This work is distributed as a Discussion Paper by the

STANFORD INSTITUTE FOR ECONOMIC POLICY RESEARCH

SIEPR Discussion Paper No. 13-023

The Effect of Prices on Nutrition: Comparing the Impact of Product- and Nutrient-Specific Taxes

By

Matthew Harding and Michael Lovenheim

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

The Stanford Institute for Economic Policy Research at Stanford University supports research bearing on economic and public policy issues. The SIEPR Discussion Paper Series reports on research and policy analysis conducted by researchers affiliated with the Institute. Working papers in this series reflect the views of the authors and not necessarily those of the Stanford Institute for Economic Policy Research or Stanford University

The Effect of Prices on Nutrition: Comparing the Impact of Product- and Nutrient-Specific Taxes*

Matthew Harding† Michael Lovenheim ‡

December 24, 2013

Abstract

This paper provides an analysis of the role of prices in determining food purchases and nutrition using very detailed transaction-level observations for a large, nationally-representative sample of US consumers over the period 2002-2007. Using product-specific nutritional information, we develop a new method of partitioning the product space into relevant nutritional clusters that define a set of nutritionally-bundled goods, which parsimoniously characterize consumer choice sets. We then estimate a large utility-derived demand system over this joint product-nutrient space that allows us to calculate price and expenditure elasticities. Using our structural demand estimates, we simulate the role of product taxes on soda, sugar-sweetened beverages, packaged meals, and snacks, and nutrient taxes on fat, salt, and sugar. We find that a 20% nutrient tax has a significantly larger impact on nutrition than an equivalent product tax, due to the fact that these are broader-based taxes. However, the costs of these taxes in terms of consumer utility are not higher. A sugar tax in particular is a powerful tool to induce healthier nutritive bundles among consumers.

KEYWORDS: food and nutrient demand, SSBs, food taxes, purchasing environments, QAIDS
JEL CLASSIFICATION: I12, I19, C33

*Research support from the Robert Wood Johnson Foundation, the UPS Foundation, and the U.S. Department of Agriculture is gratefully acknowledged. Binying Liu provided excellent research assistance. We thank Jerry Hausman, Ephraim Leibtag, Arthur Lewbel, Aviv Nevo, and conference participants at the AEA Annual Meetings, the 2013 RWJ Healthy Eating Conference, and the 9th Conference on Industrial Organization and the Food Processing Industry, Toulouse, June 21-22, 2012 for excellent comments.

†Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305; Phone: (650) 723-4116; Fax: (650) 725-5702; Email: mch@stanford.edu

‡Department of Policy Analysis and Management, College of Human Ecology, Cornell University, 135 Martha Van Rensselaer Hall, Ithaca, NY 14853 and NBER; Phone: (607)255-0705; Fax: (607)255-4071; Email: mfl55@cornell.edu

1 Introduction

Obesity is one of the largest prevailing public health concerns in the United States. National obesity rates (defined as having a body mass index (BMI) greater than 30 kg/m^2) have increased dramatically among both children and adults over the past several decades. In the late 1970s, the national obesity rate among adults was about 14.5%; it increased to 22.5% by the early 1990s (Flegal et al., 1998). Today, over 2/3 of Americans are overweight and over 36% are obese.[1] Estimates also suggest that about 30% of children are obese or overweight, although the trend appears to have flattened somewhat in recent years (Ogden et al., 2010; Cawley, 2010). The increases in obesity have been more pronounced among those with lower income, especially for women, as well as among non-Asian minorities.[2] Obesity has been linked to a higher prevalence of chronic diseases, such as arthritis, diabetes and cardiovascular disease (Must et al., 1999; Pi-Suyner, 1991). The associated cost to the U.S. medical system has been estimated at about $147 billion in 2006, with Medicare and Medicaid financing approximately half such costs (Finkelstein, Fiebelkorn and Wang, 1998). The high and rising obesity rate combined with the financial burden it places on the health care system have made it one of the leading public health problems in the United States.

The rise in obesity can be ascribed mainly to the fact that physical exercise has not increased over this time period but calories consumed have grown dramatically (Cutler, Glaeser and Shapiro, 2003). The implication is that Americans are consuming an increasingly unhealthy diet, both in terms of the total number as well as the composition of calories. With over 2/3 of household caloric intake occurring at home (Nestle, 2012), a question of central importance is why households are purchasing large quantities of more unhealthy food and what public policies could be used to support purchasing a more healthy nutritional bundle. The core type of policies that have been discussed in the economics and public health communities are price-based interventions, such as a sugar-sweetened beverage (SSB) tax or a fat tax. To date, however, there is little robust evidence on how price policies will affect the products consumers purchase and the resulting nutritional bundles of households.

The lack of understanding of how price-based policies will affect nutrition is surprising given the pervasive policy discussions throughout the world on how to use the tax system to support healthier eating. However, there are several data limitations and methodological complications

[1] The Economist. Special Report: Obesity. December 15, 2012.

[2] Non-Hispanic blacks have the highest rates of obesity (49.5%) compared to non-Hispanic whites (34.3%): `http://www.cdc.gov/obesity/data/adult.html`

prior work has faced that has limited our understanding of price-based nutritional policies. In this paper, we combine structural demand techniques with a Big Data approach to this problem that relies on linking several large, proprietary databases together to provide a comprehensive view of consumer purchases, prices, nutritional characteristics of foods purchased, and purchasing environments. Our analysis focuses on identifying the role of prices in driving the type and nutritional content of foods households purchase for home consumption, which sheds needed light on how consumption bundles and nutrition are shaped by prices consumers face in order to inform the development of appropriate price-based policies to support nutritional goals policymakers may have. The size and scope of our data allow us to partition the product space into "nutritional clusters" that represent distinct groupings of product types within each product group (e.g., diet and regular soda, less- and more-healthy packed meals). We then estimate a structural Quadratic Almost Ideal Demand System (QAIDS) model over the joint product-nutrient groups at the household-month level, accounting for the endogeneity of prices with instruments based on marginal cost variation and the endogeneity of expenditures using instruments based on local business cycle variation. Using the resulting price and expenditure elasticities, we simulate the effect of various product (e.g., soda) and nutrient (e.g., sugar) taxes on the distribution of consumer expenditures across groups and on the distribution of nutrients. We also calculate the consumer welfare costs of each tax.

The unique and detailed data we employ combined with our methodological approach allows us to advance the literature on consumer demand and expand our knowledge of the effects of price-based interventions on nutrition along several dimensions. First, a major limitation that prior research has faced is the lack of comprehensive data linking consumer purchases to prices and nutrition. The dataset we assemble for this project is unique in its size and scope: no prior work has used data that has the volume of information we possess as well as the long timeframe over which we have it at the household level. Our purchasing data are from the Nielsen Homescan Panel, which is comprised of a demographically-representative set of consumers that provides, for each purchase transaction of a grocery item, the universal product code (UPC), the price paid, and the date of purchase. Overall, we observe over 123 million transactions made by consumers between 2002-2007. In addition, the data contain rich demographic information for each household, including household composition, race, income, and the census tract in which they live. Using the UPCs, we match each product with information from another data set that contains all nutritional information included on the product nutritional label. Finally, using the census tract identifiers, we construct precise measures of the food purchasing

environment each household faces with data from a yearly census of US supermarkets, grocery stores, and convenience stores. Together, these data provide extensive information on the nutritive bundles households purchase, the prices paid for all goods, and the local purchasing environments surrounding each household.

This data set is well-suited to examine the role of prices in household purchasing behavior and nutrition and provides a significant advantage over earlier attempts at quantifying food demand. Existing researchers using scanner data have typically had access to data that is more limited in time or in the breadth of information used. For example, Dubois, Griffith and Nevo (forthcoming) use a similar scanner dataset to construct international comparisons for the year 2005-06 only. They document and examine the reasons for nutritional differences across the US, UK and France and find that price differences and product composition differences play some role in driving cross-country food purchasing variation. These results suggest that price-based manipulations could have important impacts on household nutrient intake, however they do not examine how product and nutrient taxes would affect the distribution of consumer expenditures and nutrition. Finkelstein et al. (2012) and Finkelstein et al. (2010) use similar data as well but only for 2006, and their analysis data do not include a complete set of purchased products.[3] In their examination of purchasing patterns over the lifecycle, Aguiar and Hurst (2007) use data only for Denver, covering the period January 1993 through March 1995, and do not have nutritional information about the products purchased.[4] The dataset we construct thus provides the opportunity to measure a more detailed set of information over a longer period of time than any existing demand study in the literature.

The second contribution we make is to partition the data into a joint product-nutrient space over which the demand model can be estimated. A core difficulty in estimating food demand models is that it is not possible to price the individual nutritional components of a good. That is, one cannot determine what part of the price of a good is due to the fat content and what part is due to the salt content. Prior research typically has addressed this issue

[3]Other researchers who have used scanner data to estimate demand models have focused on a specific category of products, such as beverages (Dharmasena and Capps, 2012; Lin et al., 2011; Zhen et al., 2011), snack foods (Kuchler, Tegene and Harris, 2005), and dairy (Chouinard et al., 1998; Griffith, Nesheim and O'Connell, 2010). These analyses typically contain only one or a few years of data, do not examine substitution across groups, and/or aggregate the data to the product-month level or to the household-year level, which throws out a large amount of important demographic and price information. Aggregation also renders the estimates subject to biases that might arise from local composition changes that are correlated with price changes. Another strand of the demand literature uses household survey data combined with local price estimates from the American Chamber of Commerce Research Association (ACCRA) or from respondent reports (Yen et al., 2004; Duffey et al., 2010). The prices used in these analysis are likely to be far noisier than in our data, and respondent reports of foods purchased or consumed also are measured with more error. Finally, although the specific methods used in all of these studies differ, there is little attention paid to endogeneity issues with respect to prices. The use of endogenous price variation could introduce significant biases in the resulting demand estimates.

[4]In a related study examining lifecycle expenditures, Aguiar and Hurst (forthcoming) use Consumer Expenditure Survey data that provide information about a wider variety of goods consumed but that do not have the specificity about food products purchased that characterize the Nielsen Homescan data.

either by separating different products into different nutrition groupings in an ad-hoc manner or by estimating models with the nutrients purchased from each product type as the outcome of interest in the demand model (Finkelstein et al., 2012).[5] Neither method is completely satisfying, however, because the former relies on researcher judgements about how consumers segment products and the latter does not account for the fact that nutrients are correlated with each other within products, thus introducing a bias from omitted variables. We develop a new method for partitioning the product space into relevant nutritional groupings using cluster analysis. The motivation of this method is that nutrients tend to be correlated with each other in specific ways that define a set of relevant goods for a consumer. For example, high-fat snack foods also tend to be high-salt and high-sugar, while healthier snack foods have low amounts of all of these nutrients.

This process amounts to partitioning each product type into clusters of nutrient groupings that define a nutritionally-similar set of goods using the data, rather than making judgement calls about what the relevant different food groupings are. It also is possible for there to be several nutritional clusters for each product type. Clustering in this manner thus allows us to generate a set of nutritionally-bundled goods within each product group that define consumers choice sets in a parsimonious manner and that do not require us to assign a price to each nutrient. Specifically, we categorize all purchases into 14 product categories, which are then further partitioned into 33 different product-nutrient groups. We then estimate QAIDS models at the household-month-group level, the result of which is a matrix of own- and cross-price elasticities as well as expenditure effects for the different groups. These estimates subsequently are used to simulate the effect of product- and nutrient-specific taxes on purchasing behavior and nutritional bundles. This method is straightforward but data-intensive, which highlights the value of our Big Data approach.[6]

Our analysis also contributes to a relatively large literature that focuses on estimating the price elasticity of demand for particular foods.[7] These studies can be split into two groups. The first uses structural demand techniques to identify price elasticities of demand. As discussed

[5]Alternatively, Dubois, Griffith and Nevo (forthcoming) model the joint product-nutrient space in such a way that the characteristics of products are fully subsumed into the nutrient space. While useful for the aggregate cross-country comparisons they employ, this method imposes restriction on the cross-price elasticities and does not allow one to examine product-specific price manipulations directly. Thus, it is less useful for our research questions.

[6]Another alternative would be to estimate a choice model over UPCs in the spirit of Berry, Levinsohn and Pakes (1995). Given that we have 568,637 total UPCs spanning 123 million transactions, such a model is currently not tractable to estimate. The QAIDS model estimated over the joint product-nutrient space has the benefit of being computationally feasible while still allowing for a large degree of flexibility in consumer substitution across relevant food types.

[7]See Andreyeva, Long and Brownell (2010) for a review and meta-analysis of this literature. Most of the these studies use aggregate time-series data, with a smaller percentage using household-level data. A small minority use household scanner data, and these studies typically lack credible instruments for prices and do not allow for consumer substitution across products.

above, data limitations and issues pertaining to price endogeneity have hampered these analyses. The second group uses variation from imposition of excise or sales taxes as natural experiments to examine how consumers respond to arguably exogenous price changes. While most states exempt food purchased in grocery stores from sales taxes, as of 2011, 35 states either include soda in the sales tax base and/or levy an excise tax on soda. The existing literature using these tax changes as natural experiments to identify the effect on soda demand find that soda is inelastic with respect to this tax variation (Powell et al., 2013; Fletcher, Frisvold and Tefft, 2010*a*,*b*).[8] However, these taxes tend to be quite small and they are unlikely to be salient to consumers because they usually are added onto the price at checkout (Chetty, Looney and Kroft, 2009). In general, there is a lack of tax variation with which to identify the effect of taxing most products or nutrients. The lack of such policy variation highlights the need to examine how larger and more salient price changes might affect consumption using structural demand models that incorporate methods to identify causal price parameters. We argue our estimates are the most detailed and accurate set of price and expenditure elasticities for the US produced to date, and thus they are of interest in their own right. Furthermore, our methodology allows us to estimate the welfare costs of imposing different types of taxes, which typically is not possible using reduced form methods surrounding tax changes.

