing UPC is highly substitutable with other UPCs that exist in the census tract (σ_g^w is large), then missing the UPC will not greatly affect the VA. Similarly, the

[5] As an illustration, suppose that there are 10 UPCs of brand-product *b* available nationally, but only 4 UPCs are available in census tract *c* and its contiguous census tracts. Then the S_{bc} is calculated by dividing national expenditure on the 4 UPCs of brand-product *b* by the national expenditure on all 10 UPCs of that brand-product.

brand-product and food group availability index ($S_{gc}^{\frac{1}{1-\sigma_g^a}}$ and $S_c^{\frac{1}{1-\sigma}}$) depend on the national expenditure shares (S_{gc} and S_c) and whether the brand-product/food group has close substitutes (σ_g^a and σ). In our main analysis we use the sum of the TFP weights for food groups that are available in the census tract or its contiguous census tracts instead of $S_c^{\frac{1}{1-\sigma}}$ because TFP weights have direct implications for nutrition. The lower the VA, the more goods are available in the census tract and its contiguous census tracts, and thus the less adjustment needed for the CEPI, and the closer the EPI is to the CEPI.

After constructing the census tract price indices, we compare the CEPI, VA and EPI based on the following model:

$$y_{ij} = \alpha_0 + \alpha_1 FD_{ij} + \alpha_2 LA_{ij} + \alpha_3 LI_{ij} + x_{ij}\beta + C_j + \varepsilon_{ij} \qquad (4)$$

where y_{ij} is the log of CEPI, VA or EPI for census tract *i* in county *j*; The indicator variables LA_{ij} LI_{ij} and FD_{ij} , take the value of one if the census tract *i* in county *j* is a low-access census tract, a low-income census tract or a food desert (defined as a low-income low-access census tract). The precise definitions of low-income and low-access census tracts are given in the data section. So the omitted group is high-income, high-access (HIHA) census tracts. We also control for other neighborhood characteristics (x_{ij}) i.e. population, education, gender, marital status, age and racial composition to control for the demand factors that may influence local food prices. One might be concerned that the effects of food deserts may reflect different transportation cost to ship food to local stores or that bigger counties may have more local firms that can offer more varieties at cheaper prices. Therefore we include county fixed effects (C_j) to control for county-specific heterogeneity such as county-level transportation costs and county-specific economies of agglomeration. We do not differentiate between county-level demand and supply characteristics.

Based on the regression results, we compare how CEPI, VA and EPI differ between food deserts and all types of non-food deserts. Notice that this specification provides each type of census tracts its own intercept, based on a combination of coefficients $\alpha_0, \alpha_1, \alpha_2, \alpha_3$. In the absence of any other variables x_{ij} and C_j, a combination of $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ will exactly capture the mean price index of the subgroups of census tracts. In a more fully specified model, these coefficients can be combined to construct regression-adjusted mean price index for different types of census tracts.

3. Data

We use the Information Resources, Inc. (IRI) retailer scanner data (IRI InfoScan) of 10,403 census tracts in 2012 in the United States to construct the census tract level CEPI, VA and EPI. These sales datasets are provided as a part of 2012 USDA National Household Food Acquisition and Purchases Survey (FoodAPS). These IRI data only cover stores in the fifty primary sampling units (PSUs) from the FoodAPS, where the PSUs are selected randomly from all counties in the nation.[6] The IRI InfoScan data provide the sales, quantity sold, brand and the product description of each food item at the UPC level at each store or regional market area (RMA) on a weekly basis.[7,8] The datasets cover almost all major national and regional chain stores in the 108 counties

[6] The counties in the FoodAPS are nationally representative in terms of the number of SNAP households and the number of non-SNAP households from three income groups: below 100 percent of the poverty threshold; between 100 and 184 percent of the poverty threshold, and equal to or greater than 185 percent of the poverty threshold.

[7] We include both random-weight food items (usually fresh produce) that have a pseudo UPC and non-random-weight food items (standardized food items) that have a unique UPC.

[8] Some store chains only provide weekly sales datasets at the RMA level. The RMAs of a store chain are aggregate geographical areas defined by the retailer and usually include several stores. Thus the individual prices paid for a UPC cannot be identified at each store within a RMA. Therefore, we use the average price for the whole RMA to

in the PSUs.[9] Importantly, we construct census tract level price indices based on store sales rather than consumer purchases data such as Nielsen Homescan or FoodAPS where the collection point is at households. We do so because our goal is to study whether stores in food deserts charge higher prices rather than whether consumers in food deserts pay higher prices for what they purchase. For the same reason, we do not include prices of food away from home because we want to focus on the cost of a nutritious diet offered by local food stores rather than the cost of what consumers purchase.

To build the EPI for a nutritious diet, we first need to define a nutritious diet. We use the weekly Thrifty Food Plan (TFP) for a male aged 19 to 50 as our nutritious diet (CNPP 2007). The TFP categorizes all foods into 29 categories and assigns weekly recommended consumption quantities of each food category for fifteen age and gender groups based on the Dietary Guidelines for Americans and the My Pyramid Food Guidance System. The TFP is used to estimate the cost of a nutritious but cheap or "thrifty" diet and serves as the basis for the maximum Supplemental Nutrition Assistance Program (SNAP) monthly benefit. A full list of TFP categories and the weekly recommended pounds for a male aged 19 to 50 are provided in appendix table B.

Next we categorize each UPC into different TFP food groups and brand-products within a TFP food group based on the product descriptions of UPCs. Then we use the elasticities of

---

impute for each store and assume that if a UPC is sold in the RMA, then all stores in the RMA also sell that UPC at the same price.

[9] The covered stores include stores of various types. It includes mass merchandises, drug stores, convenience stores, dollar stores, large grocery stores and club stores. One drawback is that local independent stores and farmers' market are not included. The IRI states that around 80% of nationwide food at home expenditure is spent in stores covered by the IRI. We exclude rural census tracts in the analysis where farmers' markets and independent stores may play a bigger role in households' food at home expenditure. Therefore, missing the data on farmers' markets and independent stores may not bias our results a lot.

substitution within and across brand-products (σ_g^w and σ_g^a) for each TFP food group to calculate the implicit increase in price associated with missing brand-products and groups. Because one TFP food group overlaps with several food groups from Handbury and Weinstein (2014), we use the average of σ_g^a and σ_g^w from the overlapping food groups in Handbury and Weinstein (2014).

After constructing the local TFP EPI, we generate our main explanatory variable of interest, the food deserts indicator, defined as a low-income low-access census tract (USDA 2013). A low-income census tract is defined as one that has either a poverty rate of 20 percent or higher, or a median family income at or below 80 percent of the area's median family income.

