Does a Nutritious Diet Cost More in Food Deserts?

Linlin Fan*
University of Illinois at Urbana-Champaign
E-mail: lfan3@illinois.edu

Kathy Baylis
University of Illinois at Urban-Champaign

Craig Gundersen
University of Illinois at Urbana-Champaign

Michele Ver Ploeg
Economic Research Service, USDA

Selected Paper prepared for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28

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 consumption behavior. In this paper, we compare food prices in food deserts (defined as low-income, low-access census tracts) and non-food deserts to observe whether limited access to nearby grocery stores is associated with a higher price for a complete diet. If a nutritionally complete diet costs 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). To address potential biases, we control for both product heterogeneity and variety availability. We have four central findings. First, assuming consumers shop in both their own and neighboring census tracts, prices for the 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 price differences between food deserts and all types of non-food deserts are small. The EPI in food deserts is only 3% to 8% higher than similar census tracts with higher store access and 3% to 6% higher than high-income low-access census tracts. Third, we find that if consumers are constrained to shop within their home census tract, the difference in EPI increases to 22% between food deserts and low-income high-access census tracts and 9% between food deserts and high-income low-access census tracts. Fourth, prices measured by conventional means such as average or median cost of a TFP basket are not different in food deserts. In sum, residents of food deserts do not face substantially higher prices for similar consumption bundles as those living in non-food deserts; instead, the primary difference is with respect to the smaller variety of foods available in food deserts. This finding has implications for the mechanisms through which policymakers may wish to enhance access to food in low-income, low-access areas.

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

[•] We would like to thank Paula Dutko, Joseph Llobrera, Lisa House and Parke Wilde for their help in understanding 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, Nielsen or the IRI.

Recent research has found that in the United States, limited access to healthy food is associated with a lower consumption of fruits and vegetables (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). Areas with limited food access and low average incomes are often referred to as food deserts. Several federal, state, and local initiatives have emerged in response to the challenge of food deserts including a $400 million Healthy Food Financing Initiative to encourage large grocery retailers to move into underserved areas. Since 2001, California, Nevada, New Mexico, Texas, Oklahoma, Louisiana, Illinois, New York, Pennsylvania, D.C., and Maryland have enacted legislations aimed at increasing the number of healthy food retailers or have subsidized local stores to provide fresh food to increase the availability of fruits and vegetables.[1]

A central concern about food deserts is that food deserts are thought to have higher food prices[2] and less access to foods needed for a healthy diet than non-food deserts (e.g. Chung and Meyers 1999; Block and Kouba 2005; Andreyeva et al. 2008). It is, however, unclear whether prices of a nutritious diet are really higher in food deserts. Most studies that compare food prices in food deserts and non-food deserts are case studies focus on a single community (e.g. Chung and Meyers 1999; Block and Kouba 2005; Andreyeva et al. 2008) and use prices from one or two store chains (e.g. Hatzenbuehler, Gillespie and O'Neil 2012). Other existing studies primarily compare prices of particular food stuffs, particularly fresh food such as fruits and vegetables (e.g. Hayes 2000; USDA 2009) or average prices of any food using Universal Product Codes (UPC) (Broda, Leibtag and Weinstein 2009). In this paper, we compare the price of a nutritious diet in food deserts (low-income, low-access census tracts) to non-food deserts. We

---

[1] For example, Pennsylvania has initiated Fresh Food Financing Initiatives to provide funding to qualified local food stores to supply fresh food in underserved low-income communities (Bitler and Halder 2010).
[2] We use prices and price indices interchangeably.

specifically compare food desert prices to census tracts of similar income but higher access, and higher income but similar access to attempt to differentiate the effect of access from that of income. No other research that we know of compares food prices of a broad set of food products between food deserts and non-food deserts on a national scale with a comprehensive list of store price data.

Work using food price indices often does not account for product heterogeneity or variety bias when comparing food prices. Concern about product heterogeneity arises when one uses unit costs for broad food categories instead of the specific price for a food item when calculating a price index. For example, one might use the average price of bread per pound versus the price for Wonder Classic white bread per pound. This method cannot disentangle whether stores sell higher prices of identical products in poor neighborhoods or limited higher quality-varieties of these products. Broda, Leibtag and Weinstein (2009) find after controlling for product heterogeneity using UPC or barcode fixed effects, households from poor neighborhoods, contrary to earlier findings, pay less for the same food item than households from rich neighborhoods.

A second concern about food prices when they are defined over broad categories is that they do not account for the possibility that some areas have much wider product variety than other areas.[3] Insofar as products are not perfect substitutes, consumers, in general, prefer variety. Controlling for variety bias can substantially change price indices. For example, Handbury and Weinstein (2014) show that after controlling for variety availability across cities, contrary to previous findings, larger cities have lower average food price indices than smaller cities.

[3] The variety bias may occur when there are fewer food groups, brands or UPCs within a brand in stores of food deserts compared to stores in non-food deserts.

In this paper, we address both of these potential sources of bias by calculating variety-adjusted prices for food deserts versus other census tracts. We use 2012 detailed store sales data of a nationally representative geographic sample from Information Resources, Inc. (IRI) 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 that is defined with respect to low-income households' purchasing behavior and nutritional guidelines. Then to address the product heterogeneity and variety bias, 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).[4] The price index is based on a minimum cost diet that achieves a given level of utility assuming Constant Elasticity of Substitution (CES) preferences. 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 variety bias. The VA uses both estimated elasticities of substitution and national expenditure share of each UPC to adjust for substitution between UPCs and weight its importance in households' consumption and utility. We control for the product heterogeneity by using UPC prices rather than average costs for broad food categories.

To study the differences in prices and varieties, we regress the variety-adjusted prices (EPI) against an indicator for food deserts (an interaction term between a low-income census tracts dummy variable and a low-access census tracts dummy variable) along with a low-income census tracts dummy variable and a low-access census tracts dummy variable, neighborhood

[4] Broda and Weinstein (2010) construct the EPI for all goods that consumers buy including non-food items such medicine, electronics and appliances for the whole U.S. for each year. Handbury and Weinstein (2014) devise the EPI for all food at a city level. In this paper, we focus on a census-tract level EPI for food included in the TFP.

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 dummy variables. Our regression is a hedonic regression where the food prices of a neighborhood are a function of characteristics of the neighborhood. Further, we regress the components of EPI, which are conventional or unadjusted price indices (CEPI) and variety adjustment (VA), on food deserts to see how prices for foods that are common across census tracts and access to variety differ. Specifically, we compare two sets of the CEPI, VA and EPI based on different definitions of households' shopping areas. The first set of food price indices use store prices only within the census tract, which implicitly assumes households only shop in their census tract. The second set is based on store prices not only within the census tracts but also in the contiguous census tracts, which assumes that households potentially shop in both. We restrict our analysis to urban census tracts to ensure that both the food deserts and non-food deserts are urban census tracts and the results are more comparable to each other. We compare the variety-adjusted price indices (EPI) with other commonly used measures of TFP cost and check the robustness of our results against various definitions of food deserts and different assumptions of the elasticities of substitution between foods.

Our paper makes several contributions to the literature on food deserts. First, 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 much 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 within United States which allows us to speak more broadly to the price effect of living in a food desert.

Our central findings are as follows. First, when we assume consumers can purchase groceries in both their own and neighboring census tracts, we find that the price index without adjusting for variety available (CEPI) is not statistically significantly different between food deserts and all types of non-food deserts. Second, after accounting for the variety bias, the EPI in food deserts is 3% to 8% higher than low-income high-access census tracts, 3% to 6% higher than high-income low-access census tracts and 3% to 11% higher than high-income high-access census tracts. Thus, households living in food deserts face a smaller variety of groceries and a slightly higher food price index than households in census tracts with greater access to stores or households in census tracts with higher income. Third, we find our measures are highly sensitive to the definition of the consumers' shopping area. When we assume consumers only shop within their home census tracts we find that the EPI in food deserts is 22% higher than low-income high-access census tracts, 9% higher than high-income low-access census tracts and 25% higher than high-income high-access census tracts. Thus, consumers who are constrained to purchase food within their home census tracts face a much higher price than those who can shop in both their own and neighboring census tracts. Fourth, compared with average and median TFP costs that use county and state average/median prices to impute missing food prices, the EPI captures greater differences in TFP prices between food deserts and all types of non-food deserts.