Finally, our paper contributes to the literature on commodity taxation and nutrition by examining taxes on different products and on nutrients themselves that incorporate both the ability of consumers to substitute across different goods in response to price changes but also the ability of consumers to reduce overall food consumption when prices rise. The inability to link purchases and prices together with nutritional content of food has precluded an examination of the effects of taxing specific nutrients, such as a sugar tax, in prior work, and existing studies are limited to an examination of the effects of taxing a specific product, like soda, or a group of products that have "undesirable" nutritional characteristics, like sugar-sweetened beverages. However, in order to inform the development of food-based taxes to affect nutrition, it is important to generate evidence on how consumer behavior will respond to taxing different types of products, product groupings and nutrients more broadly. For example, what will be the effect of taxing soda versus other highly-processed goods that previous work has argued contribute to rising obesity (Cutler, Glaeser and Shapiro, 2003)? How does taxing products affect consumer behavior differently than taxing nutrients? Does taxing sugar rather than taxing soda have a

[8]Another 13 states tax snack foods, such as chips and pretzels, but we are unaware of any studies that examine the effect of these taxes. Some countries, such as Hungary, also tax foods high in sugar, fat, or salt. A recent meta-analysis of the international evidence suggests that these taxes have generally affected consumption in the predicted direction, but that in order to have a significant population health impact taxes in excess of 20% are required (Mytton, Clarke and Rayner, 2012).

larger affect on nutrition, and how do fat taxes compare with sugar taxes in terms of changing consumer behavior and in terms of welfare costs of these taxes? We examine these questions directly by separately simulating the effects of a 20% tax on soda, sugar sweetened beverages (SSBs), packaged meals, snacks/candy, fat, sugar and salt on budget shares for our 33 product groups and on nutrient bundles purchased by households. By incorporating substitution effects across goods as well as the income effects due to these taxes, these simulations yield a more comprehensive picture of how these product- and nutrient-specific taxes affect nutrition than has been possible in prior research.

Our main finding from the tax simulations is that nutrient-specific taxes have much larger effects on nutrition than do product-specific taxes, without causing a larger decline in consumer utility. The intuition for this result is that nutrient-based taxes have a much broader base, so it is more difficult to substitute away from any one good in response to such taxes. For example, a 20% tax on soda decreases total purchased calories by 4.84% and decreases sugar consumption by over 10%. However, a 20% sugar tax decreases total calories by over 18% and sugar by over 16%. The larger effect of a sugar tax on nutrition comes despite the fact that it has the same effect on indirect utility as a soda tax. Due to their negative income elasticities and the patterns of own- and cross-price elasticities we find, taxes on snacks and packaged meals have very small effects on nutrition. Fat and salt taxes, on the other hand, have much larger effects, decreasing calories by 19% and 10%, respectively. SSB taxes, which can be thought of as a hybrid price policy that targets a set of products based on their nutritional content, also are quite effective, reducing caloric intake by over 9%. However, these taxes are less-effective and only slightly less-distortive than a broad-based sugar tax. Overall, our results point to large potential impacts of nutrient-based taxes on nutrition that are driven by their broad base. Among these, sugar taxes stand out as being particularly effective at supporting healthier purchasing behavior.

This paper is structured as follows. Section 2 introduces the different proprietary databases and describes how the analysis dataset was constructed. We then explain how the data are used to estimate the QAIDS model. Section 3 discusses the estimated parameters of the demand system as well as the price and expenditure elasticities. Section 4 presents counterfactual simulations of the impact of product taxes for soda, SSBs, packaged meals, and snacks/candy and also nutrient taxes on fat, sugar, and salt on purchasing behavior and nutrition, and Section 5 concludes.

2 Data and Empirical Model

2.1 Data

The data we use for this analysis is constructed from several different databases that, together, provide the most detailed information on food purchases over a multi-year period by American households to date. The core of the data set consists of records on over 123 million food purchase transactions made at grocery stores. These data come from the Nielsen Homescan database.[9] The Nielsen Homescan data are collected from a demographically-representative panel of US households who are provided with a small scanner similar to ones used at grocery stores. We observe food purchases intended for at-home consumption made at a variety of stores, from traditional grocery stores, supermarkets and convenience stores to supercenters and warehouse clubs. Most purchases are identified using a unique Universal Product Code (UPC), a 12 digit barcode scanned by the retailer at the point of purchase. Households are required to scan the UPC barcode of each item purchased and record the price, date, and location of purchase as well as the quantity of each item bought. Non-UPC label items, such as fresh fruits and vegetables, are recorded manually by the households and also are included in our data and analysis. UPC-labeled items are automatically matched by Nielsen to an additional database that contains detailed product characteristics, such as brand, size, and price of the product. Prices are checked against store data in order to account for discounts and coupons.[10] At the end of the week, households transmit their purchase data electronically to Nielsen. Participating households are not paid in cash but earn a set number of points with each data upload. Households are rewarded the longer they stay with the program, and those who already have been in the panel for at least six months receive a substantially larger number of points for each transmission. The points can be redeemed for products, so households have a strong incentive to upload their purchase data. The reward program is similar to point programs offered by major credit cards, where customers can redeem points for select merchandise from a product catalog, such as tools, electronics, toys or to enter sweepstakes. To our knowledge the rewards are not food related or likely to induce a significant income effect in the population. All households are in the sample for at least 10 months, and most stay in for under 5 years.

Participating households are recruited across 52 large markets (roughly corresponding to the largest Metropolitan Statistical Areas or MSAs in the US). As a result of the engaging system, which generously rewards participation, the panel is oversubscribed with an up to two

[9] For additional information on the Nielsen Homescan data, see Todd et al. (2010).

[10] Einav, Leibtag and Nevo (2010) discuss some of the technical issues that arise as a result of the price recording process.

year waiting list. Nielsen aims to maintain a nationally-representative panel, subject to additional weights. Thus, recruitment into the panel is ongoing, with observably similar refresher households replacing those who leave the sample.

The Nielsen Homescan data do not contain any nutritional information about the products purchased. From Gladson and FoodEssentials, we have obtained nutritional information at the UPC level for all products tracked by these companies. These companies monitor and track the various products sold in stores and record the nutritional panel and packaging characteristics for products from a wide range of manufacturers. The data contain all nutritional information on the product nutrition facts panel by UPC code, including the amount of each nutrient and the entire ingredient list. While the nutrition data cover a large proportion of the total purchases made by households in the panel, about 35% of the transactions in our data cannot be matched directly to products in the nutrition database. The products that cannot be matched directly typically consist of small, often generic, brands with a limited regional distribution network. In order to ascribe nutritional characteristics to these products, we developed a matching algorithm that assigns to each product the average per-serving characteristics within decreasingly narrow bins of similar products. We first match unmatched Nielsen products to products with nutritional information in the same "product module" with the same size, type, brand, product, flavor, variety and formula.[11] This produces matches for an additional 15% of the transactions. We then alter the product characteristics used for matching to systematically broaden the set of potential match products. Details of the matching strategy and the proportion of the sample in each year matched using each set of characteristics are show in Table A-2. This algorithm produces nutritional estimates for between 96 and 98 percent of the products recorded in the Nielsen Homescan data, depending on the year.[12] Furthermore, about 5% of the transactions are "random weight" or bulk foods, such as fruits, vegetables, and fresh meats, that are not tracked by Gladson or FoodEssentials. The nutritional content of these items were coded manually by name and integrated with the database using information from the USDA National Nutrient Database.

Although these data provide extremely detailed information about grocery purchases, there are some limitations of using these data to study nutritional outcomes. First, they do not include restaurant purchases (including fast food) or purchases made at vending machines, and

[11] Product modules are specific product designations that indicate small groups of similar products. For example, diet and regular soda are two product modules, as are canned apples and pears. There are numerous product modules in the data, and thus products within product modules are very similar to each other in terms of nutritional content.

[12] This algorithm is essentially the same as the one employed by Dubois, Griffith and Nevo (forthcoming) and resulted from a joint collaboration with the USDA.

any food consumed outside the home that is not purchased by the household at a store will not be included. Second, because we are using purchase data, we cannot observe consumption of the overall household or of different household members. Spoilage, wastage, and different eating patterns among different household members complicate extrapolations from purchasing behavior to individual consumption. Despite these limitations, we believe these data are appropriate to estimate a rich food demand model and study the effects of food taxes on nutrition. To the extent that these factors influence purchasing behavior and nutritive bundles, it is unlikely that changes in the food available within the household does not result in changes in nutrients consumed by household members. This is particularly true because over 2/3 of calories are consumed in the home (Nestle, 2012).

The data allow us to control for a rich set of observables that capture detailed information on the household and the purchasing environment in which food purchase decisions are made. Nielsen administers a yearly demographic survey to all participating households that records their home address, household income, the education and race of the head of household, the age and labor supply for both male and female heads of households, the household composition and number of children. Our data do not include addresses, but we are given the census tract in which each household resides in each year. The summary statistics for the demographic variables are presented in Table A-1. Consistent with previous work using these data (e.g., Burda, Harding and Hausman (2012)), we find that certain sampling distortions remain relative to national averages. For example, in the data 31.8% of the male heads and 36.8% of the female heads are recorded as not employed. Similarly, we find that 28.5% of the sample reports a household income above $70,000 compared to only 24.7% who report a household income of below $30,000. These sample means reflect the relative attractiveness of participation in the panel to households of different demographic compositions, but, especially since we control for these demographic characteristics in our empirical models, it is unlikely any differences between the Nielsen sample and a national sample will bias our estimates.

While the Nielsen database has some information about the stores at which purchases are made, it does not contain information on the set of stores available to consumers. In order to further characterize purchasing environments, we supplement our data with records from the National Establishment Time-Series (NETS) Database, compiled by Walls and Associates using data from the Dun and Bradstreet (D&B) archival national establishment data (DUNS Marketing Information file, DMI). This database tracks over 40 million establishments nationwide, with information on their industries and location. We obtained data on the location of all stores

at which one can buy groceries (i.e., all non-prepared foods) in each county in which a Nielsen respondent lives in the 52 markets. For each household, we construct the detailed distribution of stores by geographic distance. For simplicity, we categorize stores as: supermarkets, grocery stores and convenience stores and count the number of stores within 1/4 mile, 1/2 mile, 3/4 mile, 1 mile, 1 1/2 miles and 2 miles of the centroid of a household's residence census tract.[13]

2.2 Product Classification

One of the challenges involved in building a comprehensive food demand model consists in choosing the level of product aggregation that leads to a realistic, yet computationally feasible structural demand system. A central contribution of this paper is the use a joint product-nutrient aggregation, which models households as making joint decisions over products and nutrients. The need for such a process of aggregation is easy to see when considering the impact of a sugar tax on soda. When modeling the impact of such a tax, it is not sufficient to aggregate all soda products into one category, because people may substitute across different soda types when sugared soda become more expensive. Changes to the price of sugar-sweetened soda also could cause households to substitute towards other goods with more or less sugar in them, depending on whether sugar soda is a complement or substitute with other high-sugar goods. Furthermore, taxing sugar-sweetened soda could lead to a substitution towards other unhealthy nutrients, such as fat. A realistic model that captures the manner in which consumers choose over types of goods requires a joint product and nutrient demand system in order to accurately model households substitution patterns. At the same time, expanding the choice set too much leads to a computationally infeasible demand system, which highlights the need to balance computational feasibility with allowing a sufficient choice set for consumers.

Our approach to this problem is to classify all products into 33 mutually exclusive product-nutrient groups derived as follows. First, we categorize all UPCs into 14 product categories, roughly corresponding to the major food areas a customer is likely to encounter in the store. These are fruits/vegetables, cold beverages, warm beverages, soda, packaged meals, grains/pasta/bread, canned food/sauce, cereal/breakfast items, other dairy (such as yogurt and cheese), meat, condiments, snacks/candy, milk and baking goods. Second, we analyze the nutritional content of each UPC in each of the 14 categories in terms of the most important and most salient nutritional items recorded on the nutritional panel of each product: calories, calories from fat, total fat, cholesterol, salt, carbohydrates, sugar and protein. Using these nutrients, we perform a

[13] All stores are counted in these measures, not just stores within the household's county.

k-median clustering at the UPC level separately for each of the 14 product categories. This clustering separates the 14 initial product groups into 33 distinct product-nutritional clusters, where the separation is a direct function of the nutrients recorded for those products.

K-median clustering is a variation of the more frequently encountered k-means clustering algorithm. The k-median procedure partitions the UPCs within each of the 14 categories into k additional distinct groups, where each group is characterized by a centroid corresponding to the vector of median values of each nutrient for the UPCs in a given partition. The distance between a product and the centroid of a group is evaluated in terms of the corresponding L_1 distance between the nutrient values. We chose the k-median over the k-means procedure due to its robustness to outliers in the nutrition distribution.

We provide a partial solution to the problem of choosing the optimal number k of clusters corresponding to each of the major product categories by employing the following optimization. For each of the 14 product categories, we sequentially partition all UPCs using the k-median algorithm into $k = 1 \dots 10$ partitions. For each partition, we compute the silhouette value corresponding to each UPC. The silhouette value is a metric that captures the distance between a given UPC and the UPCs in the other partitions. Thus, a UPC with a silhouette value of $+1$ indicates that a product is classified as being part of a group that is well-separated from all other groups in the nutrition space. UPCs with silhouette values close to 0 are from a nutrition composition point of view not distinguishable from UPCs in another cluster. For each configuration, we compute the average silhouette value. Partitions with larger average silhouette values are preferred, and we choose the number of partitions that maximize the average silhouette value for each of the 14 product groups.

This clustering procedure provides a local solution to the optimal number of product clusters. Due to the computational complexities of the estimating a large non-linear demand system, we only evaluated a maximum of 10 partitions for each product group. However, the algorithm partitioned each of the 14 product categories into 2 or 3 additional groups, which suggests using more partitions would not have altered our results. The result is a total of 33 product groups that are distinct both in terms of product type and nutritional content. Figure 1 plots the silhouette values by UPCs for soda products computed for $k = 2$ partitions, which provides an illustration of how this method works. It shows that the two partitions are well separated in the nutrient space. Only a very small number of UPCs have (small) negative silhouette values, indicating that their nutritional composition is shared by both groups and thus cannot

be unambiguously allocated to either group.[14] As Figure 1 illustrates, $k = 2$ is the optimal partition for the soda group when considering partitions of 10 or fewer groups. In this case, one of the groups is clearly distinguished in the nutrient space by having 0g of sugar and 0 calories per serving, compared to the other group where the median UPC has 960 calories and 256g of sugar per serving. Thus, the k-medians method ostensibly breaks soda into diet and non-diet "clusters." While soda is a fairly simple category to divide up along nutritional lines, the success of this method in doing so suggests it also is reasonably dividing other types of product groups that are more difficult to separate, such as packaged meals and cereal.