What constitutes access is debated in the food deserts literature. The first definition of low-access census tracts we use are tracts having at least 500 people and/or 33 percent of the population residing more than one mile from a supermarket[10] in urban areas, where a supermarket is a store that has over 2 million annual sales and has all major food departments including fresh produce, fresh meat and poultry, dairy, dry and packaged foods and frozen foods. This definition of food deserts is commonly used in the literature (e.g. Thomsen et al. 2015) and is used for policy targeting. For example, projects expanding access to nutritious food in a food desert, according to this definition, can qualify for federal Healthy Food Financing Initiative.

For the second food deserts definition, we include the lack of access to a vehicle instead of distance to supermarkets. Because with the same distance to stores, households without vehicles are more restricted to access stores compared to households with vehicles. Following the USDA (2013), census tracts are defined to have low vehicle access "if at least 100 households are more than ½ mile from the nearest supermarket and have no access to a vehicle; or at least 500 people

[10] The distance from a household to the nearest supermarket is measured by the distance from the centroid of the block groups where the household resides to the nearest supermarket and aggregates to the census tract level.

or 33 percent of the population live more than 20 miles from the nearest supermarket, regardless of vehicle access."

To define food deserts, we use data from USDA FoodAPS Geography Component (FoodAPS-GC) that is based on 2012 TDLinx and STARS store lists and 2008-2012 American Community Survey (ACS). We obtain the other explanatory variables (x_{ij} in equation 4), i.e. race, gender, marital status, age, education and population from the 2008-2012 ACS.[11] Marital status and education are measured by the proportions of people who are married and have completed high school in the census tract respectively. We also include the census-tract level median age, share of male population, and an indicator variable for White, Hispanics, Black and Asians tract where the share of population in each race is over 50%.

Tables 1 and 2 present the summary statistics of the sample by income and access to supermarkets, using both food deserts definitions. All the measures of food prices, food availability and the number of supermarkets are calculated using store data both within the census tracts and contiguous census tracts. Because a high-access census tract is defined by whether the nearest supermarket is within one mile of a census tract in table 1 and vehicle access in table 2, it is possible for a high-access census tract to not have a supermarket within its boundaries or contiguous census tracts. For similar reasons, a food desert can have supermarkets. Nevertheless, in both table 1 and 2, we find that food deserts are less likely to have supermarkets than all types of non-food deserts on average. Next, we construct the average and median TFP cost by first calculating the average (the total expenditure divided by total pounds spent on the food group in the contiguous tracts) and median prices for each food group and use the county-

[11] The Hispanic population for each census tract is not available in the 2008-2012 ACS. Thus we use the Hispanic population variable from 2010 Census.

level average and median prices to impute the prices of the missing food groups.[12] Then we multiply the average/median price of each food group with the recommended pounds of consumption per week to get the average/median TFP cost. Notably, these average and median TFP costs do not address product heterogeneity or variety bias. In theory, missing food groups have infinitely positive prices in the census tract and thus it is problematic to use the average or median prices of the county to impute the missing food groups. We construct the average and median TFP cost this way to compare with our theoretically justified price index EPI to discern the degree of potential bias that is inherent in this common way of imputation for missing food groups.

When comparing food availability, we find that food deserts and non-food deserts have similar number of TFP food groups available but food deserts have almost 10,000 fewer UPCs than both HIHA and HILA tracts.[13] However, both measures of variety availability do not account for the substitutability between food items or the differential utility levels of food items. After addressing both issues, we find the variety index (VA) in food deserts is similar to that of LIHA tracts but 17% higher than HILA in table 1, with a higher VA indicating lower access to variety. The prices of common food available in food deserts and non-food deserts (CEPI) such as cheese, sweets and coffee are similar between food deserts, LIHA and HILA tracts. Therefore, the variety-adjusted prices (EPI) are similar between food deserts and LIHA tracts but 15% higher than HILA tracts in table 1. The summary statistics suggest income matters more for the variety-adjusted prices than access before controlling for other demand and cost factors. In

[12] There are 7438, 1810, 551 and 603 tracts missing one, two, three and over three food groups with non-zero recommended consumption.

[13] The maximum number of TFP food groups a census tract can have is 26 in the summary statistics because we exclude three food groups with zero recommended consumption for males aged 19 to 50.

comparison, average and median TFP costs are actually lower in food deserts than all types of non-food deserts tracts.[14] It is possible that food deserts sell more lower-quality varieties of foods at lower prices compared to non-food deserts. Thus, if ignored, product heterogeneity and variety bias may mask a great deal of information in price comparisons.

With respect to socio-demographics, food deserts have more unmarried, younger, less educated people and more African Americans and Hispanics compared to HILA and HIHA tracts. LIHA tracts are more similar to food deserts across socio-demographic characteristics. These differences in demographic composition in census tracts imply that to truly compare prices among regions, we need to control for socio-demographic characteristics that may affect preferences and demand.

4. Results

4.1 Main Results

In table 3 we present the regression results that estimate the differences in CEPI, VA and EPI between food deserts and all types of non-food deserts. The LIHA/HILA/HIHA rows in all tables show the differences in the price indices (CEPI, VA, EPI) between food deserts and LIHA/HILA/HIHA tracts holding demographic characteristics constant.

In all of the following results tables (table 3 to table 6), the first three columns show regression results using the food deserts definition 1 while the next three columns use the second definition (vehicle access). As mentioned in the methods section, in all regressions we include

[14] Notably, the average TFP cost is much lower than median TFP cost across all types of census tracts. It is because the average price for a TFP food group is calculated based on the total expenditure divided by total quantity sold, and is essentially an expenditure-weighted average price. If consumers spend most of their food expenditure on the cheaper items than more expensive items within a TFP group, then the cheaper items will have a larger weight in the average price than in the median price, resulting in a lower average price than the median price.

census tract characteristics (i.e. population, age, education, gender, marital status and race) to control for demand factors that may affect food prices along with county fixed effects.

Table 3 presents our estimates for how the CEPI, VA, and the EPI, vary across food deserts and non-food deserts. By definition 1, we find that the prices of commonly available foods (CEPI) do not differ significantly between food deserts and LIHA tracts. However, there are 4% fewer varieties available in food deserts, leading to a 5% higher EPI in food deserts than LIHA tracts. Food deserts also have 3% higher EPI than HILA tracts and 4% higher EPI than HIHA tracts. This result is again driven by the fact that food deserts have fewer varieties of foods available. The fact that limited access to variety is the primary reason for a higher EPI highlights the value of incorporating variety into a price index. Although not accounting for product heterogeneity and variety bias, recent studies (Handbury, Rahkovsky and Schnell 2015; Allcott, Diamond and Dube 2015) have also found similar prices of available foods between high and low-income neighborhoods and the lower availability of produce items in low-income areas, which is consistent with our results. In contrast to the summary statistics that show similar average EPI between food deserts and LIHA tracts (table 1 and 2), the regression-adjusted means demonstrate the importance of controlling for various demand factors and cost variables when comparing prices.