Theoretical Model

We want to estimate a price index that includes the fact that consumers gain utility from having access to variety. To account for the gains from variety, we use the theoretical model of Broda and Weinstein (2010) and Handbury and Weinstein (2014) which assumes that consumers have Nested Constant Elasticity of Substitution (CES) preferences.[5] Food is split into three tiers

[5] The CES assumption makes the price indices estimable and tractable. This assumption is also constantly used in many economic geography models such as Krugman (1991). Moreover, Feenstra and Weinstein (2010) show the

within the nested framework. All foods are, first, categorized into the 26 food groups in the TFP[6] and second, into different brand-products.[7] Every brand-product is composed of different UPCs. Detailed examples of each food tier are given in the data section. Based on Broda and Weinstein (2010), the Constant Elasticity of Substitution (CES) utility of a representative consumer who lives in census tract c and is assumed to purchase a nutritious diet following the TFP within a census tract or contiguous census tracts is denoted by equation (1).

$$U_c = \left[\sum_{g \in G_c} \left(X_{gc} \right)^{\frac{\sigma}{\sigma - 1}} \right]^{\frac{\sigma - 1}{\sigma}} \qquad (1)$$

The elasticity of substitution across food groups is denoted by σ and G_c is the set of all available food groups (g) available in census tract c and its contiguous census tracts. The consumers' consumption of food group g in census tract c and its contiguous census tracts, X_{gc} is defined as

$$X_{gc} = \left[\sum_{b \in B_{gc}} \left(X_{bgc} \right)^{\frac{\sigma_g^a}{\sigma_g^a - 1}} \right]^{\frac{\sigma_g^a - 1}{\sigma_g^a}} \qquad (2)$$

where σ_g^a is the elasticity of substitution across brand-products within a food group and the set of all brand-products (b) available of food group g in census tract c and its contiguous census tracts is denoted as B_{gc}.

The consumption of brand-product b within food group g, X_{bgc}, is defined as

---

CES utility function has similar price levels to translog (a second-order approximation of an arbitrary expenditure system). Therefore, the CES assumption is unlikely to alter the results substantively.

[6] There are 29 TFP food groups in total. But for a male aged from 19 to 50, there are three food groups with zero recommended weekly consumption. Thus we exclude those three food groups and use 26 TFP food groups in the price index calculations. The details are provided in the data section. From now on, we use TFP food groups and non-zero weight TFP food groups interchangeably.

[7] The brand-products are the same as brand-modules used in Broda and Weinstein (2010) and Handbury and Weinstein (2014).

$$X_{bgc} = \left[\sum_{u \in U_{bgc}} \left(d_{ubgc} X_{ubgc} \right)^{\frac{\sigma_g^w}{\sigma_g^w - 1}} \right]^{\frac{\sigma_g^w - 1}{\sigma_g^w}} \qquad (3)$$

where d_{ubgc} is the taste parameter or utility weight,[8] the elasticity of substitution between UPCs within brand-products is σ_g^w, the set of all UPCs in brand-product b food group g in census tract c and its contiguous census tracts is U_{bgc}, and the individual's consumption of UPC u within the brand-product b and food group g in census tract c and its contiguous census tracts is X_{ubgc}. Consistent with previous work, we assume a constant σ_g^w and σ_g^a for each food group.

Based on Broda and Weinstein (2010), the minimum cost function to achieve one unit of utility from consuming brand-product b in equation (3) is given in equation (4).[9]

$$C_{bgc} = \left[\sum_{u \in U_{bgc}} \left(\frac{P_{ubgc}}{d_{ubgc}} \right)^{1-\sigma_g^w} \right]^{\frac{1}{1-\sigma_g^w}} \qquad (4)$$

where P_{ubgc} is the price of UPC u in census tract c and its contiguous census tracts. We assume consumers have homogeneous tastes towards UPC u across census tracts, which indicates that d_{ubgc} is constant across the nation and d_{ubgc} equals d_{ubgn}.[10] Then based on the minimum unit cost in (4), the national expenditure share of a UPC in a census tract in a particular brand-product reflects the utility-adjusted price of the UPC. This is illustrated in equation (5)

$$\ln \frac{P_{ubgc}}{d_{ubgc}} = \left(\frac{\ln S_{ubgn}}{1 - \sigma_g^w} \right) + \ln C_{bgc} \qquad (5)$$

[8] Although Broda and Weinstein (2010) and Handbury and Weinstein (2014) treat d_{ubgn} as the measure of quality of the UPC u, we think of d_{ubnc} as the taste parameter or utility weight associated with UPC u. Because the variable d_{ubgn} is essentially how much utility the representative consumer can get from the consumption of one unit of UPC u.
[9] The minimum cost of achieving one unit of utility from consuming food group g and from consuming all food available in the census tract and contiguous census tracts is similar to equation (4).
[10] This assumes that individuals in different census tracts have the same taste towards foods. This is the similar assumption that Feenstra (1994), Broda and Weinstein (2010) and Handbury and Weinstein (2014) use.

where the utility-adjusted prices of UPCs are indexed as P_{ubgc} / d_{ubgc} (or the minimum cost of achieving one unit of utility from consuming UPC u). The national expenditure share of a specific UPC is given by S_{ubgn}, which is described in detail in the data section. The elasticities of substitution within brand-products are greater than one given that brand-products are substitutes rather than complements within a food group. The national expenditure share S_{ubgn} reflects the utility-adjusted prices in a census tract. For example, if there are two UPCs that exist in a brand-product with the same price, but only the UPC with lower national expenditure share is available in a census tract, this indicates that the UPC available in the census tract is less important than the other UPC in providing utility to the consumer. Based on the local prices P_{ubgc} and national expenditure shares S_{ubgn}, Feenstra's (1994) and Handbury and Weinstein's (2014) critical insight is that obtaining the relative minimum unit cost (P_{ubgc}/ P_{ubgn}) between a local geographical scale such as a census tract and the whole nation will eliminate the taste parameter d_{ubgc} and thus yield the variety-adjusted prices in each census tract.

Figure 1 illustrates the intuition behind the potential gains from variety (from Feenstra 2010). Suppose a consumer gains 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 missing good compared to the available one. The increase in 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) and widely used in trade economics.

Methods

Based on the theoretical model, we adapt Handbury and Weinstein (2014)'s methods to construct a census-tract level TFP food price index. The variety-adjusted price index, the EPI (Exact Price Index), is the relative minimum cost in a census tract of obtaining the TFP, assuming consumers purchase food only within their census tract and neighboring contiguous census tracts, and have CES preferences. Of course, many consumers will purchase food outside of their census tract and neighboring census tracts but our interest in this paper is about the prices facing consumers in nearby stores to reflect their local food market. In this we follow the food deserts literature whereby persons are assumed to only purchase food within their census tract and contiguous census tracts.

There are two major components in the EPI. One is the unadjusted price index, the CEPI (Conventional Exact Price Index) that measures prices of food that are available in consumer's census tract and contiguous census tracts is given by

$$CEPI_c = \prod_{g \in G_c} \left(CEPI_{gc} \right)^{W_g} \qquad (6)$$

where,

$$CEPI_{gc} = \prod_{u \in U_{gc}} \left(\frac{V_{uc} / Q_{uc}}{\sum_c V_{uc} / \sum_c Q_{uc}} \right)^{w_{uc}} \qquad (7)$$

V_{uc} and Q_{uc} are local expenditure and local quantity spent on UPC u across all stores in census tract c and its contiguous census tracts respectively, U_{gc} is the set of all UPCs that are available in the census tract c and its contiguous census tracts, W_{uc} is the log-ideal CES Sato (1976) and

Vartia (1976) weights. The variable W_{uc} essentially gives more weight to UPCs that have a larger local market share. A detailed definition of W_{uc} is provided in appendix A. The Conventional Exact Price Index for food group g in census tract c ($CEPI_{gc}$) is calculated as the geometric mean of local versus national relative price ratio in food group g weighted by W_{uc}. The *CEPI* for the census tract ($CEPI_c$) is the geometric mean of *CEPI* for each individual food group ($CEPI_{gc}$) adjusted by the weight of each food group g in constructing a TFP for a male age 19 to 50 (W_g). Notably, the W_g 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 the elements of a set of price, quantity and expenditure indices but may overlook the different importance in nutrition of each TFP food category. Because we are more interested in calculating the price indices of a nutritious diet, not a quantity or expenditure index of food, we use TFP weights in the EPI calculations.

The other component of the EPI, Variety Adjustment (VA) or a measure of variety availability is given by

$$VA_c = S_c^{\frac{1}{1-\sigma}} \prod_{g \in G_c} \left(VA_{gc} \right)^{W_g} \qquad (8)$$

where,

$$VA_{gc} = \left(S_{gc} \right)^{\frac{1}{1-\sigma_g^a}} \prod_{b \in B_{gc}} \left(S_{bc} \right)^{\frac{W_{bc}}{1-\sigma_g^w}} \qquad (9)$$

W_{bc} is the log-ideal CES Sato (1976) and Vartia (1976) weight defined in appendix A, σ is the elasticity of substitution between food groups, σ_g^a is the elasticity of substitution across brand-products within food group g and σ_g^w is the elasticity of substitution within a brand-product from food group g. They are assumed to be constant for each food group g. 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. 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 diving national expenditure on the 4 UPCs of brand-product b by the national expenditure on all 10 UPCs of that brand-module.