Figure 2 maps the average household budget in terms of the 14 major expenditure categories. Additional details on budget shares and prices for each of the 33 groups is available in Table 1. They demonstrate the comparatively large share of snacks and candy (15.6%), meat (12.8%), soda (5.3%), and packaged meals (7.5%), relative to fruits and vegetables (9.1%). International comparisons reveal these to be substantially larger than the shares consumed in other developed countries, such as France or the UK, where fruits and vegetables account for close to 20% of expenditure (Dubois, Griffith and Nevo, forthcoming). In Figure 2, darker rectangles correspond to more expensive product categories as measured by the price per ounce. Products with a higher share of the household food budget tend to also be more expensive. Meat/protein is twice as expensive as fruits and vegetables per ounce, while while snacks and candy are over 2 1/2 times as expensive.

Table 2 lists the median nutrient quantity per serving for each of the 33 joint product-nutrient cluster groups in the data. These values correspond to the group centroids in the nutrition space. Packaged meals stand out relative to all other food categories as having the most calories, calories from fat, total fat, salt, and protein, the second largest amount of cholesterol after meat, and the third largest amount of carbohydrates after cereal, pasta and bread. Compared to a serving of fruits and vegetables, a serving from a packaged meal has over 3 times the number of calories and 7 times the amount of fat and salt. Soda, which has almost no nutritional value, dominates all other food items in terms of sugar per serving, with 3 times the amount found in a serving of fruits and vegetables and twice the amount found in milk products. The high sugar content of soda and of other beverages explains the intense policy focus on sugar-sweetened beverage prices. However, several other categories, such as cereal and canned foods, also have substantial sugar content that suggests a SSB tax would be imperfectly

[14] Note that the clustering algorithm still assigns these products to the "closest" group, but this small set of products has characteristics that are similar to those in both groups.

targeted at high-sugar goods. Snacks and candy, which correspond to the largest share of a household's at-home food budget, also are characterized by a large number of calories and a significant amount of salt and sugar in spite of the limited nutritional value.

Table 2 also shows that within the 14 aggregate product categories there are substantial differences across nutritional groups. This is a direct implication of the k-median clustering, which partitions the product space into nutritionally-distinct clusters. Figure 3 present a bar chart of the (median) calories per serving for each of the product groups. Each aggregate category is also separately labeled and colored. While we do not label the individual product groups for each aggregate category, we can think of most categories as being divided into low, medium, and high calorie products, while some categories only have a low and a high calorie group.

The mapping of these product groups across the multivariate nutritional space is rather more complicated however, reflecting the non-linear relationships between the different nutrients when combined in the different product formulations. The soda category, for example, can easily be described as consisting of two groups of products: zero calorie / zero sugar soda ("diet-soda") and high calorie / high sugar soda ("classic" soda). The packed meals and snacks/candy categories exhibit similar clustering, with there being a "healthy" and "unhealthy" grouping of each. These results demonstrate that for these product categories, nutrients are highly correlated with each other – high-sugar snack foods tend also to be high in fat. When one taxes fat (sugar), it thus should affect sugar (fat) consumption from these groups as well. In contrast, some groups exhibit clustering around certain nutrients. For example, the baking goods category consists of a group of high-calorie but low-sugar products and another group of products containing substantially fewer calories per serving but also containing a non-negligible amount of sugar. Similarly for fruits and vegetables, there is a high-calorie and low-salt group, a low-salt and high-calorie group, and a low-calorie, low-salt group. Taxing salt therefore could lead to more calories being consumed from baking goods and fruits and vegetables. These trade-offs are likely familiar to most households who are routinely facing the problem that minimizing the amount of one nutrient (e.g. fat) requires them to consume much larger amounts of another nutrient (e.g. salt or sugar). Ultimately, such trade-offs represent the chemical reality of product formulation, where reducing one ingredient requires the manufacturer to increase the amount of another in order to preserve the attractiveness of the product to the consumer. Table 2 shows that this trade-off is important for some products but not for others, which is an interesting finding. To the extent that prior work has attempted to split up product groups

to allow for within-group substitution, they have done so in an ad-hoc manner that relies on judgements made by individual researchers. The k-medians results show that nutrient clusters vary considerably across product types, making such determinations very difficult. The ability to model the joint product-nutrient space in a parsimonious manner that also reflects the actual choices households face without having to rely on judgement calls about product classifications we believe represents an important contribution to food demand modeling.

2.3 Empirical Methodology: QAIDS Model

A large number of different empirical strategies have been employed in the literature to estimate food demand models. Since the aim of this paper is to evaluate counterfactual policy changes for which few if any natural experiments exist, it is important to adopt a utility-derived structural approach. A structural approach will allow us to use estimated price and income effects to simulate the effect of product and nutrient taxes. In addition, this approach will permit us to examine how variation in purchasing environments affect purchasing behavior.

A nonparametric analysis of the relationship between log expenditure and income, which is available upon request from the authors, indicates that all product groups have either linear or quadratic Engel curves. This result suggests a rank 3 demand model. Restricting attention to the class of exactly aggregable demand systems, the maximum rank of such demand systems is 3 (Deaton, 1981). Thus, following Banks, Blundell and Lewbel (1997), the quadratic nature of the Engel curves argues for a demand model that is linear in log real expenditure and log real expenditure squared. In particular, this class of models has the attractive feature that it allows households at different points of the income distribution to have different marginal consumption responses to the same tax.

We focus on rank 3 demand systems with indirect utility:

$$\ln V(\mathbf{p}, m) = \left\{ \left[\frac{\ln m - \ln a(\mathbf{p})}{b(\mathbf{p})} \right]^{-1} + \lambda(\mathbf{p}) \right\}^{-1}, \tag{1}$$

where $\mathbf{p}$ denotes the vector of product group prices (in dollars per ounce) for each of the $i = 1 \ldots K$ product categories (with $K = 33$), and m denotes the total expenditure on grocery purchases for at-home consumption for the household in the given time frame. Moreover, let

w_i denote the budget share of product group i. Then,

$$\ln a(\mathbf{p}) = \alpha_0 + \sum_{i=1}^{K} \alpha_i \ln p_i + \sum_{i=1}^{K} \sum_{j=1}^{K} \gamma_{ij} \ln p_i \ln p_j, \tag{2}$$

is a price index (in translog form),

$$b(\mathbf{p}) = \prod_{i=1}^{K} p_i^{\beta_i}, \tag{3}$$

is the Cobb-Douglas price aggregator, and

$$\lambda(\mathbf{p}) = \sum_{i=1}^{K} \lambda_i \ln p_i, \tag{4}$$

is a function of prices $\mathbf{p}$, with $\sum_{i=1}^{K} \lambda_i = 0$.

The system of budget share equations for the product groups is given by:

$$w_i = \alpha_i + \sum_{j=1}^{K} \gamma_{ij} \ln p_j + \beta_i \ln \frac{m}{a(\mathbf{p})} + \frac{\lambda_i}{b(\mathbf{p})} \left(\ln \frac{m}{a(\mathbf{p})} \right)^2. \tag{5}$$

In order to estimate this model, we use data on over 123 million transactions for 162,974 household years. As a result of the expansion of the consumer panel over the years, the number of participating households increases from 40,000 to over 60,000 over the sampling period. We aggregate all of the data to the household-month-product-nutrition cluster group level. This aggregation leads to 34,245,832 observations. Aggregating the data in this manner has several advantages. First, using monthly budget shares minimizes problems associated with storage that plague studies using high-frequency shopping data. Monthly data at the product-nutrition cluster group level also decreases the prevalence of zeros, whereby households have no purchases of the given group in the month. Finally, using monthly data rather than longer horizons, such as quarterly data, makes it more likely that the average prices used reflect the prices faced by the household in that time period.

Letting $w_{i,h,t}$ denote the budget share of group i for household h in month t, the model above gives rise to the following structural system of estimating equations:

$$w_{i,h,t} = \alpha_{i,h,t} + \sum_{j=1}^{K} \gamma_{ij} \ln p_{j,h,t} + \beta_i \ln \frac{m_{h,t}}{a(\mathbf{p_{h,t}})} + \frac{\lambda_i}{b(\mathbf{p_{h,t}})} \left(\ln \frac{m_{h,t}}{a(\mathbf{p_{h,t}})} \right)^2 + u_{i,h,t}, \tag{6}$$

where $u_{i,h,t}$ is an iid error term and $\mathbf{p_{h,t}}$ denotes the vector of prices faced by household h in

month t.

For computational simplicity, we approximate the term $a(\mathbf{p_{h,t}})$ by the Stone price index (Decoster and Vermeulen, 1998):

$$a(\mathbf{p_{h,t}}) = \sum_{k=1}^{K} w_{k,h,t} \ln(p_{k,h,t}), \tag{7}$$

and as above, $b(\mathbf{p_{h,t}}) = \prod_{i=1}^{K} p_{i,h,t}^{\beta_i}$ is a Cobb-Douglas price aggregator.

Household-level heterogeneity is introduced in the structural model through the group-specific intercepts $\alpha_{i,h}$, which are allowed to depend linearly on both observable and unobservable household characteristics:

$$\alpha_{i,h,t} = \tilde{\alpha}_i + \mu_h + \pi_i \mathbf{Z_h} + \delta_t. \tag{8}$$

In equation (8), $\tilde{\alpha}_i$ is a product-nutrition cluster group intercept, μ_h is a market fixed effect, and $\mathbf{Z_h}$ is an extensive set of household-level demographics, including household income, male and female head education, age, labor supply, race, household size, and family composition. These variables are shown in Table A-1. We also include in $\mathbf{Z_h}$ the purchasing environment controls discussed in Section 2.1. Common macroeconomic fluctuations are controlled for with the inclusion of month and year fixed effects δ_t.[15] Our aim in this paper is to use a flexible model specification that accounts for individual heterogeneity but that is rich enough to evaluate the relative merits of the different food taxes under consideration. A full exploration of heterogeneity in food demand is beyond the scope of this paper and would add little, since we are primarily interested in the average/aggregate impacts of the taxes as opposed to their distributional effects.[16]

Estimating this system of structural equations involves solving two econometric problems. First, instruments are required to identify the parameters of interest, since in general $E(u_{i,h,t}|p_{j,h,t}, m_{h,t}) \neq 0$. The endogeneity of prices and expenditure comes from a straightforward simultaneous equations problem. If there is unobserved demand heterogeneity or local demand shocks that are correlated with local prices or with household expenditure levels, then prices and expenditures are endogeneous. The instruments used thus need to be correlated with

[15] We also have estimated models using month-by-year fixed effects and the estimates are virtually identical. These results are available upon request.

[16] In an earlier version of this paper, we estimated different sets of parameters for different income levels. While this produces insightful results on income gradients, it adds little to evaluating the relative merits of different types of food taxes. Furthermore, we found little evidence of differential price responsiveness across the income distribution. We leave distributional questions and more detailed investigations of heterogeneity in the demand system for future research.

prices and expenditures but uncorrelated with unobserved demand heterogeneity and demand shocks, conditional on the other controls in the model. Second, due to the large sample size, a computationally efficient approach is required for solving the system of 33 structural equations.

Our choice of instruments is informed by the existing economic literature. Prices are typically set by stores at the market level. At the same time, stores purchase the majority of their products from national manufacturers who typically charge a uniform wholesale price. Thus, store level prices can be seen as the combination of market level factors and production cost factors common across markets. If we assume that the market-level pricing factors are independent across markets once the common production cost is accounted for, we can use the average price of each product across other markets as instruments (Hausman, Leonard and Zona, 1994; Nevo, 2003).[17] These prices are unlikely to be influenced by factors affecting only local demand variation, and it is important to emphasize that any national-level demand shifts are accounted for by the year and month fixed effects. We construct price instruments by computing the average price of a product identified at the UPC-level in the same month and year in all other markets other than the one the consumer has made the purchase in. These prices are then aggregated at the product-nutrition cluster group level.

Additionally, we use local macroeconomic variation to identify expenditure coefficients, which assumes that expenditure is related to a non-linear function of market real income (Blundell and Robin, 1999), an index of market housing prices (Campbell and Cocco, 2007), and market unemployment(Attanasio et al., 2009). We approximate the relationship between log expenditure and log monthly per-capita market income, log quarterly housing prices, and log monthly unemployment by a second order polynomial expansion that includes squared terms and interactions between all of these variables.[18] For both prices and expenditures, the first stage residuals are added as additional regressors to the estimating equations (Hausman, 1978).

The first stage price estimates are shown in Appendix Tables A-3, A-4, A-5, and the expenditure first stage is shown in Appendix Table A-6. For all instruments, the first stages are strong in the sense that the instruments are highly correlated with the endogeneous independent variable. For prices, the own price instruments are universally close to 1 and are highly statistically significant, even though standard errors are clustered at the census tract level. The expenditure instruments also are strong, with a first-stage F-statistic of 24.92.

[17] Finkelstein et al. (2012) uses a weighted average of prices within each respondent's market and month, where the weights are the inverse distance from the respondent's census tract. While this instrument addresses certain endogeneity issues related to individual demand heterogeneity, it does not address endogeneity coming from market-level demand shocks as well as unobserved heterogeneity correlated across space within a market.

[18] Monthly per-capita income come from the US Bureau of Economic Analysis, housing prices are from the Federal Housing Finance Agency's Home Price Index, and unemployment rates come from the US Bureau of Labor Statistics.

Computationally, the estimation of the system of 33 structural equations is inherently difficult due to the non-linearity of the specification in conjunction with the large number of observations and model parameters. However, the system of equations is conditionally linear (Blundell and Robin, 1999; Lecocq and Robin, 2006). If we condition on $b(\mathbf{p_{h,t}})$, the system of budget shares given in equation (6) is linear in the unobserved parameters $\tilde{\alpha}_i, \mu_h, \pi_i, \delta_t, \gamma_{ij}, \lambda_i$ (recall that in our specification $a(\mathbf{p_{h,t}})$ is approximated by the Stone price index and does not depend on unknown parameters). This suggests a recursive estimation approach. Given an initial guess of the β_i parameters, estimates of the $\tilde{\alpha}_i, \mu_h, \pi_i, \delta_t, \gamma_{ij}, \beta_i, \lambda_i$ parameters can be easily obtained from a linear moment estimator. We then use the estimated β_i parameters to update the value of $b(\mathbf{p_{h,t}})$ until convergence is reached across iterations. The convergence threshold we use is that the maximum change in each β_i across iterations must be less than 0.0001 in absolute value.