One may be concerned that because LA tracts (including food deserts) are places that are less likely to have supermarkets, LA tracts by definition will have a higher VA i.e. less access to variety, and thus a higher EPI. Indeed, we find LA tracts (91.4%) are less likely to have supermarkets than HA tracts (94.2%) including stores in the contiguous census tracts. However, having a supermarket does not automatically mean a lower VA because some food items/brands-products/food groups may have close substitutes available implying the VA will not be greatly affected by greater choice. Furthermore, even if more food options in supermarkets translate into higher VA in LA tracts, lack of variety may not affect the price indices in an economically meaningful way. Our article quantifies how much variety affects the cost of a nutritious diet and finds that lower access to variety translates into a 3%-5% higher EPI in food deserts than LIHA tracts. Lastly, our results are robust to vehicle access definitions of food deserts that are as likely to have supermarkets nearby as HA tracts.

Next, we compare the EPI with the average and median TFP cost in table 4. We find that food deserts do not have significantly different average TFP costs from LIHA tracts. By definition 1, food deserts even have slightly lower TFP median cost than LIHA tracts. Again the lower median TFP cost may be caused by lower-quality varieties of foods sold in food deserts compared with LIHA tracts. The average and median cost for the TFP in food deserts is either only slightly higher or insignificantly different from HILA tracts. Compared to HIHA tracts, the average TFP cost is 3% to 4% higher in food deserts, which is only half of difference found in EPI in table 3. These results suggest that simply using average or median prices of the TFP, heterogeneity bias and variety bias may mask the price effect of living in a food desert.

To check whether the difference in EPI, or essentially VA is driven by the existence of supermarkets, we compare the prices only for the census tracts with supermarkets nearby and those without supermarkets nearby. We find that CEPI, VA, EPI are similar between food deserts and non-food deserts when there are only small stores around (table 5) and are similar between those food deserts and non-food deserts with only supermarkets nearby (table 6). Thus, the higher EPI in food deserts largely comes from those food deserts without a supermarket nearby vs non-food deserts with a supermarket nearby. These results suggest that neither supermarkets nor non-supermarkets charge higher EPI in food deserts compared to non-food deserts. But food deserts are less likely to have supermarkets which, in turn, have higher VA than non-supermarkets. This finding complements a number of studies that find cheaper prices charged at supermarkets such as Wal-Mart (e.g. Courtemanche and Carden 2011). In sum, lack of supermarkets and the lower access to variety leads to higher EPI in food deserts than non-food deserts.

4.2 Robustness Tests

We conduct four robustness tests of our results in appendix C. First, we explore the food prices faced by the consumers who are constrained to shop only within their home tracts as opposed to own and contiguous census tracts as analyzed above. We find that, consumers moving from a LIHA tract to a food desert, would experience a 2% increase in food price for the identical food items, but because VA is 20% higher, the EPI would actually rise by 22% (appendix Table C1). The larger access effects illustrate that for those consumers constrained to shopping within their census tract, access has a much greater effect, and the food desert definition much greater meaning in comparison to above when we allowed consumers to access stores in neighboring census tracts. Second, we test to see if higher prices are driven by higher prices of goods in convenience stores in appendix table C2. We find our results do not merely reflect higher prices for processed foods in convenience stores in food deserts; instead they reflect higher prices in items that are not generally found in convenience stores such as fruits and vegetables. Third, we find our results are robust against different estimates of elasticities of substitution in appendix table C3 and C4. Lastly, we study the store coverage of IRI data compared with TDLinx, the most complete list of food stores at the census-tract level in the U.S. and is widely used by the industry to analyze the regional retail market in appendix table C5. We find that at the census tract level, on average IRI covers over 90% of club stores, mass merchandisers, dollar stores and drug stores. But the coverage of grocery stores (74% and 75% in store counts and sales) and convenience stores (53% and 57%) is lower. The implications of the IRI store coverage on our results are discussed in appendix C.

5. Conclusion

In this article we construct a price index that adjusts for both product heterogeneity and variety bias to compare the local cost of a nutritious diet in food deserts versus non-food deserts. We find that when consumers are assumed to shop in both their home and contiguous census tracts, prices for common goods are not significantly different between food deserts and all types of non-food deserts. But depending on the assumptions around the elasticities of substitution and the definition of food deserts, there are 2% to 4% fewer varieties available in food deserts compared to LIHA tracts. As a result, the variety-adjusted prices (the EPI) are 3% to 5% higher in food deserts than LIHA tracts, capturing the effect of access. Food deserts have 3% fewer varieties than HILA tracts and thus the EPI is 4% higher in food deserts than HILA tracts, capturing the effect of living in a low-income neighborhood. After comparing the price premium paid in food deserts with supermarkets nearby and those without, we find that food deserts prices are significantly higher because there are no supermarkets close by to provide a wide variety.

Our results are strongly affected by how we define the 'local' grocery market. For households who buy food only within their resident tracts, the EPI is 22% higher in food deserts than LIHA tracts. In contrast, when consumers can shop within the contiguous census tracts, the

EPI is only 3% to 5% higher in food deserts than LIHA tracts. This result suggests that those households who are truly geographically constrained in their shopping are much more affected by living in a food desert.

Our findings suggest that living in a food desert affects the overall food prices faced by households. The degree to which prices differ is largely driven by differences in available variety. As such, while higher food prices are associated with higher rates of food insecurity (Gregory and Coleman-Jensen 2013), the results of this article suggest that living in a food desert is unlikely to influence food insecurity to a great extent, at least in as much as substitute foods are available (For more on food insecurity in the United States see Gundersen and Ziliak 2014 and Gundersen Kreider and Pepper 2011). However, efforts to encourage the location of food stores into food deserts and/or contiguous census tracts may have an impact on variety of foods faced by the local consumers. As long as variety in diets improves health, such policies can perhaps be justified. In particular, it might be most effective to encourage supermarkets (rather than other food outlets) to locate in food deserts. This can be achieved through, among other things, removing regulations that may discourage supermarkets from locating in these areas.

We conclude with two major areas for future research. First, while we have established that households in food deserts do not face a higher price index of commonly available goods but they do when variety is incorporated in the index, we are unable to explore the reasons behind this lack of variety. For example, is how much of this difference in variety is due to lower demand for variety? If it is primarily due to lower demand, encouraging stores to carry a wider array of products may not have any effect on variety of foods purchased. Second, our work is based on the price and quantity information available in IRI but, as discussed above, not all stores are included in this data set. While we do not anticipate that including more stores would have an impact on our substantive conclusions, an expansion of the IRI data set to include more stores would help address this issue.