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 and elasticity of substitution between food groups is σ. Notably, 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 $S_c^{\frac{1}{1-\sigma}}$ to measure the importance of available food groups in VA_c. We use $S_c^{\frac{1}{1-\sigma}}$ for VA_c in our robustness tests. 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 (Exact Price Index) in census tract c is the product of $CEPI_c$ and VA_c:

$$EPI_c = CEPI_c VA_c \qquad (10)$$

The Conventional Exact Price Index ($CEPI_c$) can be thought of as the correct way to measure the price level of the census tract if all TFP foods 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 using 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} provided a utility-adjusted count of missing UPCs in census tract c and its contiguous census tracts. The exponent weight is the elasticity of substitution between UPCs within a brand-product for a food group, σ_g^w.

The UPC availability index implies that if the census tract misses a UPC that has a large national expenditure share (S_{bc}), which indicates the missing UPC is important in utility, then the variety-adjusted price (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 matter much. Similarly, the brand-product availability index is $\left(S_{gc}\right)^{\frac{1}{1-\sigma_g^a}}$. The impact of brand-product availability on the EPI depends on how large an expenditure share the brand-product has nationally (S_{gc}), implying its importance and whether the brand-product has close substitutes (σ_g^a).

Lastly, the food group availability index is $S_c^{\frac{1}{1-\sigma}}$ which combines the importance of the existing food groups (S_c) and substitutability between food groups (σ). For all three availability indices, the lower the value, the more goods are available in the census tract and its contiguous census tracts, and thus a lower variety-adjusted price index.

After constructing the census tract price indices, we compare unadjusted price indices (CEPI in equation 6), variety adjustment indices (VA in equation 8) and adjusted price indices (EPI in equation 10) based on the OLS regressions given in equation (11).

$$y_{ij} = \alpha_0 + \alpha_1 FD_{ij} + \alpha_2 LA_{ij} + \alpha_3 LI_{ij} + x_{ij}\beta + C_j + \varepsilon_{ij} \qquad (11)$$

where y_{ij} is the log of CEPI, VA or EPI for census tract i in county j; The indicator variable FD_{ij} takes the value of one if the census tract i in county j is a food desert (low-income low-access census tracts). The indicator variable LA_{ij} is equal to one if the census tract i in county j is a low-access census tract, and LI_{ij} is an indicator variable for a low-income census tract. So the reference group is high-income, high-access census tracts. The detailed definitions of low-income and low-access census tracts are provided in the data section. We also control for other neighborhood characteristics (x_{ij}) such as population, education, gender, marital status, age and racial composition to strip out some of 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 (Handbury and Weinstein 2014). 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.

Based on the regression results, we compare how the unadjusted price index (CEPI), variety adjustment (VA) and adjusted price index (EPI) differ across food deserts, low-income high-access census tracts, high-income low-access tracts and high-income high-access census tracts. 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 (census tract socio-demographic variables x_{ij} and county fixed effects C_j), a combination of $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ will exactly capture the mean price or variety index of the subgroups of census tracts. In a more fully specified model, these coefficients can be combined to construct regression-adjusted mean price or variety index for different types of census tracts. To be more specific, the regression-adjusted average price or variety index for food deserts is $\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3$, the regression-adjusted average price or variety index for low-income high-access is $\alpha_0 + \alpha_3$, for high-income low-access census tracts is $\alpha_0 + \alpha_2$, and for high-income high-access census tracts is α_0. Therefore, the regression-adjusted average difference in price indices and variety adjustment between food deserts and low-income high-access census tracts is $(\alpha_0+\alpha_1+\alpha_2+\alpha_3)-(\alpha_0+\alpha_3)=\alpha_1+\alpha_2$ and represents the access effect. Similarly, the difference between food deserts and high-income low-access census tract is $(\alpha_0+\alpha_1+\alpha_2+\alpha_3)-(\alpha_0+\alpha_2)=\alpha_1+\alpha_3$ and represents the income effect. The difference between food deserts and high-income high-access census tracts is $(\alpha_0+\alpha_1+\alpha_2+\alpha_3)-\alpha_0=\alpha_1+\alpha_2+\alpha_3$ [11]

We use two different measures of food deserts to check the robustness of our results. The first set of food deserts indicator variables are extracted from Food Access Research Atlas (FARA, USDA 2013). We create the second set of food deserts indicator variables using census tract income and access variables from National Household Food Acquisition and Purchase Survey datasets (FoodAPS), which use a newer round of data compared to the FARA. For both FoodAPS and FARA data, we use two definitions of food deserts to check the robustness of our results. One defines food deserts as low-income urban census tracts where a significant portion of people live at least one mile away from a supermarket. The other defines food deserts are low-income urban census tracts where a lot of households do not have access to vehicles. Both food deserts definitions are described in detail in the data section.

In our main analysis, we construct two sets of price indices: CEPI, VA and EPI using different definitions of shopping areas. For the first set of price indices, we use store sales within both the census tract and contiguous census tracts, implicitly assuming consumers can shop in both their home and contiguous census tracts. In the second set of price indices, we assume

[11] Alternatively, one could include three dummy variables, one for low-income high-access census tracts, one for high-income low-access census tracts and the other for high-income high-access census tracts in the regressions. Then the food deserts are the base group and the coefficients on the three dummy variables are the price differences between food deserts and the respective comparison groups. The price differences calculated from this specification will be identical to the one used in the paper.

consumers can only shop within their home census tracts and use store sales only within the census tracts to construct and compare the price indices.

In all of our analysis, we exclude census tracts where there are no stores selling any TFP food groups in both the home and contiguous census tracts because the EPI for those census tracts are infinitely positive.[12] Admittedly, omitting those census tracts will likely bias the estimated price differences between food deserts and all types of non-food deserts downwards because we would expect food deserts where none of the TFP food groups are available to have the highest EPI. In the second set of price indices where we assume consumers are constrained to shopping within their home census tract, we exclude all of the census tracts with no TFP food groups available, reducing our number of observations further.

Last, we compare other commonly used price measures with the EPI. We use the average and median TFP cost where we impute the prices of food groups missing in both the census tracts and contiguous census tracts with county and state averages and medians to test the veracity of EPI. In all price comparisons (average/median TFP cost and the EPI), we exclude census tracts that do not have any stores that sell any TFP food groups in the census tracts and contiguous census tracts, and thus we do not need to address the missing food groups in those census tracts. Notably, in all of our different EPI calculations, we do not impute any prices for missing food groups but rather use the TFP weights to measure the penalty of missing food groups on price indices. However, when we use other conventional measures of TFP cost, i.e. average and median TFP costs that do not address variety bias directly, we impute the prices of the missing food groups with the average and median state/county TFP food group prices.

Robustness Tests

---

[12] Notably, this problem primarily results from there being no IRI stores in these census tracts whereas there may actually be stores in these areas that are not included in the IRI data.

In the main regression analysis, we weight the importance of existing food groups using TFP weights. Alternatively, we assume that the elasticity of substitution between food groups is 2, which essentially uses the inverse of national expenditure share on the food groups available in the census tract and its contiguous census tracts S_c^{-1} to adjust for the availability of food groups. Additionally, because 2 is the lowest 1st percentile of the across brand-product elasticity of substitution within a food group (σ_g^a) and the elasticity of substitution between food groups is lower than the elasticities of substitution within food groups, we use 2 as the upper bound of the elasticity of substitution between food groups (σ).

Next, instead of arbitrarily defining the elasticities of substitution between food groups, we focus on census tracts that have a full set of core TFP food groups in its census tracts or contiguous census tracts (no missing core TFP food groups). We choose a set of core TFP food groups instead of all TFP food groups to retain as many as possible census tracts for the analysis. We choose the TFP food groups that have over 0.1 pounds of weekly recommended consumption for a male aged 19 to 50, which may indicate that these food groups are more important than others. As a result, seven food groups are deleted.[13] We select and run regressions on census tracts that have access to all of the remaining 19 TFP food groups in the census tracts or contiguous census tracts. Therefore, we do not need to estimate the elasticity of substitution between food groups for these census tracts. This change means we have 9,248 census tracts rather than 10,403 census tracts from our analysis. For all other price calculations, we include all

[13] To construct a core set of TFP food groups, we choose food groups that have recommended consumption of over 0.1 pounds per week. As a result, "whole grain cereals including hot cereal mixes" (0.08 lbs), "all cheese, including cheese soups and sauces" (0.07 lbs), "bacon, sausage, and lunch meats including spreads" (0.02 lbs), "coffee and tea" (0.01 lbs), "sugars, sweets and candies" (0.08 lbs), "soups (dry)" (0.02 lbs) and "frozen/refrigerated entrees including pizza, fish sticks and frozen meals" (0.01 lbs) are deleted. They account for only 0.29 lbs of the total 39.86 lbs of consumption. As a result, 9,249 census tracts have a full range of core TFP food groups either within the census tracts or in their contiguous census tracts (1,156 census tracts are excluded).

census tracts that have at least a store that sell some TFP food groups in the census tracts and contiguous census tracts when consumers are assumed to shop in their home and contiguous census tracts.