Because the budget shares must sum to one, the system of budget share equations imply the following parameter restrictions: $\sum_{i=1}^{14} \hat{\beta}_i = 0$, $\sum_{i=1}^{14} \hat{\lambda}_i = 0$ and $\sum_{i=1}^{14} \hat{\gamma}_i = 0$ for all K products. Without computational constraints, one could impose these parameter restrictions and estimate the model by maximum likelihood. Given our sample size, this approach is not feasible in our context. One also could omit one category from the estimation and use these restrictions to solve for the parameters from the omitted product group. Alternatively, we estimate the model using all of the product groups and then test the parameters restrictions. In all cases, our estimates are consistent with these restrictions.[19]

Another identification concern arises when examining the role of purchasing environments in consumer shopping behavior. Since stores do not locate randomly with respect to consumer demand, it is very likely that the proximity of different store types is related to unobserved aspects of consumer demand. To address this problem, we include in equation (6) census-tract-level means of each of the purchasing environment measures. Thus, the purchasing environment coefficients are identified off of changes in these environments from store entry and exit over time. This approach is conceptually similar to controlling for census tract fixed effects (Lecocq and Robin, 2006), but with thousands of census tracts, it is not possible to include them in the model. Census-tract-level means control for any unobserved demand heterogeneity that is fixed over time within tracts and that is correlated with store locations. The identifying assumption is that store entry and exit are uncorrelated with unobserved aspects of household demand. Given the myriad controls in our model and the fact that the households in our sample do not

[19] In particular, we find $\sum_{i=1}^{14} \hat{\beta}_i = 0, \sum_{i=1}^{14} \hat{\lambda}_i = 0, and \sum_{i=1}^{14} \hat{\gamma}_i$ is between 0 and 2×10^{-7} in absolute value.

move while they are in the panel, we believe this assumption is plausible.

The expenditure elasticities and the uncompensated price elasticities play a central role in our structural demand model. For two products i and j, the procedure described above produces estimates of $\hat{\beta}_i, \hat{\gamma}_{i,j}, \hat{\lambda}_i$. Let $\hat{\Sigma}$ be the (partial) asymptotic covariance matrix corresponding to these three coefficients. The expenditure elasticity for product i is given by:

$$\epsilon_{i,h,t} = 1 + \frac{\beta_i}{w_{i,h,t}} + \frac{2\lambda_i}{w_{i,h,t} b(\mathbf{p_{h,t}})} log(m/a(\mathbf{p_{h,t}})).$$

The estimated variance of the expenditure elasticity is then given by:

$$\widehat{Var(\epsilon_{i,h,t})} = \left(\frac{1}{w_{i,h,t}}\right)^2 \widehat{Var(\beta_i)} + \left(\frac{2\ln(m/\widehat{a(\mathbf{p_{h,t}})})}{w_{i,h,t}\widehat{b(\mathbf{p_{h,t}})}}\right)^2 \widehat{Var(\lambda_i)} + \frac{4\ln(m/\widehat{a(\mathbf{p_{h,t}})})}{w_{i,h,t}^2 \widehat{b(\mathbf{p_{h,t}})}} \widehat{Cov(\beta_i, \lambda_i)}$$

The cross-price elasticities are given by:

$$\epsilon_{i,j,h,t} = -\frac{\widehat{\beta}_i w_{j,h,t}}{w_{i,h,t}} + \frac{\widehat{\gamma_{i,j}}}{w_{i,h,t}} - \mathbb{I}(i=j) - \widehat{\lambda}_i \left(\frac{2\ln(m/\widehat{a(\mathbf{p_{h,t}})})}{\widehat{b(\mathbf{p_{h,t}})}} \frac{w_{i,h,t}}{w_{j,h,t}}\right) - \widehat{\lambda}_i \widehat{\beta}_j \frac{(\ln(m/\widehat{a(\mathbf{p_{h,t}})}))^2}{w_{i,h,t}\widehat{b(\mathbf{p_{h,t}})}},$$

where $\mathbb{I}(i=j) = 1$ if $i = j$ and 0 otherwise.
Define the following 3×1 vector $\hat{g}$:

$$\hat{g} = \begin{pmatrix} -\frac{w_{j,h,t}}{w_{i,h,t}} - \hat{\lambda}_i \frac{(\ln(m/\widehat{a(\mathbf{p_{h,t}})}))^2}{w_{i,h,t}\widehat{b(\mathbf{p_{h,t}})}} \\ \frac{1}{w_i} \\ \frac{\ln(m/\widehat{a(\mathbf{p_{h,t}})})}{w_{i,h,t}\widehat{b(\mathbf{p_{h,t}})}}(-2w_{j,h,t} - \hat{\beta}_j \ln(m/\widehat{a(\mathbf{p_{h,t}})}) \end{pmatrix}$$

By the Delta Method, the estimated variance of the cross-price elasticity is (approximately) given by:

$$\widehat{Var(\epsilon_{i,j})} \approx \hat{g}' \hat{\Sigma} \hat{g}$$

Below, we report average elasticities, which are computed over all households and time periods. Average elasticities are accompanied by average standard errors that are calculated using these variance formulas. Because the elasticities are functions of the parameters and of the data, we need to calculate standard errors that incorporate the variability coming from both of those sources. We do this by first calculating the standard error for each observation, using the standard errors from the QAIDS model estimates that are clustered at the census tract level. We then take the average of these standard errors across all households and time periods. To our knowledge, this is the first analysis to provide standard errors for the average elasticities

calculated from QAIDS model coefficients that also incorporate geographic clustering.

3 Estimation Results

3.1 Elasticities

Our structural model estimates a 33-equation demand system, which makes the discussion of over 1000 estimated elasticities inherently difficult. The full set of elasticity point estimates is available in Appendix Tables A-7, A-8, A-9, while the full set of standard errors is available from the authors.

We now take a closer look at the average own-price and expenditure elasticities. Their point estimates and standard errors are reported in Table 3. A bar chart of these elasticities is presented in Figure 4. In general, the elasticities are estimated with a very high level of precision, even though we cluster by census tract.

The own-price elasticities reveal that most products are fairly price inelastic. Soda, milk and cold beverages provide the exception. The price elasticity of the two soda groups are -2.26 and -2.20, and the elasticities for cold beverages range between -1.81 and -2.06. In a review of the literature, Powell et al. (2013) report an elasticity for sugar-sweetened beverages of about -1.21. Thus, our estimates show these products to be more elastic than has been found in previous work. The studies reviewed by these authors estimate the price elasticity from changes in state level taxes, which are very small and likely to suffer from salience issues. In contrast, our estimates are identified off of much larger price variation. Table 3 also shows fruits and vegetables to be rather elastic, with elasticities that range from -0.83 to -1.38.

At the other extreme, meat, snacks, cereal, and pasta are fairly inelastic. Fruits and vegetables are between 2-5 times more price elastic than meat, grain/pasta/bread or breakfast cereals. It also is interesting to note that the elasticities vary substantially both between and within aggregate categories. Two product groups in the cereal and meat categories (21, 23) have small positive elasticities but are economically insignificant. One group of packaged meals (group 14) has a larger own-price elasticity of 0.12, which still is small. As Figure 4 shows, this group also has a negative expenditure elasticity, suggesting that at least some of the packaged meals can be thought of as Giffen goods.

Table 3 reports average expenditure elasticities and standard errors as well, and the results highlight the importance of including expenditure elasticities in simulating consumer responses to different food-based taxes. For soda in particular, these elasticities are greater than 4.

These expenditure elasticities indicate that large income effects will generate sizable reductions in overall expenditures due to a tax on these products that would be missed by examining substitution patterns alone. We also find that one of the packaged meal categories (14) and one of the snacks/candy categories (23) are inferior goods, meaning that overall consumption might increase if you tax these goods. Depending on the size of this response, taxing packaged meals or snacks/candy could increase total purchases of different nutritional categories. All other products are normal goods, with income elasticities that vary considerably both between and within product categories.

The substantial variation in elasticities within product categories indicates that consumers respond differently to prices changes for products within a category. While there are many additional dimensions of the product space that are not explicitly accounted for here, in our model product groups are distinguished by their location in the nutrition space. Consider the packaged meals category. Within this category, the k-median algorithm partitions the products into two groups. By looking at the nutritional characteristics of these groups (Table 2), we see that these groups are very different. One of the groups has median calories per serving of 330 compared to 1518 in the other group. Similarly, the low-calorie group has substantially lower amounts of salt, sugar, and fat. Thus, one of the groups can be thought of as the "healthy" group of products while the other one consists of products which are substantially less healthy along a number of different dimensions. Interestingly, it is the healthier group of packaged meals products which is a Giffen good. The variation in elasticities appears to be related to the underlying nutritional characteristics of the product groups. This begs the questions of whether there may be a systematic relationship between nutrients and elasticities. We explored this further in a series of multivariate correlation analyses: products higher in sugar and protein are significantly more price elastic, while products higher in carbohydrates are less price elastic. Though descriptive in nature, such correlations suggest that sugar taxes in particular might be an effective method to alter consumer purchasing behavior.

In order to get a sense of the patterns of cross-price elasticties, Table 4 presents symmetric matrices of own and cross-price elasticities for each aggregate category. The diagonals in each panel show the own-price elasticities (and are identical to those in Table 3), and the off-diagonals show the cross-price elasticities among all the nutrient clusters within product type. On the whole, these estimates show little evidence of much within-product substitution behavior. Soda exhibits the most within-category substitution, with cross-price elasticities of 0.18 to 0.2. The cross-price elasticities for milk also are sizable, at 0.11 and 0.26. For the remainder of the

products, consumers do not respond to price changes by purchasing more (or less) of the same product with a different nutritional composition.

The estimates in Table 4 do not necessarily imply that cross-group substitution patterns are low, as it could be that substitution is more likely across product types than within product categories. Appendix Figure A-1 displays the full pattern of cross-price elasticities in a graphical format. The size of the rectangles corresponds to the magnitude of the cross-price elasticity, while the color denotes the sign. We notice a broad dispersion of magnitudes and signs for the off-diagonal elements of the cross-price elasticity matrix, suggesting that the substitution patterns are indeed more likely to occur across product groups. A close investigation of the standard errors revels that most of the elasticities also are statistically significant. This emphasizes the fact that it is important to model the full range of substitutions when examining the effect of price-based policy manipulations.

Furthermore, the complex pattern of complements and substitutes in this large demand system indicates that policy interventions need to be implemented with care, since the potential for unanticipated consequences resulting from a price intervention is inherent in the substitution pattern. In Appendix Figure A-2, we explore the cross-price elasticities in more detail for product groups 11-14, corresponding to soda and packaged meals. For example, soda is a substitute good for most other products including fruits and vegetables, breakfast foods, meat and packaged meals. But, it is a complement for snacks and candy. While packed meals are own-price inelastic, they are complements for fruits and vegetables and grain/pasta/bread. The opposing pattern of signs for soda and packaged meals indicates that price changes are likely to have very different effects when applied to these two product groups. Soda is very own-price elastic and an increase in the price of soda increases the demand for many other food items. In contrast, packaged meals are own-price inelastic (and possibly a Giffen good), and an increase in their price reduces the demand for many other food items.

It is important to highlight that our cross-price elasticity results do not imply that increasing the price of goods perceived to be unhealthy, such as soda or packaged meals, necessarily leads to consumers switching to more healthy options, such as fruits and vegetables. In the case of soda and packaged meals, we see responses that are polar opposites. Higher soda prices increase the demand for various other food items, while higher packaged meal prices decrease the demand for a range of food items. The response to prices increases for all other food categories is somewhere in between. If we think of the household food consumption bundle before and after a price increase, we see that a price increase alters the composition of the

bundle in a fairly complex fashion, typically increasing the demand for some product groups and decreasing the demand for others. The overall impact on nutrients, however, is unclear and needs to be simulated in a fashion similar to that in the next section of this paper. This also emphasizes the fact that it is potentially misleading to estimate single product demand models in order to infer the benefit of a nutrition policy on household nutrient consumption.

3.2 Demographic Heterogeneity and Store Density

Our model estimates also generate a large number of results relating demographic variables to preferences for the different food groups. While detailed tables are available from the authors, it is worth pointing out that we find a variety of demographic gradients for the different product groups. For example, consider income gradients that reflect the extent to which consumption of the different product groups varies over the income distribution. Our reference category consists of households with income less than $30,000. We find that shares of fruits and vegetables, packaged meals and snacks/candy increase with income, while soda, meat and milk purchases decline with income. These gradients point to a complex relationship between the distribution of nutrients and the distribution of income. Thus, wealthier households spend over 7% more of their budget both on fresh fruits and vegetables as well as on packaged meals, potentially reflecting the presence of opposing forces: increased health awareness but also higher opportunity cost of time.

Shares of fruit and vegetables increase with education, while the soda share is uncorrelated with education, and snack/candy and meat purchases are negatively correlated with educational attainment. The product shares in the household food budget vary with age as well, with younger households consuming more fruits and vegetables. The share of consumption decreases with age but then increases again after retirement. Overall, however, the fruit and vegetable shares remains lower for households above 65 than for households below 35 years of age. The shares of soda, packaged meals and snacks consumption increase with age but then decrease again at retirement. The share of meat consumption increases with age throughout.

Expenditure shares are strongly associated with female labor supply, with large differences between households with a working and a non-working female head.[20] Households with a non-employed female head purchase more soda, fewer packaged meals, and less meat and snacks/candy. In contrast, non-employment of the male head of household does not have a statistically significant impact on the budget shares of the different products consumed. The

[20] Note that households in our data can have both a male and a female head.

presence of young children (under 12 years of age) indicates an increased share of consumption spent on fruits/vegetables, cereals and milk and a lower share of expenditure allocated to soda, packaged meals or canned foods. This pattern may reflect parental concerns to provide healthier food options to their children. Note, however, that the higher share of consumption allocated to cereal is implicitly associated with a higher sugar intake.