Tables and Figures

Table 1. Summary Statistics (Food Deserts Definition 1)

	Food Deserts	Low Income High Access	High Income Low Access	High Income High Access
Average TFP Cost	58.37	62.22	59.95	61.41
	(16.03)	(18.44)	(13.85)	(13.57)
Median TFP Cost	97.32	103.73	105.14	108.70
	(17.61)	(18.26)	(14.06)	(13.41)
Number of TFP Groups	25.33	25.12	25.59	25.65
	(1.39)	(2.01)	(1.28)	(1.00)
Number of UPCs	29677.73	27498.86	38856.69	38346.90
	(20078.12)	(18653.43)	(19044.73)	(17784.91)
VA	1.46	1.49	1.25	1.23
	(0.58)	(1.37)	(0.41)	(0.34)
CEPI	1.02	1.04	1.03	1.05
	(0.06)	(0.07)	(0.06)	(0.06)
EPI	1.50	1.56	1.30	1.30
	(0.64)	(1.42)	(0.47)	(0.41)
Having supermarkets	0.88	0.92	0.93	0.96
	(0.32)	(0.27)	(0.26)	(0.19)
Population	4358.9	4083.64	4952.96	4184.67
	(1923.88)	(1727.47)	(2198.02)	(1708.02)
Married population share	0.37	0.37	0.56	0.49
	(0.12)	(0.11)	(0.11)	(0.11)
Median age	33.59	32.74	40.99	39.22
	(8.66)	(6.28)	(7.11)	(6.37)
Proportion of population	0.75	0.70	0.93	0.90
who complete high school	(0.14)	(0.15)	(0.06)	(0.08)
Proportion of population	0.49	0.49	0.49	0.49
who are male	(0.06)	(0.05)	(0.04)	(0.04)
Black tract	0.21	0.17	0.03	0.04
	(0.41)	(0.37)	(0.16)	(0.20)
White tract	0.64	0.52	0.91	0.81
	(0.48)	(0.50)	(0.29)	(0.40)
Hispanic tract	0.30	0.44	0.04	0.08
	(0.46)	(0.50)	(0.20)	(0.27)
Asian tract	0.00	0.03	0.02	0.04
	(0.05)	(0.16)	(0.13)	(0.19)
Observations	935	3483	2119	3866

Note: Standard deviations are in the parenthesis. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live at least 1 mile away from supermarkets.

Table 2. Summary Statistics (Food Deserts Definition 2)

	Food Deserts	Low Income High Access	High Income Low Access	High Income High Access
Average TFP Cost	59.20	60.56	59.61	60.49
	(15.07)	(16.62)	(12.25)	(12.78)
Median TFP Cost	98.21	101.59	104.44	107.04
	(16.72)	(18.02)	(13.09)	(13.79)
Number of TFP Groups	25.25	25.12	25.61	25.63
	(1.49)	(2.06)	(1.05)	(1.11)
Number of UPCs	27965.50	27957.35	37713.41	39099.32
	(19892.59)	(18536.31)	(18760.17)	(18172.71)
VA	1.50	1.48	1.26	1.24
	(0.59)	(1.46)	(0.38)	(0.37)
CEPI	1.03	1.03	1.05	1.04
	(0.06)	(0.07)	(0.06)	(0.06)
EPI	1.55	1.55	1.32	1.29
	(0.66)	(1.51)	(0.43)	(0.43)
Having supermarkets	0.89	0.92	0.94	0.95
	(0.32)	(0.27)	(0.23)	(0.22)
Population	4272.74	4079.53	4978.53	4377.56
	(1830.71)	(1743.47)	(2013.78)	(1906.07)
Married population share	0.33	0.38	0.47	0.52
	(0.11)	(0.11)	(0.10)	(0.11)
Median age	33.51	32.64	40.81	39.70
	(7.62)	(6.45)	(7.37)	(6.58)
Proportion of population	0.74	0.70	0.90	0.91
who complete high school	(0.14)	(0.16)	(0.06)	(0.08)
Proportion of population	0.48	0.50	0.48	0.49
who are male	(0.05)	(0.06)	(0.04)	(0.04)
Black tract	0.30	0.12	0.07	0.03
	(0.46)	(0.32)	(0.25)	(0.18)
White tract	0.33	0.28	0.80	0.80
	(0.47)	(0.45)	(0.40)	(0.40)
Hispanic tract	0.28	0.48	0.04	0.07
	(0.45)	(0.50)	(0.19)	(0.25)
Asian tract	0.01	0.03	0.01	0.03
	(0.09)	(0.17)	(0.11)	(0.18)
Observations	1426	2992	788	5197

Note: Standard deviations are in the parenthesis. Food deserts definition 2 is low-income urban census tracts where a significant proportion of households are far away from supermarkets and do not have access to vehicle. Detailed food deserts definitions are provided in the data section.

Table 3. Regressions on EPI–With Contiguous Census Tracts Sales

	Definition 1			Definition 2		
	CEPI	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income	0.002	0.04***	0.05***	0.01***	0.02***	0.03***
High-access	(0.002)	(0.010)	(0.011)	(0.002)	(0.009)	(0.010)
High-income	0.001	0.002	0.003	0.01***	0.03**	0.04***
Low-access	(0.002)	(0.012)	(0.013)	(0.003)	(0.013)	(0.014)
High-access	0.01***	0.07***	0.08***	0.01***	0.04***	0.05***
High-income	(0.002)	(0.011)	(0.012)	(0.002)	(0.010)	(0.011)
Observations	10403	10403	10403	10403	10403	10403

Note: *, **, *** denotes significance levels at 0.1, 0.05 and 0.01 respectively. Standard errors are in parenthesis. All dependent variables are in logs. The values in LIHA/HILA/HIHA rows denote the price differences between food deserts and LIHA/HILA/HIHA census tracts. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live at least 1 mile away from supermarkets. Vehicle access is used for food deserts definition 2. Race, gender, marriage, age, education, population in the census tracts and county fixed effects are also included.

Table 4. Regressions on Log Average and Log Median TFP Cost

	Log Average TFP Cost		Log Median TFP Cost	
	Definition 1	**Definition 2**	**Definition 1**	**Definition 2**
Low-income High-access	0.01	0.01	-0.02***	-0.002
	(0.008)	(0.007)	(0.005)	(0.004)
High-income Low-access	-0.01	0.04***	-0.01	0.01**
	(0.01)	(0.01)	(0.006)	(0.006)
High-access High-income	0.04***	0.03***	-0.01	0.01
	(0.009)	(0.008)	(0.006)	(0.005)
Observations	10403	10403	10403	10403

Note: *, **, *** denotes significance levels at 0.1, 0.05 and 0.01 respectively. Standard errors are in parenthesis. All dependent variables are in logs. The values in LIHA/HILA/HIHA rows denote the price differences between food deserts and LIHA/HILA/HIHA census tracts. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live at least 1 mile away from supermarkets. Vehicle access is used for food deserts definition 2. Race, gender, marriage, age, education, population in the census tracts and county fixed effects are also included.