We estimate the elasticity of substitution across brands and within brands for a TFP food group from the averages of overlapping food groups defined by Handbury and Weinstein (2014). For example, one TFP food group overlaps with several food groups used in Handbury and Weinstein (2014). Therefore, we use other estimates of elasticity of substitution across brands and within brands from existing marketing literature (4 for elasticity of substitution across brands and 7 for the elasticity of substitution within brands) to check the robustness of our results (Dube and Manchanda 2005).

Data

We use the Information Resources, Inc. (IRI) retailer scanner data 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 National Household Food Acquisition and Purchases Survey (FoodAPS) conducted by the USDA. Fifty primary sampling units (counties or a group of counties) are selected randomly from all counties in the nation. The IRI retailer scanner data have weekly sales for each food item at the UPC level for almost all major national and regional chain stores in the 108 counties in the primary sampling units.[14] Among the 10,687 census tracts in the 108 counties, 239 census tracts do not have any stores that sell any TFP food groups within the census tracts or their contiguous census tracts. Additionally, 45 census tracts with

[14] The covered stores include mostly national and regional chain stores of various store types. It includes mass merchandiser stores, drug stores and convenience stores that sell some food, dollar stores, large grocery stores and club stores. One drawback is that local independent stores, farmers' market, and mom and pop stores are not included. The USDA (2008) found that around 80% of food at home expenditure is spent in chain stores. 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.

stores that sell TFP food groups in the census tracts or contiguous census tracts are deleted because the population of the census tracts is zero. As a result, 10,403 census tracts remain in the sample.

As noted earlier, we also calculate another set of CEPI, EPI and VA based on store prices only within the census tracts. In this set of price indices, in additional to the 239 census tracts where none of the TFP food groups exist in both the census tracts and the contiguous census tracts, we exclude 5,605 census tracts that do not have any stores selling any of the TFP food groups within the census tract but have TFP food groups in the contiguous census tracts. Additionally, 8 census tracts where there are stores selling TFP food groups inside but have zero population are excluded in this second analysis. The IRI weekly sales datasets have the total value sold, total quantity sold, brand, product description of the sold UPCs at each store or regional market area (RMA).[15]

We define a nutritionally complete diet following the TFP. The TFP is one of the USDA-designed food plans specifying food categories and quantities that provide an affordable and adequate nutrition based on the Dietary Guidelines for Americans and the My Pyramid Food Guidance System. For all of the 29 TFP food categories, the TFP has recommended consumption of each food category for fifteen age and gender groups. As noted earlier, three of the 29 TFP categories, namely “popcorn and other whole grain snacks”, “milk drinks and milk desserts”, and “soft drinks, sodas and fruit drinks” have zero recommended weekly consumption for a male aged 19 to 50. So, we end up using 26 categories from the TFP. We use the TFP food consumption recommendation for a male aged 19 to 50. The TFP is commonly used in U.S.

[15] 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 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.

government policies to estimate the cost of a nutritious but cheap or "thrifty" diet and serve as the basis for the maximum Supplemental Nutrition Assistance Program (SNAP) monthly benefit. A full list of TFP categories and recommended pounds is provided in appendix table B.

As the first step to construct a local TFP EPI, each UPC is categorized into different food groups defined by the TFP using the brand and product description of the UPCs. The structure of the food categorization is illustrated in figure 2. Each food group is then split into several brand-products that are finer food categories within a food group. For example, Yoplait Original yogurt is defined as a brand-product within the whole milk, yogurt and cream food group. Yoplait Original is the brand or the manufacturer and yogurt is a product. A brand-product incorporates several UPCs. For example, the Yoplait Original yogurt brand-product includes 0.6 oz Yoplait Original yogurt strawberry flavor and 0.6 oz Yoplait Original yogurt blueberry flavor, each with distinct UPCs. Food items with different sizes or flavors will also have different UPCs. But changing the slogan on the same 0.6 oz Yoplait Original yogurt strawberry flavor will not result in a change in the UPC. So if there is no meaningful quality or size change in the food item, there will be no changes in UPCs (Broda and Weinstein 2010).

The second step to construct the price indices is to estimate elasticities of substitution between foods. We obtain estimated elasticities of substitution from Handbury and Weinstein (2014). Handbury and Weinstein (2014) have 63 food groups in total and each of the 63 food groups have an estimated elasticity of substitution across brands (σ_g^a) and within brands (σ_g^w) respectively. One TFP food group overlaps with several food groups from Handbury and Weinstein (2014). For example, the TFP food group "all potato products" overlaps with six food groups in Handbury and Weinstein (2014), namely "prepared food-ready-to-serve" that includes potato salad-canned, "prepared food-dry mixes" that includes potatoes-specialty-dehydrated and potatoes-mashed-dehydrated, “vegetables-canned” that include potatoes-canned and sweet potatoes & yams-canned, “snacks” that includes potato chips and potato sticks, “vegetables-frozen” that includes potato-frozen, and “fresh produce” that includes fresh potatoes. The elasticities of substitution across brands and within brands are constant for each food group in Handbury and Weinstein (2014). As a robustness check, we use 4 and 7 as the across the within brand elasticity of substitution that, as discussed above, are commonly used in marketing literature (Dube and Manchanda 2005).

We extract and create food deserts indicator variables from two different datasets. One set of food deserts indicator variables are from FARA, which uses 2010 TDLinx and STARS store lists, 2006-2010 American Community Survey (ACS) and 2010 Census of Population and Housing to define food deserts. The other set of food deserts indicator variables are from FoodAPS, which uses 2012 TDLinx and STARS store lists and 2008-2012 American Community Survey (ACS) to define census tract median income, poverty rate, the number of people with low access to supermarkets and the number of housing units without access to vehicle. Then we use the FoodAPS income, poverty rate and access variables to define food deserts.

Although the two sets of food deserts indicator variables are based on different datasets, the food deserts definitions are similar. For both FARA and FoodAPS, food deserts are defined as a low-income low-access census tract following USDA (2013). A low-income census tract is defined as a census tract with a poverty rate greater than or equal to 20 percent; or a census tract with median income less than or equal to 80 percent of state median income; or a metropolitan census tract with median family income less than or equal to 80 percent of metropolitan area’s median income. A low-access census tract is defined as a census tract where 500 people in the census tract or at least 33% of the people in the census tract are at least 1 mile away from a supermarket in urban areas. A supermarket is defined as a store which has over 2 million annual sales and has all major food departments such as fresh produce, fresh meat and poultry, dairy, dry and packaged foods and frozen foods. The FARA calculates the distance from a household to the nearest supermarket by first assigning each household to a ½ km grid and then measuring the distance from the centroid of the grid to the nearest supermarket. In contrast, the FoodAPS uses the centroid of block groups where the households reside.

Following the USDA (2013), we also test alternative definitions of food deserts. We include lack of access to a vehicle instead of distance to supermarkets as an alternative definition of food deserts. 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 or 33 percent of the population live more than 20 miles from the nearest supermarket, regardless of vehicle access."

The demographic variables such as race, gender, marital status, age, education and population used in the regression analysis are taken from the 2008-2012 American Community Survey.[16]

Summary Statistics

Table 1 presents the summary statistics of the sample by income and access to supermarkets. We define food deserts in table 1 using being more than 1 mile from a supermarket for urban census tracts with income and store access variables from FoodAPS dataset. The summary statistics using vehicle access and FARA datasets are similar to that in table 1. All the food prices (average/median TFP cost, the CEPI and the EPI) and measures of food availability (the number

[16] 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.

of TFP food groups, the number of UPCs and the VA) in this table are all calculated using store sales both within the census tracts and contiguous census tracts. As noted before, because the EPI would suggest the census tracts where none of the TFP food groups are available both within the census tracts and contiguous census tracts, have infinitely positive food prices, we exclude those census tracts in all of our analysis including summary statistics and all regressions.

We find that the average and median TFP cost are the lowest in food deserts among the four types of census tracts. It suggests that the average and median TFP cost that do not address variety bias are actually lower in food deserts than all non-food deserts. We construct the average TFP cost by first calculating the average price for each food group (the total expenditure divided by total quantity spent on the food group in the census tract and its contiguous census tracts) and use the county-level and state-level average price to impute the price for the missing food groups. To build the median TFP cost, we first obtain the median price for each food group in the census tract and its contiguous census tracts and use the county-level and state-level median price to impute the missing food groups. Then we multiply the average and median price of each food group with the recommended pounds of consumption per week to get the average and median TFP cost. As noted before, the TFP average and median costs are the only cases where we impute prices for missing food groups. The CEPI, VA and EPI are not imputed.