A byproduct of our analysis is the estimate of the role of purchase environments in driving budget share allocations. The role of purchasing environments in explaining variation in food purchasing behavior recently has been featured extensively in public policy discussions. While this paper is not aiming to answer the question of the impact of food environments on nutrition, it is interesting to note that our estimates do not support a systematic relationship between store availability and purchased food bundles. Our model estimates the effect of one additional store of the given type in the given distance from the census tract centroid on the budget share, conditional on all of the other controls in the model. We find little relationship between the distance to stores and the consumption shares of the different products. Despite the large number of observations and the small standard errors, very few coefficients are even statistically different from zero at conventional levels. Thus, our estimates can rule out even small effects of purchasing environments on budget shares of these goods.

Since our estimates are capturing statistical averages for a large, nationally-representative, mostly urban population with a certain degree of technical sophistication that is required to participate in the sample, this indicates that the food purchasing environment, as measured by the store density, is not a primary driver of food consumption and nutrition trends at this level of aggregation and for these consumers. It is important to note that our paper does not directly address the impact of "food deserts," which have been documented in a number of focused case studies, as they primarily impact low-income consumers.[21] Quantifying the "food desert" problem is substantially more complicated than measuring store density by different types of stores, even though store density is one particular aspect of it. In particular, this study does not capture more subtle aspects of product quality and food choice availability within store types, which are likely to be an important component of the purchasing environment for low-income consumers.

[21] An overview of the "food desert" literature that examines the relationship between purchasing environments and shopping behavior can be found in Bitler and Haider (2011). They conclude that the lack of convincing data has precluded an understanding of how food deserts affect consumer purchasing behavior and nutrition.

4 Tax Simulations

Using the estimated structural demand model presented above, we construct a number of policy counterfactuals that have received attention by policymakers. We consider the introduction of three types of taxes aimed at changing the behavior of households and encouraging healthier consumption. One type of tax is a product-specific tax, where a product group considered to have negative health impacts is identified and all products that are part of that group are taxed proportionally. Currently, legislatures in Hawaii, Oregon and Vermont are considering bills aimed at taxing soda, and 35 states either have such taxes or include soda as part of the sales tax base. Given the important role played by packaged meals and snacks in the American diet and the low nutritional content of these goods, we evaluate the introduction of taxes on these product groups as well. Similar taxes already exist in Europe.

We also examine the effect of introducing nutrient-specific taxes, specifically taxes on fat, sugar, and salt. The taxable amount is directly proportional to the share of the specific ingredient in the product formulation. Consider a 12 fl. oz. can of Coca-Cola. It has 39g of sugar, 45mg of sodium and 0g of fat. This also corresponds to 140 calories. Now consider a 12 fl. oz. can of Diet Coke. It has 0g of sugar, 40mg of sodium, 0g of fat and 0 calories. A fat tax will not affect the price of either product, a sugar tax will affect the price of the first product but not the second product, while a salt tax will affect the price of both products in proportion to their respective sodium contents. Nutrient-specific taxes have the potential benefit that they target the harmful nutrients more directly than product-specific taxes, but they also have the drawback that they might reduce consumption of healthier products due to substitution effects or due to correlations among different nutrients in these products. The third type of tax we consider is a tax on SSBs, which can be thought of as a hybrid that targets a subset of products that contain large amounts of sugar. This tax has received considerable attention in prior work, and we include it here in order to be able to compare our results with the existing literature.

Throughout, we simulate the effect of a 20% tax on products and nutrients. Because we lack data that allow us to model the supply-side, we are not able to consider the incidence of these taxes. Thus, when we are referring to a "20% tax on soda," one can view this as a tax on soda that generates a 20% price increase. While the incidence of any such taxes we simulate are important in their own right, the existing literature on the incidence of commodity taxation tends to find tax pass-through rates close to 1 (Harding, Leibtag and Lovenheim, 2012; Kenkel, 2005; Besley and Rosen, 1999). That the incidence of these taxes tends to fall on consumers

suggests that ignoring the supply-side is not a large drawback in the short-run.[22] We also do not consider issues pertaining to the administration of these taxes. This paper is focused on identifying the expenditure and nutritional effects of these taxes, but we acknowledge that how such taxes are structured and their administration could be very important in driving any impacts they may have (Chetty, Looney and Kroft, 2009; Slemrod and Kopczuk, 2002). For the purposes of this analysis, we assume that taxes are remitted by wholesalers or on the raw nutrients themselves (e.g., a tax on refined sugar or iodized salt). This would largely eliminate salience issues because the tax amount would be reflected in the shelf price, but there still could be complications with evasion and avoidance, particularly if the taxes are not national in scope.[23]

Throughout, we use the estimated structural model to simulate the impact of the taxes under consideration. We simulate the effect of product-specific taxes by calculating, for each household month, the effect of a 20% increase in soda, packaged meals or snacks/candy on expenditures on each product-nutrient cluster group, using the estimates from our demand model. We then calculate new budget shares and, for each of the 8 nutrient groups, calculate

$$\Delta Nutrient = \Delta Expenditure * \frac{\$}{Nutrient}. \tag{9}$$

That is, we multiply the estimated change in expenditure for the tax by the per-nutrient price of the good, separately for each product group. The expenditure changes are adjusted for the fact that the price of the taxed good is now higher, so we are not attributing increased expenditure on a good from a price increase to higher consumption. For nutrient-specific taxes, we engage in a similar exercise, but change the price of the products proportionally with respect to the amount of the nutrient included in its formulation. For example, a 20% sugar tax will increase the price of fruits and vegetables by 7.84%, the price of soda by 14.16%, and the price of cereal by 15.76%.

Using the average amount of each nutrient in each product-nutrient cluster group, we first find the average change in price for the group if the price of the given nutrient increased by 20%. We then estimate the effect on expenditures in each group from these price changes

[22] It also is possible that companies might reformulate their products in response to these taxes. We do not have the capability of including such endogenous responses in our model. Our estimates thus ignore any effects on nutrition that would come from making given products healthier. One can interpret our results as the effect of a small government or entity (such as a state) imposing a tax that would be too small to induce producers to reformulate their products. We also note that the nutrients we consider are very difficult to replace without altering the taste and desirability of most products. Any reformulation likely would increase the effect of these taxes on nutrition, however.

[23] See Lovenheim (2008); Goolsbee, Lovenheim and Slemrod (2010); Baughman et al. (2001); Lovenheim and Slemrod (2010); Harding, Leibtag and Lovenheim (2012) for estimates of the extent of cross-state tax avoidance with respect to cigarette and alcohol regulations.

and calculate equation (9). The net effect of a given nutrient-specific tax contains own- and cross-price effects as well as expenditure effects from changes in several different product prices at once. Without a structural model that allows estimation of these substitution and income effects linked to nutrition data, simulating nutritional taxes would not be possible.

In Panel A of Table 5, we present the simulation results for a counterfactual 20% soda tax. We calculate expenditure, price and nutritional changes for all 33 categories, but we aggregate them to the original 14 product groups to present results for parsimony. Estimates for the full set of 33 groups are available from the authors upon request. We report the percent changes in expenditures and budget shares for each product as well as the overall effect on expenditure after the imposition of the tax. We further show the changes in nutrients purchased for each product group and for the overall shopping basket. In the panel titles, we report the average indirect utility cost of the tax, which is calculated from equation (1) for every household under both the original and new tax regimes. These indirect utility changes provide a measure of the utility cost of the product substitutions and changes in expenditure from each tax.[24] All changes shown are averages over all households and time periods, which are computed at the individual level and use the full demand system. Thus, we would expect the simulations to provide more information than is contained in the average price and expenditure elasticities, as a result of both the aggregation process and the interactions between all products in the demand system.

Because of the large expenditure elasticity for soda, a 20% soda tax reduces soda expenditure by 3.92% and the share of soda in the household budget by 3.43%. This reduces caloric intake by 2,231 calories and sugar intake by 675g. These changes translate to a reduction in caloric intake from soda by the equivalent of about 16 cans of Coca-Cola per month. As suggested by our results on the cross-price elasticities, soda taxes also induce a shift towards milk and snacks/candy and away from packaged meals and meat. These changes, along with a 1.66% reduction in total expenditures due to the soda tax, lead to declines in the purchase of each nutrient group. In particular, calories decline by 4.84% (4,779 calories) and sugar purchases decline by 10.35% (763.6g), the majority of which come from reductions in soda purchases. Declines in fat, salt, cholesterol, carbohydrates and protein also range from 2-4%. The expenditure and price changes from this tax lead to an average decline in indirect utility of 2.5%.

In Panel B of Table 5, we compare these simulation results to those obtained from imposing

[24] Our welfare calculations implicitly use a social welfare function that weights each household equally. See Banks, Blundell and Lewbel (1997) for a thorough treatment of welfare analysis with QAIDS models.

a 20% tax on SSBs. We model this tax as a 20% price increase for all groups that include sugar-sweetened beverages, as such taxes typically are modeled and levied as a per-product tax rather than as a proportion of sugar in the product (Brownell et al., 2009; Finkelstein et al., 2012; Kamerow, 2010). This also makes the SSB and soda taxes more directly comparable. Because it impacts more products, the SSB tax has a larger impact on the types of products consumers purchase and on nutrition. A 20% SSB tax reduces total calories purchased by 8.4%, most of which is occurring due to the almost 20% decline in sugar purchases.[25] This price change also induces a small decline in the other nutrient categories, most of which is due to income effects. Interestingly, despite the large impact of the SSB tax on consumer purchasing behavior and nutrition, its effect on indirect utility compared to the soda tax is small, at 2%. The reason for this result is that by increasing the prices of a broader set of goods, this tax induces less distortionary substitution behavior that has high utility cost. Taxing the suite of sugary beverages limits the ability to substitute across these similar products, which produces larger changes in nutrition at a lower utility cost to consumers.

The effect of taxing another "unhealthy good," packaged meal taxes, has a very different effect than taxing soda. The simulated effect of a 20% packaged meal tax is shown in Panel A of Table 6. Recall that packaged meal purchases are very own-price inelastic. Thus, households end up spending 7.96% more on packaged meals, and the share of packaged meals in the overall budget increases by 5.62%. Due to the patterns of substitution and the fact that our estimates suggest packaged meals are inferior goods, packaged meal taxes have a small but positive effect on both total expenditure and on nutritional intake of the household. They also reduce indirect utility by 1.4%. Total calories increase by about 2.11% and sugar intake increases by 2.25%. Note, however, that nutrients from packaged meals decline substantially. This reflects two forces: first, the tax leads to less overall expenditure on packaged goods and second, there is a sizable shift in expenditure within this category to the healthier group (as it has a positive own-price elasticity). Thus, packaged meal taxes would reduce the nutrients purchased coming from packaged meals. However, it would not support the purchase of more nutritious food bundles overall, even though this product group is quite unhealthy.

In Panel B, we show the effect of a 20% snack and candy tax. Recall that snacks and candy are also relatively price inelastic, although less so than packaged meals. A 20% tax on these goods causes an overall increase in budget share and expenditure, but there is a decline in

[25] This estimate is substantially larger than the closest prior estimate in the literature. Finkelstein et al. (2012) estimates a 20% SSB tax would decrease per-capita daily calories by 24. Our estimates translate into a daily per-capita reduction of 133 calories. We believe much of the difference in our estimates can be accounted for by the fact that their model does not allow for income effects of taxes, which we show are very important, especially for beverages.

quantity of these items purchased. Calories, fat, sugar and salt, in particular, decline in this product group due to the tax. One small 3.3g Reese's Peanut Butter Cup™ package contains 693 calories, 42.9g of fat, and 495mg of sodium. This tax therefore is the equivalent of reducing consumption of 1-2 of these products per month. There also is a small shift towards fruits and vegetables and away from other dairy, canned food/sauces and soda. The overall impact of the candy/snack tax is to slightly reduce calories by 1.21% (1194.5 calories), fat by 0.63%, and salt by 1.12%. However, sugar and cholesterol consumption also increase slightly. The overall effect of this tax on nutrition therefore is somewhat ambiguous in sign, but it is likely to be small. As with packaged meal taxes, the indirect utility cost of this tax also is modest, at 2%. Compared with sizable nutritional effects of SSB and soda taxes, taxes on packaged meals and candy/snacks produce much smaller changes at a similar utility cost.

One potential complication with our tax simulation approach is that if consumers switch to lower-priced versions of the same goods, such as generic brands, in response to product or nutrient taxes, expenditures could decline without affecting nutrition. However, we would be attributing the resulting expenditure declines to changes in total purchases within each product group rather than changes in the types of products consumers buy within each group. This would cause us to over-state the effect of the taxes we study on nutrition. Additional robustness checks (not reported here), which split the product categories by the median price suggest that our results are not over-stating nutritional effects due to consumer substitution to cheaper goods with the same nutritive content.

We now turn to simulation results from fat, sugar and salt taxes, which are shown in Table 7. The overall findings shown in the table point to these taxes being much more effective at altering nutritional bundles than are product-specific taxes, without generating a larger impact on total expenditures or indirect utility. This is suggestive that these taxes have far more potential to impact nutrition without causing much larger deadweight loss from distorting purchasing behavior.

In Panel A of Table 7, we show the effect of a 20% tax on fat. The fat tax reduces caloric intake by over 19% (18,984 calories), reduces fat purchases by 30.25% (1,424.5g), decreases cholesterol by 10.83% (1,223.5g), and reduces sodium by 10.63% (19,913.5mg). Despite these large changes, average indirect utility only declines by only 2%, which is comparable to the SSB, soda and candy/snack tax welfare losses. Purchase of dairy and packaged meals are reduced the most due to a fat tax. Households however increase their expenditure on soda by 3.16% and on other cold beverages by 1.76%. Such substitution patterns lead fat taxes to only decrease

sugar intake by 6%. Expenditure on fruits and vegetables decrease as well by 0.7%, which together with the substitution to soda suggests fat taxes do not unambiguously support better nutrition. Again, this result highlights the value of considering substitution effects across goods in response to food-based taxes.