Table 5. Regressions on EPI–On Census Tracts without Supermarkets Nearby

	Definition 1			Definition 2		
	CEPI	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income High-access	-0.009** (0.004)	0.000004 (0.017)	-0.009 (0.016)	-0.002 (0.003)	-0.018 (0.013)	-0.020 (0.013)
High-income Low-access	-0.006 (0.005)	-0.038* (0.023)	-0.045** (0.022)	-0.004 (0.005)	0.004 (0.025)	0.0002 (0.024)
High-access High-income	-0.013*** (0.004)	0.024 (0.021)	0.011 (0.020)	-0.005 (0.004)	-0.010 (0.018)	-0.015 (0.018)
Observations	2163	2163	2163	2163	2163	2163

Note: *, **, *** denotes significance levels at 0.1, 0.05 and 0.01 respectively. Standard errors are in parenthesis. All dependent variables are in logs. The values in LIHA/HILA/HIHA rows denote the price differences between food deserts and LIHA/HILA/HIHA census tracts. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live at least 1 mile away from supermarkets. Vehicle access is used for food deserts definition 2. Race, gender, marriage, age, education, population in the census tracts and county fixed effects are also included.

Table 6. Regressions on EPI–On Census Tracts with Supermarkets Nearby

	Definition 1			Definition 2		
	CEPI	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income	0.0001	0.008**	0.008*	0.006***	-0.001	0.005
High-access	(0.002)	(0.003)	(0.004)	(0.002)	(0.003)	(0.004)
High-income	0.005**	0.004	0.009*	0.010***	0.001	0.011**
Low-access	(0.002)	(0.004)	(0.004)	(0.002)	(0.004)	(0.005)
High-access	0.004**	0.009**	0.013***	0.008***	0.002	0.009**
High-income	(0.002)	(0.004)	(0.005)	(0.002)	(0.003)	(0.004)
Observations	8240	8240	8240	8240	8240	8240

Note: *, **, *** denotes significance levels at 0.1, 0.05 and 0.01 respectively. Standard errors are in parenthesis. All dependent variables are in logs. The values in LIHA/HILA/HIHA rows denote the price differences between food deserts and LIHA/HILA/HIHA census tracts. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live at least 1 mile away from supermarkets. Vehicle access is used for food deserts definition 2. Race, gender, marriage, age, education, population in the census tracts and county fixed effects are also included.

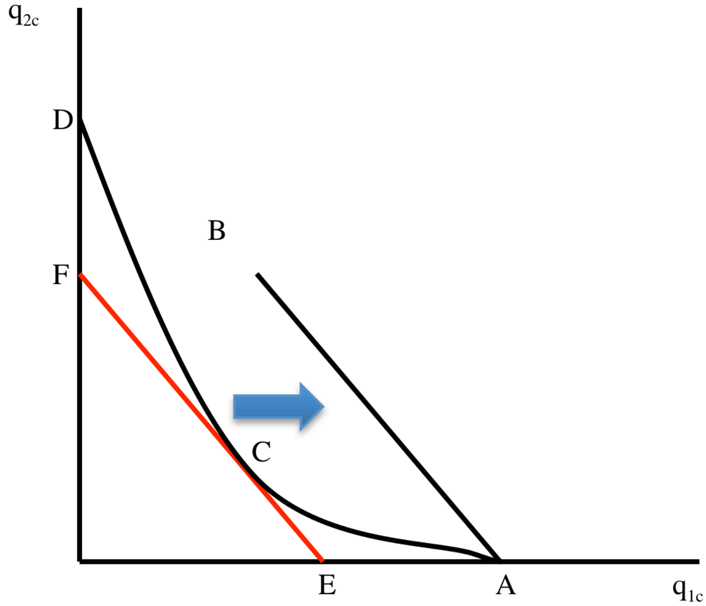

Figure 1. Grains from Variety

References

Allcott, H., Diamond, R., Dube, J., 2015. The Geography of poverty and nutrition: food deserts and food choices across the United States. NBER Working Paper.

Andreyva, T., Blumenthal, D.M., Schwartz, M.B., Long, M.W., Brownell, K.D.. 2008. Availability and prices of food across stores and neighborhoods: the case of New Haven, Connecticut. Health Affairs 27(5), 1381-1388.

Block, D., Kouba, J., 2006. A comparison of the availability and affordability of a market basket in two communities in the Chicago area. Public Health Nutrition 9(7), 837-845.

Bodor. N., Rose, D., Farley, T.A., Swalm, C. Scott, S.K.. 2008. Neighbourhood fruit and vegetable availability and consumption: the role of small food stores in an urban environment. Public Health Nutrition 11(4), 413-420.

Broda, C., Weinstein, D.E., 2010. Product creation and destruction: evidence and price implications. American Economic Review 100, 697-723.

Carroll-Scott, A., Gilstad-Hayden, K., Rosenthal, L., Peters, S.M., McCaslin, C., Joyce, R., Ickovics, J.R., 2013. Disentangling neighborhood contextual associations with child body mass index, diet, and physical activity: the role of built, socioeconomic, and social environments. Social Science and Medicine 95, 106-114.

Center for Nutrition Policy and Promotion, U.S. Department of Agriculture, 2007. Thrifty Food Plan, 2006. Washington DC.

Chung, C., Myers, S.L., 1999. Do the poor pay more for food? an analysis of grocery store availability and food price disparities. Journal of Consumer Affairs 33(2), 276-296.

Courtemanche, C., Carden, A., 2011. Supersizing supercenters? The impact of Walmart supercenters on body mass index and obesity. Journal of Urban Economics 69, 165-181.

Feenstra, R. C. 1994. New product varieties and the measurement of international prices. American Economic Review 84, 157-157.

Feenstra R.C. 2010. Product Variety and the Gains from International Trade. Cambridge: MIT Press.

Gregory, C., Coleman-Jensen, A., 2013. Do high food prices increase food insecurity in the United States? Applied Economic Perspectives and Policy 35, 679–707.

Gundersen, C., Kreider, B., Pepper, J., 2011. The economics of food insecurity in the United States. Applied Economic Perspectives and Policy, 33(3), 281-303.

Gundersen, C., Ziliak, J., 2014. Childhood food insecurity in the U.S.: trends, causes, and policy options. The Future of Children.

Handbury, J., Rahkovsky, I., Schnell, M., 2015. Is the focus on food deserts fruitless? Retail access and food purchases across the socioeconomic spectrum." NBER Working Paper 21126.