When comparing the food variety availability between food deserts and non-food deserts, we find that food deserts and non-food deserts have similar number of TFP food groups.[17] This measure of variety availability does not account for the degree of substitutability among UPCs within a TFP food group or between food groups, and the importance of each UPC in

[17] The count of TFP food groups for each census tract accounts for the stores sales that exist in the contiguous census tracts too. Because a high-access census tract is defined by whether the nearest supermarket is within a radius of a census tract, thus the high-access census tract may not have a supermarket within its boundaries or contiguous census tracts. Therefore, high-access census tracts may not have all TFP food groups.

consumers' utility. To address both issues, we construct the unadjusted price indices CEPI from equation (6) and variety adjustment VA (8) based on national expenditure shares and estimated elasticities of substitution. We multiply CEPI with VA to obtain the adjusted price indices EPI that accounts for variety bias. We find that the variety adjustment (VA) based on equation 8 in food deserts is similar to that of low-income high-access census tracts (1.46 vs 1.49). But the VA is 17% higher in food deserts compared to high-income low-access census tracts (1.46 vs 1.25). Note that a higher VA implies lower access to variety. The prices of common food available in food deserts and non-food deserts (CEPI) such as packaged food are similar between food deserts and low-income high-access census tracts (1.02 vs 1.04). The CEPI is also similar between food deserts and high-income low-access census tracts (1.02 vs 1.05). Therefore, the variety-adjusted price indices EPI are similar between food deserts and low-income high-access census tracts (1.50 vs 1.56). Compared to high-income low-access census tracts, the variety-adjusted prices EPI are 15% higher (1.50 vs 1.30) in food deserts because of different access to variety. The differences in the CEPI, VA and EPI between food deserts and high-income high-access census tracts are similar to those between food deserts and high-income low-access census tracts. This suggests income matters more for the variety-adjusted price indices than access.

When comparing socio-demographic variables of census tracts in table 1, we find that the average census tract family income[18] in food deserts is less than half that of high-income census

[18] The average tract family income is the average of tract median family income in the sample. This is different from median family income in the sample. The median family income is expected to be lower than the average of tract median family income mechanically. When calculating the median family income, because there are more poor families than rich families, the median family income is likely to be skewed towards the income of poor families. However, when calculating the average of tract median family income, all tracts are put on the same weight automatically and are not skewed towards poor families or poor census tracts. Therefore, the median family income is likely to be lower than the average of tract median family income. For example, the average tract median family income for food deserts (low-income low-access census tracts) is 42715.42 which is similar to the family median income 47299 (Handbury, Rahkovsky and Schnell, 2015).

tracts and the poverty rate is 27% in food deserts compared to 7%-8% in high-income census tracts. Food deserts also have more unmarried, younger, less educated people and more African Americans and Hispanics compared to high-income census tracts.[19] All census tracts are from urban areas. Among the non-food deserts, low-income high-access census tracts are more similar to food deserts than high-income low-access neighborhoods in terms of their population, education, age, marital status and racial composition. These differences in demographic composition in census tracts implies that to truly compare prices among regions, we need to control for demographic characteristics that may affect taste and demand.

Table 1 presents the summary statistics in which all food prices and measures of food availability are created using stores sales within both home census tracts and contiguous census tracts. To compare how the measures change, we create the second set of summary statistics in table 2 using store sales only within the home census tracts and thus assume that consumers are constrained to shop only within their home census tracts. Therefore, as mentioned before, we exclude an additional number of census tracts where none of the TFP food groups are available but some or all of the TFP food groups are available in the contiguous census tracts in table 2. Similar to table 1, food deserts in table 2 are low-income urban census tracts where a significant portion of people in the census tracts are over 1 mile away from a supermarket based on FoodAPS datasets.

We find that contrary to the results in table 1, when we limit ourselves to prices only within the census tract itself, the TFP average and median cost is higher in food deserts compared to all types of non-food deserts and there are fewer food groups in food deserts than all types of non-

[19] Marital status is measured by the proportion of people who are married in the census tract. Education is measured by the proportion of households who have completed high school. We define the census tract as a white census tract where more than 50% of all population are white. Similarly African American and Asian American census tracts are defined as census tracts where over 50% of all population are African Americans and Asian Americans respectively.

food deserts. The variety adjustment (VA) is 13% higher in food deserts than low-income high-access census tracts. In other words, there are 13% fewer food varieties in food deserts than low-income high-access census tracts after adjusting for the elasticity of substitution between UPCs and the utility importance of each UPC. The difference in VA is even larger between food deserts and high-income census tracts. The VA is 19% and 39% higher in food deserts than high-income low-access census tracts and high-income high-access census tracts respectively. Similar to table 1, the price indices of foods commonly available in all census tracts are similar between food-deserts and all types of non-food deserts. However, because food deserts have much higher VA than all types of non-food deserts, the EPI in food deserts is 13% higher than low-income high access census tracts, 18% higher than high-income low-access census tracts and 37% higher than high-income high-access census tracts. In contrast, in table 1, the EPI in food deserts is similar to that in high-access low-income census tracts while the EPI is only 15% higher in food deserts compared to both high-income low-access census tracts and high-income high-access census tracts. The larger difference in EPI between food deserts and all types of non-food deserts in table 2 compared to table 1 suggests that food deserts matter much more for consumers who are constrained to shop within their home census tracts than consumers who can access stores in their contiguous census tracts.

The averages and disparity in socio-demographic variables between food deserts and all types of non-food deserts are similar in table 2 to those in table 1.

Results

In table 3 we present the differences in CEPI, VA and EPI between food deserts and all types of non-food deserts by different food deserts definitions. The low-income high-access rows in all tables show the differences in the price indices (CEPI, VA, EPI) between food deserts and low-income high-access census tracts, which we call the access effect. Similarly, the high-income low-access rows show the price index differences between food deserts and high-income low-access census tracts, which we call income effect. Last, the high-income high-access rows show the price index differences between food deserts and high-income high-access census tracts, which we call combined access and income effect.

In all of our tables for regressions results, namely table 3 to table 8, the first three columns show regression results with the food deserts indicator variables from FoodAPS. Because the FoodAPS food deserts indicator variables uses data from the 2012 TDLinx and STARS and 2008-2012 ACS, we denote this set of results as 2012 FoodAPS. The next three columns use the food deserts indicator variables from FARA that uses 2010 TDLinx and STARS, 2006-2010 ACS and the 2010 Census of Population and Housing. Therefore, we denote these results as 2010 FARA. As mentioned in the methods and data section, in all of our regression tables, we use two definitions of food deserts based on both sets of data. Definition 1 defines food deserts Food deserts definition 1 as low-income urban census tracts where a significant portion of people live at least 1 mile away from a supermarket. Definition 2 defines food deserts using income variables and vehicle access. In all regressions, we include census tract characteristics to control for demand factors that may affect food prices. Specifically, we control for population, age, education, gender, marital status, age and race of the census tract.

Table 3 presents the results that use store sales in both home and contiguous census tracts. The EPI is a product of VA and CEPI. Thus we can attribute the differences in EPI between food deserts and all types of non-food deserts to the differences in prices of commonly available foods and the differences in varieties of available foods separately. We find that except for the FARA food deserts definition 2 (based on vehicle access), food deserts have slightly higher variety-adjusted price indices (EPI) than low-income high-access census tracts. This result is different from what we find in the summary statistics in table 1 that the EPI is higher in low-income high-access census tracts compared to food deserts. The prices of foods that are commonly available in both food deserts and low-income high-access census tracts (CEPI) do not differ significantly between these two types of census tracts. However, there are slightly fewer varieties available in food deserts after adjusting the importance of UPCs in utility and substitutability between UPCs. Thus, the VA is higher in food deserts. Therefore, the variety-adjusted price index (EPI) is slightly higher in food deserts than low-income high-access census tracts. Specifically, for FoodAPS food deserts definition 1, the EPI is 5% higher in food deserts compared to low-income high-access census tracts because the VA is 4% higher in food deserts.

As for income effect that measures the difference in the price index between food deserts and high-income low-access census tracts, except for the FoodAPS food deserts definition 1 (at least 1 mile away from supermarkets), food deserts have 3% to 4% higher variety-adjusted price indices (EPI). Similar to the access effect, this result is mostly driven by the fact that food deserts have 3% to 4% fewer varieties of food available after adjusting the elasticities of substitution between UPCs and the importance of each UPC on utility (the VA is higher). The vehicle access food deserts definition (definition 2) seems to better capture the income effect compared to food deserts definition 1. Compared to high-income high-access census tracts, the CEPI is not statistically different in food deserts. But the VA is 3% to 7% higher in food deserts and thus the variety-adjusted price indices (EPI) are 4% to 8% higher in food deserts than high-income high-access census tracts.

We next use store sales only within the census tract to calculate the price indices, implicitly assuming consumers only shop within their census tracts, as opposed to the above results in table

3 which assume consumers shop both in their own and neighboring tracts. These results are given in table 4. As mentioned earlier, in all of our analysis (summary statistics and regressions), we exclude census tracts where no TFP food groups are available in both home and contiguous census tracts and thus 10,403 census tracts remain in the sample. In table 4, we exclude an additional number of the census tracts where none of the TFP food groups are available in home census tracts but some or all TFP food groups are available in the contiguous census tracts. Notably, table 4 and table 2 (summary statistics) are the only cases where we exclude an additional number of census tracts and thus have 4,835 census tracts in the analysis.