Unlike a tax on fat, a 20% sugar tax has a very strong and unambiguously positive effect on nutrition at a small increase in utility cost of 2.6%. It also has a very small overall effect on expenditure of -0.64%. Compared to the soda tax, a sugar tax has a large impact on nutrients purchased. Our estimates show that monthly sugar consumption would drop by 16.41% (1,211.1g), with a 18.54% (18,302.4) decline in total calories. The effect on fat and salt also are sizable, at 12% and 9.63% respectively. Total consumption of cholesterol and protein declines significantly as well. As Panel B illustrates, only about 14% of the reduction in sugar purchases comes from soda, which underscores the limits of taxing only this product. Instead, households reduce their sugar intake from all product groups. This broad-based reduction reflects both income and substitution effects: consumers purchase less overall and they also purchase healthier products. Recall that many goods high in sugar also have sizable amounts of other nutrients, such as fat and salt. We find that sugar taxes cause households to substitute both to healthier product groups and to healthier versions of the same product (e.g., diet versus regular soda).

The reason that sugar taxes are so effective is the same reason why the nutritional taxes in general are much more effective at altering nutritive bundles without causing a larger utility decline: they are very broad based. The vast majority of product groups and of products in general contain sugar, fat and salt. Sugar in particular is highly prevalent in American foods. Taxing these nutrients does not allow consumers to substitute to other goods that also contain these nutrients. Rather, they shift to healthier products and decrease overall consumption. Despite the fact that these taxes are predicted to cause changes in purchasing behavior that lead to large nutritional gains, their broad-based nature requires there to be much less of a behavioral response for a given amount of change in the nutritional bundle than for product-specific taxes. Thus, they are able to exert a large influence on nutrition at a similar welfare cost to the product-specific taxes we consider.

In the last panel of Table 7, we investigate the counterfactual of imposing a 20% salt tax. Similar to the other nutrient taxes, the overall impact on nutrition is sizable: total salt purchased by the household decreases by 10.0% (18,792mg) per month, which is equivalent to about 8 teaspoons of salt. While the biggest share of the reduction comes from condiments and spices

(including salt products), substantial savings also are achieved from baking goods, snacks, and grains and pasta. This tax reduces total calories and fat by over 10% as well. However, soda expenditures increase by 8.3%, which significantly reduces the effect of a salt tax on sugar. Again, this result underscores the importance of considering substitution behavior in assessing the effects of a given food-based tax. The cost of this tax in terms of indirect utility also is low, at 1.2%, suggesting that substantial nutritional gains in many areas could be achieved at minimal cost with a salt tax.

The discussion above suggests that when considering the choice between product- and nutrient-specific taxes, policy makers can achieve substantially larger improvements in nutrition by implementing nutrition taxes, rather than product taxes. Both sets of taxes alter indirect utility similarly. But, nutrient taxes have far larger impacts on nutrition, which is the ultimate goal of these taxes. While there are some unintended consequences of some of these nutrient taxes, such as higher soda consumption and decreased fruit and vegetable consumption, on the whole they lead to large reductions in purchases of most nutrient categories. One tax simulation stands out, however, which is the sugar tax. Not only is the indirect utility cost modest, it also leads to substantial reductions in the purchased amount of calories, sugar, fat, salt and cholesterol.

5 Conclusions

This paper investigates the degree to which food taxes can be used to support better nutritional choices among US households. The last few decades have seen a sharp rise in obesity and overweight in the US, which is now being referred to as the "obesity epidemic." With a large proportion of calories being consumed in the home, there has been a surprising lack of empirical examination of policy tools that can be used to induce consumers to purchase more nutritious food.

One reason for this gap in the literature is the need for detailed data on food purchases, prices, demographics, local store availability and nutrition as well as a parsimonious way to model the joint product-nutrient space. We construct the first data set of which we are aware that contains all of these components over a longer time horizon. We use 7 years of Nielsen Homescan data, merged with the nutritional content of foods purchased and with store density information that allows us to examine these policy issues with a level of detail and technical sophistication that previously were not feasible. One of the central contributions of this paper

is to use the UPC-level nutritional information to find distinct nutrient clusters of goods within each product group that define the types of goods over which consumers can substitute. We then estimate a large utility-derived demand system, which identifies the price and expenditure elasticities across 33 product groups at the household-month level. Using our structural demand estimates, we simulate the effect of 20% product taxes on soda, SSBs, packaged meals, and candy/snacks and 20% nutrient taxes on fat, salt, and sugar.

Our main finding is that nutrient-specific taxes have much larger effects on nutrition than do the product-specific taxes we study. However, they do not cost more in terms of consumer utility. While a 20% soda tax reduces sugar purchases by 10.35%, it only reduces overall caloric intake by 4.84%. Taxing packaged meals actually increases slightly overall caloric intake, even though this product group is the unhealthiest per serving. Taxing snacks/candy also has at most small impacts on the purchased nutritive bundle.

In contrast, taxing nutrients has large impacts on nutrition without producing larger welfare losses than product-specific taxes, largely because of the broad-based nature of these taxes. Among the three nutrient-specific taxes, sugar taxes stand out as particularly effective. A 20% sugar tax reduces sugar consumption by 16.41% while also reducing caloric intake by 18.54% and salt consumption by 9.63%. Consumer indirect utility declines by 2.6% as a result, which is very similar to the utility cost of a soda tax and SSB tax.

Overall, our estimates suggest that the use of nutrient-specific taxes could have an important effect in inducing healthier purchasing behavior among consumers. The results we present are consistent with theoretical arguments of the value of having broad-based taxes that are harder to avoid. Nutrients provide such a broad base, and our findings indicate that taxes on nutrients would do much more to support healthier nutritional choices than would taxes on products. The focus among public health officials and policy makers on the product-specific taxes (such as soda taxes) and even on nutrient-specific taxes on a small set of goods (such as an SSB tax) may be misplaced; broadening the tax base would achieve larger nutritional gains without causing larger overall distortions in household utility.

References

Aguiar, M., and E. Hurst. 2007. "Lifecycle Prices and Production." *American Economic Review*, 97(5): 1533–1559.

Aguiar, M., and E. Hurst. forthcoming. "Deconstructing Lifecycle Expenditure." *Journal of Political Economy*. forthcoming.

Andreyeva, T., M.W. Long, and K.D. Brownell. 2010. "The Impact of Food Prices on Consumption: A Systematic Review of Research on the Price Elasticity of Demand for Food." *American Journal of Public Health*, 100(2): 216–222.

Attanasio, O., L. Blow, R. Hamilton, and A. Leicester. 2009. "Booms and Busts: Consumption, House Prices and Expectations." *Economica*, 76: 20–50.

Banks, J., R. Blundell, and A. Lewbel. 1997. "Quadratic Engel Curves and Consumer Demand." *The Review of Economics and Statistics*, LXXIX(4): 527–539.

Baughman, R., M. Conlin, S. Dickert-Conlin, and J. Pepper. 2001. "Slippery When Wet: The Effects of Local Alcohol Access Laws on Highway Safety." *Journal of Health Economics*, 20.

Berry, S., J. Levinsohn, and A. Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–890.

Besley, T., and H. Rosen. 1999. "Sales Taxes and Price: An Empirical Analysis." *National Tax Journal*, 52(2): 157–178.

Bitler, M., and S. Haider. 2011. "An Economic View of Food Deserts in the United States." *Journal of Policy Analysis and Management*, 30(1): 153–176.

Blundell, R., and J.-M. Robin. 1999. "Estimation in Large and Disaggregated Demand Systems: An Estimator for Conditionally Linear Systems." *Journal of Applied Econometrics*, 14: 209–232.

Brownell, K.D., T. Farley, W.C. Willett, B.M. Popkin, F.J. Chaloupka, J.W. Thompson, and D.S. Ludwig. 2009. "The Public Health and Economic Benefits of Taxing Sugar-Sweetened Beverages." *The New England Journal of Medicine*, 361: 1599–1605.

Burda, M., M. Harding, and J. Hausman. 2012. "A Poisson Mixture Model of Discrete Choice." *Journal of Econometrics*, 166: 184–203.

Campbell, John, and J. Cocco. 2007. "How do house prices affect consumption? Evidence from micro data." *Journal of Monetary Economics*, 54(3): 591–621.

Cawley, John. 2010. "The Economics of Childhood Obesity." *Health Affairs*, 29(3): 364–371.

Chetty, R., A. Looney, and K. Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review*, 99(4): 1145–1177.

Chouinard, H.H., D.E. Davis, J.T. LaFrance, and J.M. Perloff. 1998. "Fat Taxes: Big Money for Small Change." *Forum for Health Economics & Policy*, 10(2).

Cutler, D., E. Glaeser, and J. Shapiro. 2003. "Why Have Americans Become More Obese?" *Journal of Economic Perspectives*, 17(3): 93–118.

Deaton, A., ed. 1981. "Theory and Management of Consumer Behavior." , ed. A. Deaton, Chapter Some Engel Curves. Cambridge.

Decoster, A., and F. Vermeulen. 1998. "Evaluation of the Empirical Performance of Two-Stage Budgeting AIDS, QUAIDS and Rotterdam Models based on Weak Separability." Discussion Paper, Katholieke Universiteit Leuven.

Dharmasena, S., and O. Jr. Capps. 2012. "Intended and Unintended Consequences of a Proposed National Tax on Sugar-sweetened Beverages to Combat the U.S. Obesity Problem." *Health Economics*, 21(6): 619–756.

Dubois, P., R. Griffith, and A. Nevo. forthcoming. "Do Prices and Attributes Explain International Differences in Food Purchases?" *American Economics Review*. forthcoming.

Duffey, K.J., P. Gordon-Larson, J.M. Shikany, D. Guilkey, D.R. Jr. Jacobs, and B.M. Popkin. 2010. "Food Price and Diet and Health Outcomes: 20 Years of the CARDIA Study." *Archives of Internal Medicine*, 170(5): 420–426.

Einav, L., E. Leibtag, and A. Nevo. 2010. "Recording Discrepancies in Nielsen Homescan Data: Are they Present and Do They Matter?" *Quantitative Marketing and Economics*, 8(2): 207–239.

Finkelstein, E., C. Zhen, J. Nonnemaker, and J. Todd. 2010. "Impact of Targeted Beverage Taxes on Higher- and Lower-income Households." *Archives of Internal Medicine*, 170(2): 2028–2034.

Finkelstein, E., C. Zhen, M. Bilger, J. Nonnemaker, A. Farooqui, and J. Todd. 2012. "Implications of a Sugar-sweetened Beverage (SSB) Tax When Substitutions to Non-beverage Items are Considered." *Journal of Health Economics*, 32: 219–239.

Finkelstein, E., I. Fiebelkorn, and G. Wang. 1998. "Annual Medical Spending Attributable to Obesity: Payer and Service-Specific Estimates." *Health Affairs*, 28: w822w831.

Flegal, K., M. Carroll, R. Kuczmarski, and C. Johnson. 1998. "Overweight and Obesity in the United States: Prevalence and Trends, 1960-1994." *International Journal of Obesity and Related Metabolic Disorders*, 22: 39–47.

Fletcher, J., D. Frisvold, and N. Tefft. 2010*a*. "Can Soft Drink Taxes Reduce Population Weight?" *Contemporary Economic Policy*, 28(1): 23–35.

Fletcher, J., D. Frisvold, and N. Tefft. 2010*b*. "The Effects of Soft Drink Taxation on Soft Drink Consumption and Weight for Children and Adolescents." *Journal of Public Economics*, 94(11-12): 967–974.

Goolsbee, A., M.F. Lovenheim, and J. Slemrod. 2010. "Playing With Fire: Cigarettes, Taxes, and Competition from the Internet." *American Economic Journal: Economic Policy*, 2(1): 131–154.

Griffith, R., L. Nesheim, and M. O'Connell. 2010. "Sin taxes in differentiated product oligopoly: an application to the butter and margarine market." CEMMAP Working Paper No. CWP37/10.

Harding, M., E. Leibtag, and M. Lovenheim. 2012. "The Heterogeneous Geographic and Socioeconomic Incidence of Cigarette Taxes: Evidence from Nielsen Homescan Data." *American Economic Journal: Economic Policy*, 4(4): 169–198.

Hausman, J. 1978. "Specification Tests in Econometrics." *Econometrica*, 46: 1251–1271.

Hausman, J., G. Leonard, and J. Zona. 1994. "Competitive Analysis with Differentiated Products." *Annals of Economics and Statistics*, 34: 159–180.

Kamerow, D. 2010. "The Case of Sugar Sweetend Beverage Tax." *BMJ*, 341: 132–133.

Kenkel, D. 2005. "Are Alcohol Tax Hikes Fully Passed through to Prices? Evidence from Alaska." *American Economic Review Papers and Proceedings*, 95(2): 273–277.

Kuchler, F., A. Tegene, and M.J. Harris. 2005. "Taxing Snack Foods: Manipulating Diet Quality or Financing Information Programs?" *Applied Economic Perspectives and Policy*, 27(1): 4–20.

Lecocq, S., and J.-M. Robin. 2006. "Estimating Demand Response with Panel Data." *Empirical Economics*, 31: 1043–1060.

Lin, B.-H., T.A. Smith, J.-Y. Lee, and K.D. Hall. 2011. "Measuring Weight Outcomes for Obesity Intervention Strategies: The Case of a Sugar-sweetened Beverage Tax." *Economics & Human Biology*, 9(4): 329–341.

Lovenheim, M.F. 2008. "How Far to the Border? The Extent and Impact of Cross-border Cigarette Smuggling." *National Tax Journal*, 61(1): 7–33.

Lovenheim, M.F., and J. Slemrod. 2010. "The Fatal Toll of Driving to Drink: The Effect of Minimum Legal Drinking Age Evasion on Traffic Fatalities." *Journal of Health Economics*, 29(1): 62–77.

Must, A., J. Spadano, E. Coakley, A. Field, G. Colditz, and W. Dietz. 1999. "The Disease Burden Associated with Overweight and Obesity." *Journal of the American Medical Association*, 282: 1523–1529.

Mytton, O., D. Clarke, and M. Rayner. 2012. "Taxing Unhealthy Food and Drinks to Improve Health." *British Medical Journal*, 344: 1–7.