Handbury, J., Weinstein, D. E., 2014. Goods prices and availability in cities. The Review of Economic Studies 81(4), 1-39.

Hatzenbuehler, P. L., Gillespie, J. M., O'Neil, C. E., 2012. Does healthy food cost more in poor neighborhoods? An analysis of retail food cost and spatial competition. Agricultural and Resource Economics Review 41(1), 43.

Hayes, L. R, 2000. Do the poor pay more? An empirical investigation of price dispersion in food retailing. Industrial Relations Section Working Paper 446, Dept. of Econ., Princeton University.

Larson, N. I., Story, M.T., Nelson, M.C., 2009. Neighborhood environments: disparities in access to healthy foods in the US. American Journal of Preventive Medicine 36(1), 74-81.

MacDonald, J. M., Nelson, P. E. Jr., 1991. Do the poor still pay more? Food price variations in large metropolitan areas. Journal of Urban Economics 30, 344-359.

Michimi, A., Wimberly, M.C., 2010. Associations of supermarket accessibility with obesity and fruit and vegetable consumption in the conterminous United States. International Journal of Health Geographics 9(1), 49.

Morland, K., Wing S., Roux, A. D., 2002. The contextual effect of the local food environment on residents' diets: the atherosclerosis risk in communities study. American Journal of Public Health 92, 1761–1767.

Sato, K. 1976. The ideal log-change index number. The Review of Economics and Statistics, 223-228.

Thomsen M. R., Nayga, R.M., Alviola, P. A., Rouse, H. L., 2015. The effect of food deserts on the body mass index of elementary school children. American Journal of Agricultural Economics 98(1), 1-18.

U.S. Department of Agriculture, 2009. Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences. Report to Congress. Washington DC.

U.S. Department of Agriculture, 2013. Economic Research Service. Food Access Research Atlas, http://www.ers.usda.gov/data-products/food-access-research-atlas.aspx. (Accessed on October 8, 2015).

Vartia, Y. O., 1976. Ideal log-change index numbers. Scandinavian Journal of Statistics, 121-126.

Zenk, S. N., Lachance, L. L., Schulz, A.J., Mentz, G., Kannan, S., Ridella, W., 2009. Neighborhood retail food environment and fruit and vegetable intake in a multiethnic urban population. American Journal of Health Promotion 23(4), 255-264.

APPENDIX A. SATO AND VARTIA WEIGTHS AND NATIONAL EXPENDITURE SHARES

The log ideal CES Sato and Vartia (1976) weights are

$$W_{uc} = \frac{\dfrac{M_{uc} - M_u}{\ln M_{uc} - \ln M_u}}{\sum_{u \in U_b} \left(\dfrac{M_{uc} - M_u}{\ln M_{uc} - \ln M_u} \right)} \qquad (A.1)$$

$$W_{bc} = \frac{\dfrac{M_{bc} - M_b}{\ln M_{bc} - \ln M_b}}{\sum_{b \in B_g} \left(\dfrac{M_{bc} - M_b}{\ln M_{bc} - \ln M_b} \right)} \qquad (A.2)$$

where M_{uc} and M_{bc} are local market shares of UPC u and brand-product b. The set U_b is the set of all UPCs that belong to the brand-product b while B_g is the set of all brand-products that belong to the food group g. We define M_{uc} and M_{bc} as

$$M_{uc} = \frac{V_{uc}}{\sum_{u \in U_b} V_{uc}} \ , M_{bc} = \frac{\sum_{u \in U_b} V_{uc}}{\sum_{b \in B_g} \sum_{u \in U_b} V_{uc}}$$

where V_{uc} is the sales on UPC u in census tract c and its contagious census tracts. Similarly, the national market shares of UPC u and brand-product b are

$$M_u = \frac{V_u}{\sum_{u \in U_b} V_u} \ , M_b = \frac{\sum_{u \in U_b} V_u}{\sum_{b \in B_g} \sum_{u \in U_b} V_u}$$

The national expenditure shares on UPCs of brand-product b that are available in census tract c and its contagious census tracts are

$$S_{bc} = \frac{\sum_{u \in U_{bc}} \sum_{c} V_{uc}}{\sum_{u \in U_{b}} \sum_{c} V_{uc}} \qquad \text{(A.3)}$$

where the variable U_{bc} denotes the set of all UPCs that belong to brand-product b and exist in census tract c and its contagious census tracts and U_b is the set of all UPCs in brand-product b nationally. The variable V_{uc} is the sales of UPC u in census tract c and its contagious census tracts.

Similarly, the national expenditure shares on brand-products that belong to food group g and are available in census tract c or its contagious census tracts is

$$S_{gc} = \frac{\sum_{b \in B_{gc}} \sum_{c} V_{bc}}{\sum_{b \in B_{g}} \sum_{c} V_{bc}} \qquad \text{(A.4)}$$

where the variable B_{gc} is the set of all brand-products in food group g in census tract c or its contagious census tracts and B_g is the set of all brand-products in food group g nationally. The sales on brand-product b in census tract c and its contagious census tracts is V_{bc}.

To simplify the calculation of variety adjustment (VA), Handbury and Weinstein (2014) aggregate the expenditure across UPCs within a food group and estimate a common S_{bc} within each food group. In other words, $\overline{S_{gc}}$ is a measure of the average availability of UPCs within a brand-product in census tract c and its contagious census tracts. Therefore the group-specific variety adjustment is given by

$$VA_{gc} = (S_{gc})^{\frac{1}{1-\delta_g^a}} (\overline{S_{gc}})^{\frac{1}{1-\delta_g^w}} \qquad \text{(A.5)}$$

where we do not need to use W_{bc} to weight the national expenditure shares on brand-products available in census tract c and its contagious census tracts, S_{bc}.

The national expenditure shares on food groups available in census tract c and its contagious census tracts is

$$S_c = \frac{\sum_{g \in G_c} \sum_c V_{gc}}{\sum_{g \in G} \sum_c V_{gc}} \qquad \text{(A.6)}$$

where the variable G_c is the set of all food groups in census tract c and its contagious census tracts and G is the set of all 26 non-zero weight TFP food groups. The sales on food group g in census tract c and its contagious census tracts is V_{gc}.

References

Sato, K. 1976. The ideal log-change index number. The Review of Economics and Statistics, 223-228.

Vartia, Y. O. 1976. Ideal log-change index numbers. Scandinavian Journal of Statistics, 121-126.