We find both stronger access and income effects in table 4 compared to table 3. Using all definitions of food deserts, the variety-adjusted prices (EPI) are higher in food deserts than low-income high-access census tracts because food deserts have significantly higher VA. Using FoodAPS food deserts definition 1, consumers moving from a food desert to a low-income high-access census tract, would experience a 2% increase in food price for the same foods (CEPI), but because variety adjustment (VA) is 20% higher, the food price index (EPI) would actually rise by 22%. We also observe a larger income effect and combined access and income effect in table 4 than table 3. Using FoodAPS food deserts definition 1, we find food deserts have 9% higher EPI than high-income low-access census tracts because food deserts have 9% fewer varieties (VA is 9% higher). The EPI in food deserts is 25% higher than high-income high-access census tracts because food deserts have 23% fewer varieties (VA is 23% higher). The larger access and income effects in table 4 compared to table 3 shows the importance of the definition of shopping area to detect the price differences between food deserts and non-food deserts. Further, they illustrate that for those consumers constrained to shopping within their census tract, both access and income has a much greater effect, and the food desert definition much greater meaning, than for those consumers who can access stores in neighboring census tracts.

Last, we compare the variety-adjusted prices with the average and median TFP cost (detailed definitions described in data section) in table 5. We find that using any of our definitions of food deserts, food deserts do not have significantly different average TFP costs than low-income high-access census tracts. For some definitions, food deserts even have slightly lower TFP median cost than low-income high-access census tracts. The average and median cost for the TFP in food deserts is either only slightly higher or insignificantly different than in high-income low-access census tracts. Compared to high-income high-access census tracts, the average TFP cost is 2% to 4% higher in food deserts, which is smaller than the 4% to 8% we find in table 3. The median TFP cost in food deserts is either slightly lower than or statistically insignificantly different from high-income high-access census tracts. 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. Thus, the EPI is more sensitive to capturing the price differences between food deserts and all types of non-food deserts.

Robustness Tests

In all of our robustness tests, we use store sales in both home and contiguous census tracts. Because one might be concerned that our results are driven by the elasticities of substitution used in creating our price indices, in table 6, 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 census tracts or contiguous census tracts (S_c^{-1}) to measure the importance of available food groups. The results are similar to table 3.

Table 7 demonstrates the difference in price indices for census tracts that have a full set of core TFP food groups. In all of other tables, we use TFP weights to weight the importance of available food groups in the census tracts. However, Handbury and Weinstein (2014) uses the national expenditure shares of food groups available in the census tracts and contiguous census tracts (S_c in equation 8) and the elasticities of substitution between food groups (σ) to weight the food groups' importance. To circumvent the problem of finding the correct elasticity of substitution between food groups, we focus on census tracts that have a full set of core TFP food groups either in the census tracts or in contiguous census tracts in this robustness test. By doing that, we lose 1,156 census tracts that do not have all TFP food groups, most of which are food deserts. We would expect food deserts without all TFP food groups have much higher variety-adjusted prices than non-food deserts if included. Thus we expect the variety-adjusted price differences to be smaller between food deserts and non-food deserts in the restricted sample of 9,249 census tracts that have all core TFP food groups. Table 7 shows that by FoodAPS food deserts definition 1, the price of common food is not significantly different between food deserts and low-income high-access census tracts but there are 2% fewer varieties available in food deserts. As a result, the variety-adjusted prices are 2% higher in food deserts than low-income high-access census tracts (access effect), slightly smaller than the 5% reported in table 3. When compared to high-income low-access census tracts (income effect), food deserts have 2% higher variety-adjusted food prices. Food deserts have 3% higher EPI than high-income high-access census tracts because food deserts have 3% fewer varieties. Using FoodAPS food deserts definition 2 yields similar results as using FoodAPS food deserts definition 1 while FARA food deserts definitions show no significant access effect but significant income effect.

In table 8, 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) to test the robustness of our results (specifically the VA and EPI since the CEPI will not be changed). In the previous tables, we estimate the elasticity of substitution across brands (σ_g^a) and within brands (σ_g^w) of a TFP food group by taking the average of the elasticities of substitution from several overlapping food groups defined by Handbury and Weinstein (2014). In table 8, by FoodAPS food deserts definition 1, we find that the average variety difference between food deserts and low-income high-access census tracts increases to 7% (access effect) when assuming a smaller elasticity of substitution across brands and within brands. As a result, the variety-adjusted prices are 7% higher in food deserts, which is higher than the 5% we find in table 3. Using FoodAPS food deserts definition 2, the variety-adjusted prices in food deserts are 6% higher than high-income low-access census tracts (income effect) because there are 5% fewer varieties in food deserts. The EPI is 6% to 12% higher in food deserts than high-income high-access census tracts because there are 5% to 11% fewer varieties in food deserts depending on the definitions of food deserts, which are larger than the differences we find in table 3. Similar as previous tables, prices using the vehicle access definitions of food deserts are more sensitive to the income effect.

Conclusion

In this paper 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 can easily shop in their home and contiguous census tracts, unadjusted prices for common goods are not significantly different between food deserts and all types of non-food deserts. But depending on the assumptions used around the elasticity of substitution between food groups and within food groups, and definitions of food deserts, there are 2% to 7%

fewer varieties available in food deserts compared to low-income high-access census tracts. As a result, the variety-adjusted prices (the EPI) are 3% to 8% higher in food deserts than low-income high-access census tracts, capturing the effect of access. Similarly, food deserts have 2% to 5% fewer varieties than high-income low-access census tracts and thus the EPI is 3% to 6% higher in food deserts than high-income low-access census tracts, capturing the effect of living in a low-income neighborhood. In addition, the food deserts have 3% to 11% fewer varieties than high-income high-access census tracts and thus have 4% to 12% higher EPI.

These results are strongly affected by how we define the ‘local’ grocery market. When we use store sales only within the census tracts and assuming households buy food only within their census tracts, the EPI is 22% higher in food deserts than low-income high-access census tracts (access effect). In contrast, when assuming consumers can shop both within the home and contiguous census tracts, we find the EPI is only 4% to 7% higher in food deserts than low-income high-access census tracts. Similarly, the EPI in food deserts is 9% higher than low-access high-income census tracts when consumers are constrained to shop within their home census tracts (income effect) compared to 3% to 8% when consumers are allowed to shop in contiguous census tracts to. This result suggests that those households who are truly geographically constrained in their shopping are much more affected by living in a food desert. Additionally, compared with conventional prices measures such as CEPI or TFP average and median cost, EPI is more sensitive in capturing the changes in prices across census tracts.

We control for demand factors of a nutritious diet by including a number of neighborhood socio-demographic characteristics and county fixed effects. Consistently, we find that food deserts have higher variety-adjusted prices (EPI) compared to non-food deserts.

Our findings suggest that living in a food desert affects the overall food prices faced by households to a small extent when consumers can shop both within their home census tracts and contiguous census tracts. This effect 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 paper suggest that living in a food desert is unlikely to influence food insecurity to a great extent. (For more on food insecurity in the United States see Gundersen and Ziliak 2014 and Gundersen et al. 2011.) However, living in a food desert does increase the food prices a lot (by 22%) for those who cannot shop outside their immediate area. One caveat is that we exclude 239 census tracts that do not have any stores selling TFP food groups in the census tracts or the contiguous census tracts for the main analysis. For those who are constrained to shop within their immediate area, we also exclude 5,605 census tracts that do not have any stores selling TFP food groups within the census tracts. So the estimated price difference between food deserts and non-food deserts is likely a lower bound of the true price difference.

Future work is needed to ask why are prices defined by varieties of food higher in food deserts: is it a question of higher cost, barriers to entry or lower demand? Depending on these causes, inducing greater store entry may or may not have an effect on the foods purchased by households living in food deserts.

References

Alcaly, R., and A. Klevorick. 1971. "Food Prices in Relation to Income Levels in New York City." *Journal of Business* 44: 380-397.

Alexis, M., and L. Simon. 1967. "The Food Marketing Commission and Food Prices by Income Groups." *Journal of Farm Economics* 49: 436-446.

Andreyva, T., D.M. Blumenthal, M.B. Schwartz, M.W. Long, and K.D. Brownell. 2008. "Availability and Prices of Food across Stores and Neighborhoods: the Case of New Haven, Connecticut." *Health Affairs* 27(5): 1381-1388.

Block, D., and J. Kouba. 2006. "A Comparison of the Availability and Affordability of a Market Basket in Two Communities in the Chicago Area." *Pulic Health Nutrition* 9(7): 837-845.