Nestle, M. 2012. "New Data on Calories Reported as Consumed." *Food Politics Blog*, http://www.foodpolitics.com/2012/07/new-data-on-calories-reported-as-consumed/.

Nevo, A. 2003. "New Products Quality Changes, and Welfare Measures Computed From Estimated Demand Systems." *The Review of Economics and Statistics*, 85: 255–275.

Ogden, C., M. Carroll, L. Curtin, M. Lamb, and K. Flegal. 2010. "Prevalence of High Body Mass Index in US Children and Adolescents, 2007-2008." *Journal of the American Medical Association*, 303(3): 242–249.

Pi-Suyner, F. 1991. "Health Implications of Obesity." *American Journal of Clinical Nutrition*, 53: 1595S1603S.

Powell, L., J.F. Chriqui, T. Khan, R. Wada, and F.J. Chaloupka. 2013. "Assessing the Potential Effectiveness of Food and Beverage Taxes and Subsidies for Improving Pub-

lic Health: a Systematic Review of Prices, Demand and Body Weight Outcomes." *Obesity Reviews*, 14(2): 110–128.

Slemrod, J., and W. Kopczuk. 2002. "The Optimal Elasticity of Taxable Income." *Journal of Public Economics*, 84(1): 91–112.

Todd, J., L. Mancino, E. Leibtag, and C. Tripodo. 2010. "Methodology Behind the Quarterly Food-at-Home Price Database." USDA Technical Bulletin 1926.

Yen, S.T., B.-H. Lin, D.M. Smallwood, and M. Andrews. 2004. "Demand for Nonalcoholic Beverages: The Case of Low-income Households." *Agribusiness*, 30(3): 309–321.

Zhen, C., M.K. Wohlgenant, S. Karns, and P. Kaufman. 2011. "Habit Formation and Demand for Sugar-Sweetened Beverages." *American Journal of Agricultural Economics*, 93(1): 175–193.

Table 1: Budget Shares and Prices

Group	Budget Share	Price Per Ounce	Aggregate Category	Budget Share
1	0.025	0.107		
2	0.043	0.101	Fruits & Vegetables	0.091
3	0.023	0.078		
4	0.016	0.029		
5	0.025	0.047	Cold Beverages	0.062
6	0.009	0.035		
7	0.012	0.055		
8	0.002	0.091		
9	0.021	0.308	Warm Beverages	0.034
10	0.011	0.181		
11	0.031	0.029	Soda	0.053
12	0.022	0.021		
13	0.020	0.170	Packaged Meals	0.075
14	0.055	0.211		
15	0.037	0.107		
16	0.015	0.130	Grain, Pasta, Bread	0.085
17	0.033	0.150		
18	0.008	0.130	Canned Food, Sauce	0.040
19	0.032	0.170		
20	0.034	0.161	Cereal, Breakfast	0.053
21	0.019	0.221		
22	0.107	0.171	Other Dairy	0.107
23	0.064	0.198		
24	0.014	0.159	Meat, Protein	0.127
25	0.050	0.191		
26	0.020	0.481	Condiments, Spices	0.040
27	0.019	0.210		
28	0.061	0.317	Snacks, Candy	0.155
29	0.095	0.193		
30	0.029	0.024	Milk	0.047
31	0.018	0.043		
32	0.006	0.064	Baking Goods	0.032
33	0.027	0.206		

Source: Nielsen Homescan data from 2002-2007. Tabulations include only those in the 52 identified Nielsen markets.

Table 2: Median Nutritional Content (per Serving) of Product-Nutrition Cluster Groups

Group	Aggregate Category	Calories	Total Fat (g)	Salt (mg)	Sugar (g)	Cholest-erol (g)	Protein (g)	Calories from Fat (g)	Carbs (g)
1	Fruits & Vegetables	660	0	50	111	0	8	0	160
2	Fruits & Vegetables	140	0	35	12	0	4	0	30
3	Fruits & Vegetables	210	0	1260	10	0	7	0	42
4	Cold Beverages	0	0	5	0	0	0	0	0
5	Cold Beverages	880	0	120	192	0	3	0	219
6	Cold Beverages	1528	0	995	352	0	0	0	405
7	Cold Beverages	240	0	298	56	0	0	0	60
8	Warm Beverages	5000	43	36	684	0	0	656	896
9	Warm Beverages	200	0	79	23	0	0	0	24
10	Warm Beverages	1280	33	400	160	0	7	299	224
11	Soda	0	0	135	0	6	0	0	0
12	Soda	960	0	287	254	25	0	0	272
13	Packaged Meals	1518	78	4070	16	193	79	713	122
14	Packaged Meals	330	10	750	7	33	16	90	40
15	Grain, Pasta, Bread	1540	23	3240	36	0	52	207	288
16	Grain, Pasta, Bread	1360	10	138	18	0	40	96	267
17	Grain, Pasta, Bread	960	16	1788	20	0	29	150	176
18	Canned Food, Sauce	600	0	7564	100	0	0	0	150
19	Canned Food, Sauce	225	5	1869	5	2	8	50	36
20	Cereal, Breakfast	1760	20	2450	130	0	45	200	372
21	Cereal, Breakfast	1080	21	720	54	0	21	189	180
22	Other Dairy	760	48	600	16	75	16	420	25
23	Meat, Protein	480	21	920	3	100	38	182	7
24	Meat, Protein	3077	1333	851	0	1227	357	11995	0
25	Meat, Protein	1040	80	3700	8	240	54	724	14
26	Condiments, Spices	540	0	1120	23	0	0	2	64
27	Condiments, Spices	1040	72	5250	23	0	0	657	32
28	Snacks, Candy	433	11	180	28	0	5	100	72
29	Snacks, Candy	1727	80	1950	18	0	27	720	216
30	Milk	2052	80	2056	192	320	129	720	201
31	Milk	874	24	1000	96	101	64	209	98
32	Baking Goods	11520	1344	0	0	0	0	11520	0
33	Baking Goods	1800	23	937	100	0	12	210	308

Source: Nielsen Homescan data combined with Gladson and FoodEssentials food label information. Tabulations include only those in the 52 identified Nielsen markets.

Table 3: Mean Own Price Elasticities of Demand and Expenditure Elasticities

Group	Aggregate Category	Own Price Elasticity	Standard Error	Expenditure Elasticity	Standard Error
1	Fruits & Vegetables	-1.128	0.004	1.425	0.014
2	Fruits & Vegetables	-0.830	0.004	1.191	0.015
3	Fruits & Vegetables	-1.379	0.004	1.491	0.012
4	Cold Beverages	-1.835	0.004	3.096	0.032
5	Cold Beverages	-1.895	0.004	2.151	0.018
6	Cold Beverages	-1.812	0.004	1.719	0.012
7	Cold Beverages	-2.055	0.006	1.902	0.018
8	Warm Beverages	-1.521	0.014	1.210	0.010
9	Warm Beverages	-0.518	0.005	0.322	0.017
10	Warm Beverages	-0.665	0.010	0.564	0.017
11	Soda	-2.260	0.006	4.485	0.044
12	Soda	-2.197	0.005	4.122	0.041
13	Packaged Meals	-0.332	0.004	0.582	0.008
14	Packaged Meals	0.120	0.005	-0.756	0.018
15	Grain, Pasta, Bread	-0.845	0.004	1.363	0.010
16	Grain, Pasta, Bread	-0.647	0.006	0.849	0.013
17	Grain, Pasta, Bread	-0.292	0.004	0.716	0.010
18	Canned Food, Sauce	-1.083	0.004	1.158	0.009
19	Canned Food, Sauce	-0.907	0.003	0.953	0.011
20	Cereal, Breakfast	-0.244	0.004	0.883	0.011
21	Cereal, Breakfast	0.022	0.005	0.135	0.013
22	Other Dairy	-0.731	0.003	0.728	0.008
23	Meat, Protein	0.011	0.005	0.155	0.015
24	Meat, Protein	-0.696	0.005	0.846	0.006
25	Meat, Protein	-0.253	0.004	0.692	0.011
26	Condiments, Spices	-0.431	0.002	0.269	0.013
27	Condiments, Spices	-0.669	0.002	0.877	0.010
28	Snacks, Candy	-0.270	0.004	-0.876	0.015
29	Snacks, Candy	-0.295	0.004	0.404	0.009
30	Milk	-1.793	0.003	2.296	0.015
31	Milk	-1.972	0.004	1.693	0.017
32	Baking Goods	-1.631	0.004	1.253	0.006
33	Baking Goods	-0.575	0.003	0.634	0.012

Source: Authors' calculations from 2002-2007 Nielsen Homescan data. The sample includes only consumers in the 52 identified Nielsen markets. The table shows average elasticities of demand and standard errors that are calculated from the QAIDS model parameter estimates.

Table 4: Own and Cross Price Elasticities by Product Categories and Within Categories Groups.

Aggregate		Within Category Group				Aggregate		Within Category Group		
Category	Group	1	2	3	4	Category	Group	1	2	3
Fruits & Vegetables	1	-1.128	-0.149	-0.058		Canned Food, Sauce	18	-1.083	0.024	
Fruits & Vegetables	2	0.023	-0.830	-0.025		Canned Food, Sauce	19	0.064	-0.907	
Fruits & Vegetables	3	-0.035	-0.105	-1.379						
						Cereal, Breakfast	20	-0.244	-0.017	
Cold Beverages	4	-1.835	0.037	0.027	0.019	Cereal, Breakfast	21	0.017	0.022	
Cold Beverages	5	0.048	-1.895	0.031	0.031					
Cold Beverages	6	0.079	0.046	-1.812	0.032	Other Dairy	22	-0.731		
Cold Beverages	7	-0.035	0.023	0.027	-2.055					
						Meat, Protein	23	0.011	0.079	-0.149
Warm Beverages	8	-1.521	-0.057	-0.092		Meat, Protein	24	0.175	-0.696	0.044
Warm Beverages	9	0.019	-0.518	0.010		Meat, Protein	25	0.028	0.019	-0.253
Warm Beverages	10	0.043	-0.097	-0.665						
						Condiments, Spices	26	-0.431	0.073	
Soda	11	-2.260	0.171			Condiments, Spices	27	0.030	-0.669	
Soda	12	0.201	-2.197							
						Snacks, Candy	28	-0.270	-0.060	
Packaged Meals	13	-0.332	0.031			Snacks, Candy	29	0.080	-0.295	
Packaged Meals	14	-0.020	0.120							
						Milk	30	-1.793	0.255	
Grain, Pasta, Bread	15	-0.845	-0.001	0.110		Milk	31	0.107	-1.972	
Grain, Pasta, Bread	16	-0.019	-0.647	0.055						
Grain, Pasta, Bread	17	0.111	0.094	-0.292		Baking Goods	32	-1.631	-0.268	
						Baking Goods	33	-0.003	-0.575	

Source: Authors' calculations from 2002-2007 Nielsen Homescan data. The sample includes only consumers in the 52 identified Nielsen markets. The table shows average own and cross-price elasticities of demand that are calculated from the QAIDS model parameter estimates. The table reports the diagonal matrix elements of the full elasticity matrix corresponding to the cross price elasticities of a given product group with respect to other product groups in the same aggregate product category.

Table 5: Tax Simulations – Product Taxes on Soda and Sugar-Sweetened Beverages

	Panel A: 20% Soda Tax (Indirect Utility Cost = 2.5%)										
Category	% Δ Expend.	% Δ Share	% Δ Price	Calories	Calories from Fat	Total Fat	Carbo-hydrates	Sugar	Pro-tein	Chole-sterol	Salt
Fruits & Vegetables	-1.15	1.13	0.00	-41.6	0.0	0.0	-9.4	-4.3	-0.8	0.0	-10.2
Cold Beverages	-0.72	0.95	0.00	-38.4	0.0	0.0	-11.2	-9.5	-0.1	0.0	-29.9
Warm Beverages	-1.69	-1.47	0.00	-270.5	-53.6	-3.2	-36.5	-25.9	-0.3	0.0	-29.2
Soda	-3.92	-3.43	20.00	-2231.1	0.0	0.0	-634.6	-674.7	0.0	-66.5	-869.3
Packaged Meals	-2.98	-1.84	0.00	-180.3	-66.9	-7.3	-17.6	-3.0	-9.2	-22.4	-472.0
Grain, Pasta, Bread	-1.09	0.81	0.00	-154.4	-27.2	-2.9	-28.4	-3.6	-4.7	0.0	-331.0
Canned Food, Sauce	-1.02	-0.15	0.00	-28.7	-3.3	-0.4	-5.3	-2.1	-0.6	-0.2	-336.7
Cereal, Breakfast	-2.92	-1.36	0.00	-330.9	-37.6	-4.1	-64.0	-20.5	-7.5	0.0	-399.4
Other Dairy	-0.94	0.53	0.00	-76.0	-35.7	-4.0	-5.0	-3.3	-2.5	-7.4	-59.1
Meat, Protein	-1.36	-1.39	0.00	-902.7	-3019.4	-335.7	-3.4	-1.6	-94.3	-301.6	-896.4
Condiments, Spices	-0.99	0.48	0.00	-43.8	-15.4	-1.7	-3.1	-1.5	0.0	0.0	-179.6
Snacks, Candy	-0.88	0.89	0.00	-92.3	-34.9	-3.9	-11.9	-2.5	-1.5	0.0	-93.5
Milk	-0.26	1.41	0.00	-12.4	-2.7	-0.3	-1.5	-1.1	-0.9	-1.6	-11.9
Baking Goods	-1.61	0.24	0.00	-376.6	-213.2	-24.4	-29.3	-10.1	-1.2	0.0	-81.3
Total	1.66			-4779.8	-3510.0	-387.9	-861.3	-763.6	-123.6	-399.7	-3799.5
Percent				-4.84	-7.37	-7.31	-5.67	-10.35	-3.70	-3.54	-2.03
	Panel B: 20% Tax on SSBs (Indirect Utility Cost = 2.0%)										
Category	% Δ Expend.	% Δ Share	% Δ Price	Calories	Calories from Fat	Total Fat	Carbo-hydrates	Sugar	Pro-tein	Chole-sterol	Salt
Fruits & Vegetables	-0.49	0.92	0.00	-17.6	0.0	0.0	-4.0	-1.9	-0.3	0.0	-1.4
Cold Beverages	-0.76	0.03	14.90	-1602.3	0.0	0.0	-422.7	-368.5	-2.6	0.0	-1224.6
Warm Beverages	-0.57	-1.22	20.00	-3258.5	-652.9	-40.1	-442.4	-313.2	-4.1	0.0	-365.0
Soda	-7.26	-7.60	8.34	-2832.2	0.0	0.0	-797.5	-750.5	0.0	-73.3	-790.7
Packaged Meals	-0.62	-0.33	0.00	-39.8	-15.1	-1.7	-3.8	-0.6	-2.0	-5.0	-104.9
Grain, Pasta, Bread	-0.33	0.71	0.00	-44.4	-8.0	-0.9	-8.2	-1.1	-1.4	0.0	-105.0
Canned Food, Sauce	-0.36	-0.36	0.00	-10.7	-1.1	-0.1	-2.0	-0.9	-0.2	-0.1	-131.4
Cereal, Breakfast	-0.70	0.02	0.00	-78.6	-8.6	-0.9	-15.2	-4.9	-1.8	0.0	-97.6
Other Dairy	-0.23	0.37	0.00	-18.8	-8.8	-1.0	-1.2	-0.8	-0.6	-1.8	-14.6
Meat, Protein	-0.28	0.11	0.00	-239.6	-845.6	-94.0	-0.8	-0.4	-25.4	-81.5	-212.9
Condiments, Spices	-0.21	0.39	0.00	-10.1	-4.3	-0.5	-0.6	-0.3	0.0	0.0	-45.7
Snacks, Candy	-0.13	0.76	0.00	-16.4	-6.4	-0.7	-2.1	-0.4	-0.3	0.0	-17.5
Milk	-0.06	0.72	0.00	6.6	2.9	0.3	0.6	0.8	0.4	1.1	8.4
Baking Goods	-0.51	0.47	0.00	-174.5	-121.5	-14.0	-8.4	-2.9	-0.3	0.0	-23.2
Total	1.66			-8337.0	-1669.3	-153.4	-1708.2	-1445.4	-38.7	-160.6	-3126.2
Percent				-8.44	-3.50	-2.89	-11.25	-19.59	-1.16	-1.42	-1.67