APPENDIX B. TFP FOOD GROUPS AND WEIGHTS

Food Type	Food Category	Pounds Per Week for Males Age 19-50
Grains	Whole grain bread, rice, pasta, pastries (incl whole grain flours)	2.82
Grains	Whole grain cereals incl hot cereal mixes	0.08
Grains	Popcorn and other whole grain snacks	0
Grains	Non-whole grain breads, cereal, rice, pasta, pies, pastries, snacks, and flours	1.66
Vegetables	All potato products	2.48
Vegetables	Dark green vegetables	1.24
Vegetables	Orange vegetables	0.98
Vegetables	Canned and dry beans, lentils, and peas or legumes	1.87
Vegetables	Other vegetables	2.7
Fruit	Whole fruit	6.65
Fruit	Fruit juices	1.76
Milk products	Whole milk, yogurt, and cream	0.55
Milk products	Low-fat and skim milk and low-fat yogurt	10.75
Milk products	All cheese, incl cheese soups and sauces	0.07
Milk products	Milk drinks and milk desserts	0
Meat and beans	Beef, pork, veal, lamb, and game	0.63
Meat and beans	Chicken, turkey, and game birds	2.55
Meat and beans	Fish and fish products	0.17
Meat and beans	Bacon, sausage, and lunch meats including spreads	0.02
Meat and beans	Nuts, nut butters, and seeds	0.26
Meat and beans	Egg and egg mixtures	0.36
Other foods	Table fats, oils, and salad dressings	0.99
Other foods	Gravies, sauces, condiments, and spices	0.99
Other foods	Coffee and tea	0.01
Other foods	Soft drinks, sodas, fruit drinks, and ades incl rice beverages	0
Other foods	Sugars, sweets, and candies	0.08
Other foods	Soups (ready-to-serve and condensed)	0.16
Other foods	Soups (dry)	0.02
Other foods	Frozen/refrigerated entrees incl pizza, fish sticks, and frozen meals	0.01

APPENDIX C. ROBUSTNESS TESTS

In the first robustness test in appendix table C1, we explore the food prices faced by the consumers who are constrained to shop only within their home tracts. As a result, we exclude the census tracts without any IRI stores and 4,835 census tracts remain in this analysis. For the tracts remain in the sample, we find IRI covers most of stores including both supermarkets (73%in store counts, 77% in store sales) and non-supermarkets (60% in store counts, 75% in store sales) that exist in TDLinx data.

In all following robustness tests, we use store data in both home and contiguous census tracts. In our second robustness test, we test to see if higher prices are driven by higher prices of goods in convenience stores. We calculate and compare the EPI for the six most commonly available TFP food groups.[15] The six food groups are mostly processed foods and drinks, that are likely available in both traditional grocery stores and convenience stores. Results show that even after accounting for different access to variety, the EPI of commonly available food groups in food deserts is almost the same as all types of non-food deserts (table C2). Thus, higher prices are driven by higher EPI of goods not available in convenience stores such as fruits and vegetables.

Third, in table C3, we choose 2 as the elasticity of substitution between food groups, which essentially uses the inverse of national expenditure shares on available food groups in the contiguous census tracts (S_c^{-1}) to measure the importance of available food groups. In table C4,

[15] The six TFP food groups are "non-whole grain breads, cereal, rice, pasta, pies, pastries, snacks, and flours", "fruit juice", "all cheese, including cheese soups and sauces", "nuts, nut butters, and seeds", "coffee and tea" and "sugars, sweets, and candies".

we use 4 and 7 as the across and within brand elasticities of substitution that are commonly used in marketing literature (Dube and Manchanda 2005). Our results remain effectively unchanged.

In our fourth robustness test, we study the representativeness of IRI data and explore the effect of IRI store coverage on the results. In table C5, we compare the numbers and sales in IRI and TDLinx stores by census tract types in 2012. The TDLinx contains the names, characteristics, annual sales and geo-coded locations of 269,674 food stores across the U.S. in 2012. We find that the IRI has a good representation of store counts of club stores (99%), mass merchandisers (98%), dollar stores (93%) and drug stores (86%). But the coverage of grocery stores (74%) and convenience stores (53%) is lower.

Specifically, convenience stores are more underrepresented in food deserts than all types of non-food deserts (41% vs 49%, 60% and 56%). It is less of a concern in our price comparisons because all of the census tracts in our contiguous sample have at least one IRI store nearby. Suppose the existing IRI store is a convenience store, then due to the effects of competition and law of one price, both the CEPI and VA of the food desert based on one IRI convenience store will not greatly differ from that based on two convenience stores if we had data on the non-IRI convenience store. If the existing IRI store is a supermarket that offers much wider variety and lower prices, then missing the non-IRI convenience store will not greatly affect the local food price.

However, under-coverage of grocery stores will have a big impact on our price indices since grocery stores have a wide selection of foods at lower prices. We find that in table C5 the grocery stores coverage (74%) in food deserts is higher than LIHA tracts (64%). Because we find that supermarkets charge lower VA and EPI from table 5 and 6, the better coverage of grocery stores in food deserts suggest that we may underestimate the variety and price difference between food deserts and LIHA tracts in all of our analysis. In contrast, the variety and price difference between food deserts and HILA tracts is likely to be overestimated because IRI tends to cover fewer grocery stores that exist in food deserts than HILA tracts (74% vs 85%).

The use of IRI data does affect the size of our sample census tracts but because of the small portion of census tracts we drop, it does not change our results. We drop 556 census tracts (5.1%) of the 10,959 urban census tracts with non-zero population in our sample because there are no IRI stores within the census tracts or their contiguous census tracts. Among the 556 dropped census tracts, 267 do not have any TDLinx or IRI stores while 289 (2.6% of the total) have TDLinx stores but no IRI stores. Because the methods we use indicate the areas without any foods have positively infinite prices, these census tracts are, therefore, not included in our analyses. The share of food deserts, LIHA, HILA, HIHA tracts remain the same in the sample after wen drop 556 tracts, suggesting we do not oversample or undersample any type of tracts, or affect our results because of dropping the 556 tracts.

References

Dube, J., Manchanda, P., 2005. Differences in dynamic brand competition across markets: an empirical analysis. Marketing Science 24(1), 81–95.

Table C1. Regressions on EPI–Without Contiguous Census Tracts Sales

	Definition 1			Definition 2		
	CEPI	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income	0.02***	0.20***	0.22***	0.01***	0.12***	0.13***
High-access	(0.004)	(0.024)	(0.025)	(0.003)	(0.022)	(0.023)
High-income	0.001	0.09***	0.09***	-0.004	0.10***	0.09***
Low-access	(0.004)	(0.029)	(0.031)	(0.005)	(0.030)	(0.032)
High-access	0.01***	0.23***	0.25***	0.01	0.15***	0.15***
High-income	(0.013)	(0.027)	(0.028)	(0.004)	(0.025)	(0.026)
Observations	4835	4835	4835	4835	4835	4835

Note: *, **, *** denotes significance levels at 0.1, 0.05 and 0.01 respectively. Standard errors are in parenthesis. All dependent variables are in logs. The values in LIHA/HILA/HIHA rows denote the price differences between food deserts and LIHA/HILA/HIHA census tracts. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live at least 1 mile away from supermarkets. Vehicle access is used for food deserts definition 2. Race, gender, marriage, age, education, population in the census tracts and county fixed effects are also included.