Bodor. N., D. Rose, T.A. Farley, C. Swalm, and S.K. Scott. 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., E. Leibtag, and D.E. Weinstein. 2009. "The Role of Prices in Measuring the Poor's Living Standards." *The Journal of Economic Perspectives* 23(2): 77-97.

Broda, C., and D.E. Weinstein. 2010. "Product Creation and Destruction: Evidence and Price Implications." *American Economic Review* 100: 697-723.

Carroll-Scott, A., K. Gilstad-Hayden, L. Rosenthal, S.M. Peters, C. McCaslin, R. Joyce, and J.R. Ickovics. 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.

Chung, C., and S.L. Myers. 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.

Cotterill, R. 1986. "Market Power in the Food Industry: Evidence from Vermont." *Review of Economics and Statistics* 68 : 379-86.

Crockett, K. G., Clancy, and J. Bowering. 1992. "Comparing the Cost of a Thrifty Food Plan Market Basket in Three Areas of New York State." *Journal of Nutrition Education* 24: 715-785.

Cummins, S., and S. Macintyre. 2006. "Food Environments and Obesity—Neighbourhood or Nation?" *International Journal of Epidemiology* 35(1): 100-104.

Dube, J., and P. Manchanda. 2005. "Differences in Dynamic Brand Competition across Markets: An Empirical Analysis." *Marketing Science* 24(1): 81–95.

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.

Feenstra R.C., and D. E.Weinstein. 2010 "Globalization, Markups, and the U.S. Price Level." NBER Working Paper 15749.

Gregory, C., and A. Coleman-Jensen. 2013. "Do High Food Prices Increase Food Insecurity in the United States?" *Applied Economic Perspectives and Policy* 35: 679–707.

Gundersen, C., and J. Ziliak. 2014. "Childhood Food Insecurity in the U.S.: Trends, Causes, and Policy Options". *The Future of Children*. Fall.

Gundersen, C., B. Kreider, and J. Pepper. 2011. "The Economics of Food Insecurity in the United States." *Applied Economic Perspectives and Policy*, 33(3), 281-303.

Handbury, J., I. Rahkovsky, and M. Schnell. 2015. "What Drives Nutritional Disparities? Retail Access and Food Purchases across the Socioeconomic Spectrum." NBER Working Paper.

Handbury, J., and D.E. Weinstein. 2014. "Goods Prices and Availability in Cities." *The Review of Economic Studies* 81(4): 1-39.

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.

Krugman, P. 1991. "Increasing Returns and Economic Geography." *Journal of Political Economy* 99 (3): 483–499.

Larson, N. I., M.T. Story and M.C. Nelson. 2009. "Neighborhood Environments: Disparities in Access to Healthy Foods in the US." *American Journal of Preventive Medicine* 36(1): 74-81.

Li N. 2012. "Store Variety and Consumer Welfare in Canada-US Retail." Working paper 15, Dept. of Econ., University of Toronto.

Michimi, A., and M.C. Wimberly. 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.

Misra, K. 2008. "Understanding Retail Assortments in Competitive Markets." PhD dissertation, Northwestern University.

Montgomery, A. L., and P. E. Rossi. 1999. "Estimating Price Elasticities with Theory-Based Priors." *Journal of Marketing Research* 36(4): 413–23.

Morland, K., A.V. Diez Roux and S. Wing. 2006. "Supermarkets, Other Food Stores, and Obesity: the Atherosclerosis Risk in Communities Study." *American Journal of Preventive Medicine* 30(4): 333-339.

Morland K., S. Wing and A. D. Roux. 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.

Morris, P.M., L. Neuhauser, and C. Campbell. 1992. “Food Security in Rural America: A Study of the Availability and Costs of Food.” *Journal of Nutrition Education*: 525-585.

Sato, K. 1976. “The Ideal Log-Change Index Number.” *The Review of Economics and Statistics*: 223-228.

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 July 5, 2014).

Vartia, Y. O. 1976. “Ideal Log-Change Index Numbers.” *Scandinavian Journal of Statistics*: 121-126.

Zenk, S. N., L. L. Lachance, A.J. Schulz, G. Mentz, S. Kannan, and W. Ridella. 2009. “Neighborhood Retail Food Environment and Fruit and Vegetable Intake in a Multiethnic Urban Population.” *American Journal of Health Promotion* 23(4): 255-264.

Table 1. Summary Statistics with Contiguous Census Tracts Sales

	Food Deserts (Low Income Low Access)	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)
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)
Poverty Rate	0.27	0.27	0.07	0.08
	(0.14)	(0.12)	(0.04)	(0.05)
Tract Family Income ($)	42715.42	42531.64	102402.28	92838.55
	(14364.27)	(14772.72)	(40441.68)	(33300.74)
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)
Non-Hispanic 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

Table 2. Summary Statistics without Contiguous Census Tracts Sales

	Food Deserts (Low Income Low Access)	Low Income High Access	High Income Low Access	High Income High Access
Average TFP Cost	81.98	75.78	79.07	75.24
	(38.96)	(32.68)	(28.63)	(27.17)
Median TFP Cost	116.60	110.55	115.39	115.19
	(63.22)	(42.21)	(40.85)	(29.92)
Number of TFP Groups	23.02	23.56	23.53	24.08
	(3.22)	(3.20)	(3.27)	(3.08)
VA	2.41	2.14	2.02	1.74
	(2.57)	(3.11)	(3.07)	(2.10)
CEPI	1.06	1.06	1.07	1.07
	(0.07)	(0.07)	(0.07)	(0.07)
EPI	2.57	2.28	2.17	1.87
	(2.59)	(3.14)	(3.09)	(2.14)
Poverty Rate	0.26	0.26	0.07	0.08
	(0.13)	(0.12)	(0.05)	(0.04)
Tract Family Income ($)	41987.17	43259.43	95002.60	91049.13
	(13671.65)	(14945.68)	(34110.29)	(32236.67)
Population	4773.79	4359.4	5521.24	4495.59
	(2010.07)	(1765.71)	(2471.48)	(1776.56)
Married population share	0.38	0.37	0.56	0.49
	(0.11)	(0.10)	(0.10)	(0.10)
Median age	33.85	33.03	40.25	39.27
	(8.50)	(6.57)	(6.69)	(6.51)
Proportion of population	0.75	0.73	0.92	0.90
who complete high school	(0.14)	(0.14)	(0.06)	(0.08)
Proportion of population	0.49	0.49	0.49	0.49
who are male	(0.05)	(0.05)	(0.03)	(0.04)
Black tract	0.20	0.14	0.02	0.03
	(0.40)	(0.35)	(0.14)	(0.17)
Non-Hispanic White tract	0.43	0.35	0.90	0.80
	(0.50)	(0.48)	(0.30)	(0.40)
Hispanic tract	0.30	0.40	0.03	0.07
	(0.46)	(0.49)	(0.18)	(0.25)
Asian tract	0.00	0.02	0.01	0.03
	(0.00)	(0.13)	(0.09)	(0.17)
Observations	436	1466	914	2019

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

	2012 FoodAPS			2010 FARA		
Food Desert Definition 1	**CEPI**	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income High-access	0.002 (0.002)	0.04*** (0.01)	0.05*** (0.011)	-0.001 (0.002)	0.03*** (0.011)	0.03*** (0.012)
High-income Low-access	0.001 (0.002)	0.002 (0.012)	0.003 (0.013)	0.002 (0.003)	0.03* (0.013)	0.03* (0.014)
High-access High-income	0.01*** (0.002)	0.07*** (0.011)	0.08*** (0.012)	0.01*** (0.002)	0.06*** (0.012)	0.07*** (0.013)
Food Desert Definition 2	**CEPI**	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income High-access	0.01*** (0.002)	0.02*** (0.009)	0.03*** (0.10)	0.001 (0.002)	0.01 (0.010)	0.01 (0.010)
High-income Low-access	0.01*** (0.003)	0.03** (0.013)	0.04*** (0.014)	0.01** (0.003)	0.03** (0.014)	0.04** (0.02)
High-access High-income	0.01*** (0.002)	0.04*** (0.010)	0.05*** (0.011)	0.01*** (0.002)	0.03*** (0.010)	0.04*** (0.012)
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 low-income high-access rows denote access effect and the values in high-income low-access rows denote income effect. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live 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 EPI–Without Contiguous Census Tracts Sales