Table C2. Regressions on EPI–Commonly Available Foods

	Definition 1			Definition 2		
	CEPI	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income High-access	0.0001 (0.0002)	0.001*** (0.0002)	0.001*** (0.0003)	0.0004*** (0.0001)	0.0004** (0.0002)	0.001*** (0.0003)
High-income Low-access	0.0001 (0.0002)	-0.0001 (0.0002)	-0.00003 (0.0004)	0.001*** (0.0002)	0.001** (0.0002)	0.001*** (0.0004)
High-access High-income	0.0004 (0.0002)	0.001*** (0.0002)	0.002*** (0.0004)	0.001*** (0.0002)	0.001*** (0.0002)	0.001*** (0.0003)
Observations	10403	10403	10403	10403	10403	10403

Note: *, **, *** denotes significance levels at 0.1, 0.05 and 0.01 respectively. Standard errors are in parenthesis. All dependent variables are in logs. The values in LIHA/HILA/HIHA rows denote the price differences between food deserts and LIHA/HILA/HIHA census tracts. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live at least 1 mile away from supermarkets. Vehicle access is used for food deserts definition 2. Race, gender, marriage, age, education, population in the census tracts and county fixed effects are also included.

Table C3. Regressions on EPI– Elasticity of Substitution between Food Groups $\sigma = 2$

	Definition 1			Definition 2		
	CEPI	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income	0.002	0.04***	0.04***	0.01***	0.02**	0.03***
High-access	(0.002)	(0.010)	(0.011)	(0.002)	(0.008)	(0.009)
High-income	0.001	0.002	0.003	0.01***	0.03**	0.04***
Low-access	(0.002)	(0.012)	(0.013)	(0.003)	(0.012)	(0.014)
High-access	0.01***	0.07***	0.07***	0.01***	0.04***	0.05***
High-income	(0.002)	(0.010)	(0.012)	(0.002)	(0.010)	(0.011)
Observations	10403	10403	10403	10403	10403	10403

Note: *, **, *** denotes significance levels at 0.1, 0.05 and 0.01 respectively. Standard errors are in parenthesis. All dependent variables are in logs. The values in LIHA/HILA/HIHA rows denote the price differences between food deserts and LIHA/HILA/HIHA census tracts. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live at least 1 mile away from supermarkets. Vehicle access is used for food deserts definition 2. Race, gender, marriage, age, education, population in the census tracts and county fixed effects are also included.

Table C4. Regressions on EPI– Within Brand Elasticity of Substitution (σ_g^w) =7, Across Brand Elasticity of Substitution (σ_g^a)=4

	Definition 1			Definition 2		
	CEPI	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income	0.002	0.07***	0.08***	0.01***	0.04***	0.04***
High-access	(0.002)	(0.016)	(0.017)	(0.002)	(0.014)	(0.015)
High-income	0.001	0.004	0.005	0.01***	0.05**	0.06***
Low-access	(0.002)	(0.019)	(0.020)	(0.003)	(0.021)	(0.022)
High-access	0.01***	0.11***	0.12***	0.01***	0.06***	0.07***
High-income	(0.002)	(0.018)	(0.019)	(0.002)	(0.016)	(0.017)
Observations	10403	10403	10403	10403	10403	10403

Note: *, **, *** denotes significance levels at 0.1, 0.05 and 0.01 respectively. Standard errors are in parenthesis. All dependent variables are in logs. The values in LIHA/HILA/HIHA rows denote the price differences between food deserts and LIHA/HILA/HIHA census tracts. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live at least 1 mile away from supermarkets. Vehicle access is used for food deserts definition 2. Race, gender, marriage, age, education, population in the census tracts and county fixed effects are also included.

Table C5. IRI Coverage in Store Counts and Sales at Census-Tract Level

	Overall	Food Deserts (Low Income Low Access)	Low Income High Access	High Income Low Access	High Income High Access
IRI Coverage in Store Counts					
Club stores	0.99	0.99	0.99	0.99	0.99
	(0.10)	(0.07)	(0.08)	(0.09)	(0.11)
Dollar stores	0.93	0.90	0.93	0.94	0.93
	(0.25)	(0.28)	(0.24)	(0.24)	(0.26)
Convenience stores	0.53	0.41	0.49	0.60	0.56
	(0.49)	(0.47)	(0.49)	(0.48)	(0.49)
Grocery stores	0.74	0.74	0.64	0.85	0.77
	(0.43)	(0.43)	(0.47)	(0.35)	(0.41)
Mass merchandisers	0.98	0.98	0.97	0.99	0.98
	(0.15)	(0.15)	(0.15)	(0.12)	(0.15)
Drug stores	0.86	0.88	0.83	0.91	0.87
	(0.33)	(0.32)	(0.36)	(0.28)	(0.32)
IRI Coverage in Store Sales					
Club stores	0.99	0.99	0.99	0.99	0.99
	(0.10)	(0.07)	(0.08)	(0.09)	(0.11)
Dollar stores	0.93	0.91	0.93	0.93	0.93
	(0.28)	(0.49)	(0.25)	(0.24)	(0.26)
Convenience stores	0.57	0.43	0.50	0.66	0.63
	(0.92)	(0.51)	(0.58)	(1.01)	(1.17)
Grocery stores	0.75	0.75	0.66	0.86	0.79
	(0.44)	(0.43)	(0.47)	(0.34)	(0.44)
Mass merchandisers	0.98	0.98	0.98	0.99	0.98
	(0.14)	(0.15)	(0.15)	(0.12)	(0.15)
Drug stores	0.89	0.90	0.86	0.92	0.89
	(0.41)	(0.38)	(0.43)	(0.27)	(0.47)
Observations	10,959	983	3,777	2,204	3,995

Note: Standard deviations are in the parenthesis. Food deserts definition 1 is used. When there are zero IRI and TDLinx stores in a census tract, then the IRI sales/counts coverage is 1. The calculations of IRI store coverage exclude tracts that have zero TDLinx stores but non-zero IRI stores (i.e. these tracts have missing IRI store coverage). As a result, 63, 17, 37, 5 and 13 tracts are excluded for IRI shares calculations of convenience stores, drug stores, dollar stores, mass merchandisers and grocery stores. So the IRI coverage in the table is lower bounds of the actual IRI coverage. The number of observations for each type of census tracts does not count the number of excluded tracts by each store type.