	2012 FoodAPS			2010 FARA		
Food Desert Definition 1	**CEPI**	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income High-access	0.02*** (0.004)	0.20*** (0.024)	0.22*** (0.025)	0.01* (0.004)	0.14*** (0.026)	0.15*** (0.027)
High-income Low-access	0.001 (0.004)	0.09*** (0.029)	0.09*** (0.031)	-0.002 (0.005)	0.06* (0.030)	0.06* (0.032)
High-access High-income	0.01*** (0.013)	0.23*** (0.027)	0.25*** (0.028)	0.01 (0.004)	0.17*** (0.028)	0.17*** (0.030)
Food Desert Definition 2	**CEPI**	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income High-access	0.01*** (0.003)	0.12*** (0.022)	0.13*** (0.023)	0.01** (0.003)	0.08*** (0.023)	0.09*** (0.024)
High-income Low-access	-0.004 (0.005)	0.10*** (0.030)	0.09*** (0.032)	-0.003 (0.005)	0.046 (0.033)	0.042 (0.035)
High-access High-income	0.01 (0.004)	0.15*** (0.025)	0.15*** (0.026)	0.01* (0.004)	0.10*** (0.026)	0.11*** (0.027)
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 low-income high-access rows denote access effect and the values in high-income low-access rows denote income effect. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live 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 Log Average and Log Median TFP Cost

	Log Average TFP Cost		Log Median TFP Cost	
Food Desert Definition 1	**2012 FoodAPS**	**2010 FARA**	**2012 FoodAPS**	**2010 FARA**
Low-income High-access	0.01 (0.008)	-0.007 (0.009)	-0.02*** (0.005)	-0.03*** (0.005)
High-income Low-access	-0.01 (0.01)	-0.001 (0.010)	-0.01 (0.006)	-0.01 (0.006)
High-access High-income	0.04*** (0.009)	0.03*** (0.009)	-0.01 (0.006)	-0.01* (0.006)
Food Desert Definition 2	**2012 FoodAPS**	**2010 FARA**	**2012 FoodAPS**	**2010 FARA**
Low-income High-access	0.01 (0.007)	-0.004 (0.007)	-0.002 (0.004)	-0.01** (0.005)
High-income Low-access	0.04*** (0.01)	0.02** (0.011)	0.01** (0.006)	0.01 (0.007)
High-access High-income	0.03*** (0.008)	0.02*** (0.008)	0.01 (0.005)	0.01 (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. The values in low-income high-access rows denote access effect and the values in high-income low-access rows denote income effect. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live 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– Elasticity of Substitution between Food Groups $\sigma = 2$

	2012 FoodAPS			2010 FARA		
Food Desert Definition 1	**CEPI**	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income High-access	0.002 (0.002)	0.04*** (0.010)	0.04*** (0.011)	-0.001 (0.002)	0.03*** (0.010)	0.03** (0.011)
High-income Low-access	0.001 (0.002)	0.002 (0.012)	0.003 (0.013)	0.002 (0.003)	0.02* (0.012)	0.02* (0.014)
High-access High-income	0.01*** (0.002)	0.07*** (0.010)	0.07*** (0.012)	0.01** (0.002)	0.06*** (0.011)	0.06*** (0.012)
Food Desert Definition 2	**CEPI**	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income High-access	0.01*** (0.002)	0.02** (0.008)	0.03*** (0.009)	0.001 (0.002)	0.005 (0.009)	0.01 (0.010)
High-income Low-access	0.01*** (0.003)	0.03** (0.012)	0.04*** (0.014)	0.01** (0.003)	0.03** (0.013)	0.03** (0.015)
High-access High-income	0.01*** (0.002)	0.04*** (0.010)	0.05*** (0.011)	0.01** (0.002)	0.03*** (0.010)	0.04*** (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 CEPI, VA and EPI are calculated based on store sales from the census tracts and contagious census tracts. The values in low-income high-access rows denote access effect and the values in high-income low-access rows denote income effect. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live 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 7. Regressions on EPI (A Core TFP Basket, 19 Food Groups)

	2012 FoodAPS			2010 FARA		
Food Desert Definition 1	**CEPI**	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income High-access	-0.0003 (0.002)	0.02*** (0.01)	0.02** (0.008)	-0.003 (0.002)	0.005 (0.008)	0.002 (0.009)
High-income Low-access	0.003 (0.002)	0.01 (0.008)	0.02* (0.009)	0.003 (0.002)	0.02** (0.009)	0.02** (0.01)
High-access High-income	0.003 (0.002)	0.03*** (0.008)	0.03*** (0.009)	0.001 (0.002)	0.02*** (0.008)	0.03*** (0.009)
Food Desert Definition 2	**CEPI**	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income High-access	0.01*** (0.002)	0.02*** (0.006)	0.03*** (0.007)	0.003 (0.002)	0.01 (0.006)	0.01 (0.007)
High-income Low-access	0.01*** (0.002)	0.03*** (0.009)	0.04*** (0.010)	0.01*** (0.003)	0.03*** (0.009)	0.03*** (0.011)
High-access High-income	0.01*** (0.002)	0.03*** (0.007)	0.03*** (0.008)	0.01*** (0.002)	0.02*** (0.007)	0.03*** (0.008)
Observations	9248	9248	9248	9248	9248	9248

Note: *, **, *** denotes significance levels at 0.1, 0.05 and 0.01 respectively. Standard errors are in parenthesis. The sample only includes census tracts with all core TFP food groups. The CEPI, VA and EPI are calculated based on store sales from the census tracts and contagious census tracts. All dependent variables are in logs. The values in low-income high-access rows denote access effect and the values in high-income low-access rows denote income effect. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live 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 8. Regressions on EPI– Within Brand Elasticity of Substitution (σ_g^w) =7, Across Brand Elasticity of Substitution (σ_g^a)=4

	2012 FoodAPS			2010 FARA		
Food Desert Definition 1	**CEPI**	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income	0.002	0.07***	0.08***	-0.001	0.06***	0.06***
High-access	(0.002)	(0.016)	(0.017)	(0.002)	(0.017)	(0.018)
High-income	0.001	0.004	0.005	0.002	0.04*	0.04*
Low-access	(0.002)	(0.019)	(0.020)	(0.003)	(0.020)	(0.021)
High-access	0.01***	0.11***	0.12***	0.01***	0.06***	0.11***
High-income	(0.002)	(0.018)	(0.019)	(0.002)	(0.017)	(0.020)
Food Desert Definition 2	**CEPI**	**VA**	**EPI**	**CEPI**	**VA**	**EPI**
Low-income	0.01***	0.04***	0.04***	0.001	0.01	0.01
High-access	(0.002)	(0.014)	(0.015)	(0.002)	(0.015)	(0.016)
High-income	0.01***	0.05**	0.06***	0.01**	0.05**	0.05**
Low-access	(0.003)	(0.021)	(0.022)	(0.003)	(0.022)	(0.024)
High-access	0.01***	0.06***	0.07***	0.01***	0.05***	0.06***
High-income	(0.002)	(0.016)	(0.017)	(0.002)	(0.017)	(0.018)
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 CEPI, VA and EPI are calculated based on store sales from the census tracts and contagious census tracts. The values in low-income high-access rows denote access effect and the values in high-income low-access rows denote income effect. Food deserts definition 1 is low-income urban census tracts where a significant proportion of households (33% or 500 people) live 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.

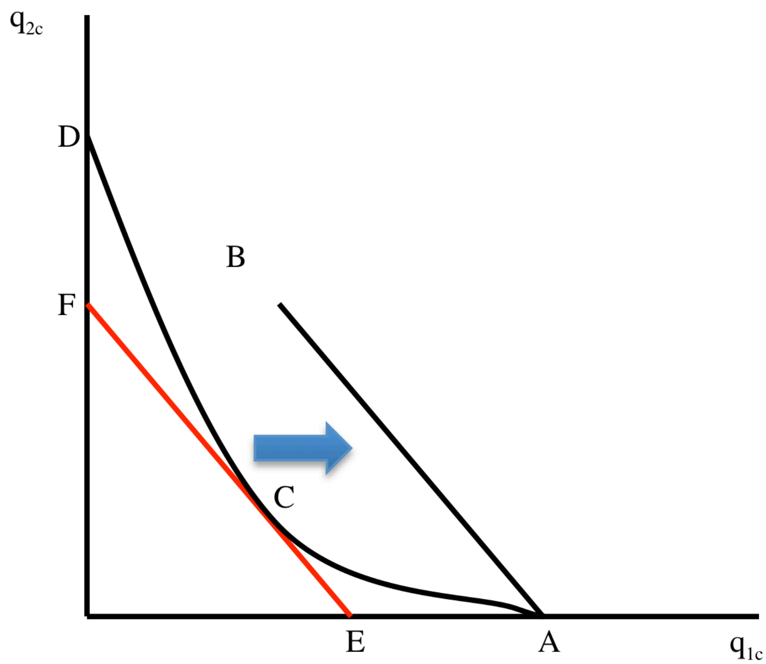

Figure 1. Grains from Variety

Food group g	$\supset$	Brand-product	$\supset$	UPC or barcode
$g \in G$		$b \in B_g$		$u \in A_{bg}$
whole milk, yogurt, and cream		Yoplait Original (brand) Yogurt (product)		Yoplait Original Yogurt strawberry flavor 6 oz
N=26		N=124,513		N=881,893

Figure 2. Categorization of Food

APPENDIX A. SATO AND VARTIA WEIGTHS AND NATIONAL EXPENDITURE SHARES

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

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

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

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}.

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