Source: Nielsen Homescan data. Indirect utility costs are calculated using equation (1). We calculate V for each household under the old ar
new price regimes; percent change in average indirect utility over all households and time periods are shown.

Table 6: Tax Simulations – Product Taxes on Packaged Meals and Snacks

	Panel A: 20% Packaged Meal Tax (Indirect Utility Cost = 1.4%)										
Category	% Δ Expend.	% Δ Share	% Δ Price	Calories	Calories from Fat	Total Fat	Carbo-hydrates	Sugar	Pro-tein	Chole-sterol	Salt
Fruits & Vegetables	3.61	2.51	0.00	171.6	0.0	0.0	38.1	15.8	3.7	0.0	412.0
Cold Beverages	1.61	-0.09	0.00	122.8	0.0	0.0	34.2	29.2	0.2	0.0	98.8
Warm Beverages	0.56	-2.54	0.00	214.1	40.0	1.8	28.2	18.7	0.1	0.0	10.3
Soda	2.74	-0.14	0.00	212.2	0.0	0.0	60.6	66.9	0.0	6.6	90.8
Packaged Meals	7.96	5.62	20.00	-897.8	-347.6	-38.0	-82.3	-13.4	-45.1	-113.1	-2381.7
Grain, Pasta, Bread	1.87	0.41	0.00	363.9	48.0	5.1	66.3	7.3	10.3	0.0	430.3
Canned Food, Sauce	0.25	-2.21	0.00	10.9	0.4	0.0	2.2	1.2	0.1	0.0	155.7
Cereal, Breakfast	2.10	0.32	0.00	220.2	26.5	2.9	42.3	13.4	5.0	0.0	252.5
Other Dairy	-0.42	-2.27	0.00	-33.7	-15.9	-1.8	-2.2	-1.5	-1.1	-3.3	-26.2
Meat, Protein	0.76	-0.46	0.00	308.4	1058.6	117.7	1.1	0.5	33.6	105.4	248.3
Condiments, Spices	-0.62	-2.47	0.00	-17.9	4.4	0.5	-2.7	-1.1	0.0	0.0	-8.9
Snacks, Candy	-0.11	-1.68	0.00	-37.0	-15.7	-1.8	-4.5	-0.2	-0.7	0.0	-46.4
Milk	2.38	0.66	0.00	222.3	65.5	7.3	25.1	22.6	14.9	30.4	237.8
Baking Goods	1.80	0.29	0.00	1220.7	1022.3	118.3	18.8	6.5	0.8	0.0	52.2
Total	-1.70			2080.5	1886.5	212.2	225.1	166.0	21.9	26.0	-474.4
Percent				2.11	3.96	4.00	1.48	2.25	0.65	0.23	-0.25
	Panel B: 20% Snack, Candy Tax (Indirect Utility Cost = 2.0%)										
Category	% Δ Expend.	% Δ Share	% Δ Price	Calories	Total Calories from Fat	Fat	Carbo-hydrates	Sugar	Pro-tein	Chole-sterol	Salt
Fruits & Vegetables	0.79	0.67	0.00	69.9	0.0	0.0	16.0	7.0	1.2	0.0	177.9
Cold Beverages	1.00	0.25	0.00	68.4	0.0	0.0	18.7	16.2	0.2	0.0	34.2
Warm Beverages	0.41	-1.76	0.00	278.2	51.5	2.2	36.5	23.7	0.1	0.0	9.6
Soda	0.34	-1.55	0.00	7.7	0.0	0.0	2.4	4.3	0.0	0.4	8.8
Packaged Meals	-0.28	-1.52	0.00	22.7	12.2	1.3	0.9	0.0	1.0	3.3	67.7
Grain, Pasta, Bread	-0.24	-0.73	0.00	-11.6	-5.3	-0.6	-2.4	-0.6	-0.6	0.0	-110.8
Canned Food, Sauce	0.37	-1.16	0.00	10.3	1.2	0.1	1.9	0.7	0.2	0.1	120.5
Cereal, Breakfast	0.39	-0.42	0.00	42.8	4.8	0.5	8.3	2.7	1.0	0.0	52.0
Other Dairy	-1.00	-1.92	0.00	-81.0	-38.1	-4.2	-5.4	-3.5	-2.7	-7.9	-63.0
Meat, Protein	-0.95	-1.31	0.00	115.2	789.3	87.7	-1.1	-0.4	14.6	50.1	-73.2
Condiments, Spices	-0.64	-1.56	0.00	-23.2	-2.5	-0.3	-2.4	-1.1	0.0	0.0	-61.1
Snacks, Candy	4.71	4.04	20.00	-2149.3	-847.0	-94.7	-271.6	-42.9	-35.8	-0.2	-2359.1
Milk	1.28	0.52	0.00	104.3	29.5	3.3	11.9	10.4	7.0	14.1	109.8
Baking Goods	0.11	-0.44	0.00	351.0	330.6	38.4	-3.0	-1.0	-0.1	0.0	-8.4
Total	-0.74			-1194.5	326.3	33.6	-189.5	15.6	-14.0	60.0	-2095.1
Percent				-1.21	0.68	0.63	-1.25	0.21	-0.42	0.53	-1.12

Source: Nielsen Homescan data. Indirect utility costs are calculated using equation (1). We calculate V for each household under the old ar new price regimes; percent change in average indirect utility over all households and time periods are shown.

Table 7: Tax Simulations – Nutrient Specific Taxes

Category	% Δ Expend.	% Δ Share	% Δ Price	Calories	Calories from Fat	Total Fat	Carbo-hydrates	Sugar	Pro-tein	Chole-sterol	Salt
Panel A: 20% Fat Tax (Indirect Utility Cost = 2.0%)											
Fruits & Vegetables	-0.70	-0.29	8.37	-284.9	0.0	0.0	-61.1	-24.4	-7.4	0.0	-521.9
Cold Beverages	1.76	1.56	0.10	132.5	0.0	0.0	36.7	31.4	0.2	0.0	117.9
Warm Beverages	0.34	-1.30	4.90	-1921.4	-382.4	-20.9	-260.2	-173.4	-1.8	0.0	-138.6
Soda	3.16	1.77	0.00	207.7	0.0	0.0	59.5	69.0	0.0	6.8	99.2
Packaged Meals	-0.23	-0.93	6.59	-581.3	-233.0	-25.4	-50.5	-7.9	-28.8	-74.2	-1559.4
Grain, Pasta, Bread	0.00	0.04	12.14	-2061.9	-308.3	-32.9	-378.2	-44.1	-60.7	0.0	-3315.2
Canned Food, Sauce	0.37	-0.63	2.79	-68.2	-7.7	-0.8	-12.7	-5.0	-1.3	-0.5	-805.4
Cereal, Breakfast	-0.03	-0.29	6.72	-723.0	-87.4	-9.6	-138.9	-44.0	-16.5	0.0	-826.4
Other Dairy	-0.67	-1.06	18.76	-1561.6	-734.0	-81.5	-103.1	-67.2	-52.4	-152.3	-1214.0
Meat, Protein	-0.35	-0.45	16.07	-2903.5	-6691.1	-744.5	-21.4	-10.3	-266.5	-860.5	-4940.0
Condiments, Spices	0.56	0.17	14.93	-648.8	-248.8	-27.6	-43.3	-21.9	0.0	0.0	-2788.5
Snacks, Candy	-0.06	-0.16	11.77	-2375.8	-972.2	-108.9	-294.3	-30.8	-40.7	-0.1	-2800.2
Milk	0.42	0.20	11.91	-1046.8	-304.2	-33.9	-118.6	-105.8	-70.5	-142.7	-1114.0
Baking Goods	1.74	1.73	7.10	-5147.6	-4487.8	-519.8	-38.6	-13.3	-1.6	0.0	-107.0
Total	-0.20			-18984.4	-14456.9	-1605.9	-1424.5	-447.6	-548.0	-1223.5	-19913.5
Percent				-19.23	-30.35	-30.25	-9.38	-6.07	-16.38	-10.83	-10.63
Panel B: 20% Sugar Tax (Indirect Utility Cost = 2.6%)											
Fruits & Vegetables	-2.53	-2.55	7.84	-745.7	0.0	0.0	-174.1	-82.2	-10.9	0.0	-864.2
Cold Beverages	6.44	5.75	14.06	-1326.3	0.0	0.0	-349.5	-302.6	-1.5	0.0	-1319.5
Warm Beverages	1.24	-0.86	13.90	-2041.9	-462.7	-35.9	-294.2	-209.8	-4.4	0.0	-339.8
Soda	11.52	9.53	14.16	-606.2	0.0	0.0	-170.8	-162.0	0.0	-15.8	-173.2
Packaged Meals	-0.81	-1.94	1.84	-135.3	-48.2	-5.3	-13.9	-2.4	-7.0	-16.5	-349.9
Grain, Pasta, Bread	0.00	-0.40	7.98	-1479.5	-206.1	-21.9	-269.4	-30.1	-42.1	0.0	-1910.4
Canned Food, Sauce	1.34	-0.11	5.84	-220.2	-5.7	-0.6	-45.5	-25.8	-1.0	-0.4	-3248.8
Cereal, Breakfast	-0.09	-0.80	15.76	-1749.8	-196.1	-21.3	-338.7	-108.5	-39.9	0.0	-2133.1
Other Dairy	-2.43	-3.24	5.24	-620.4	-291.6	-32.4	-41.0	-26.7	-20.8	-60.5	-482.3
Meat, Protein	-1.25	-1.71	5.00	-317.8	936.4	103.7	-8.6	-3.7	-18.5	-45.8	-1321.9
Condiments, Spices	2.05	1.20	12.63	-426.2	-106.6	-11.8	-36.1	-16.8	0.0	0.0	-1488.1
Snacks, Candy	-0.86	-1.40	10.47	-1979.3	-800.2	-89.6	-246.8	-30.2	-33.6	-0.1	-2280.3
Milk	1.52	0.86	20.29	-1748.4	-511.0	-57.0	-197.7	-177.2	-117.6	-238.8	-1865.0
Baking Goods	6.57	6.09	13.18	-4905.4	-4018.5	-464.9	-96.2	-33.2	-4.0	0.0	-266.7
Total	-0.64			-18302.4	-5710.3	-636.9	-2282.5	-1211.1	-301.2	-378.0	-18043.3
Percent				-18.54	-11.99	-12.00	-15.03	-16.41	-9.00	-3.34	-9.63
Panel C: 20% Salt Tax (Indirect Utility Cost = 1.2%)											
Fruits & Vegetables	-1.83	-1.67	3.64	-294.5	0.0	0.0	-61.2	-18.4	-9.4	0.0	-1714.0
Cold Beverages	4.63	4.15	1.60	121.3	0.0	0.0	35.7	30.3	0.3	0.0	47.2
Warm Beverages	0.89	-1.02	1.75	3.9	-15.8	-3.6	-4.7	-5.2	-0.6	0.0	-43.1
Soda	8.29	6.55	1.19	400.2	0.0	0.0	115.1	137.3	0.0	13.6	204.0
Packaged Meals	-0.58	-1.54	6.64	-568.7	-224.9	-24.5	-50.5	-8.0	-28.4	-72.2	-1518.9
Grain, Pasta, Bread	-0.01	-0.23	6.16	-1186.7	-160.2	-17.0	-214.0	-23.1	-32.6	0.0	-1229.8
Canned Food, Sauce	0.96	-0.30	13.87	-379.4	-39.7	-4.2	-71.2	-29.4	-6.7	-2.8	-4591.1
Cereal, Breakfast	-0.07	-0.59	9.69	-1102.0	-116.3	-12.5	-214.5	-69.3	-25.0	0.0	-1405.7
Other Dairy	-1.75	-2.39	5.85	-613.2	-288.2	-32.0	-40.5	-26.4	-20.6	-59.8	-476.7
Meat, Protein	-0.90	-1.21	8.70	-1002.8	-164.0	-19.0	-14.4	-7.2	-61.3	-213.6	-3331.0
Condiments, Spices	1.48	0.82	9.56	-405.3	-197.1	-21.9	-21.4	-11.9	0.0	0.0	-1994.7
Snacks, Candy	-0.62	-0.98	7.17	-1755.8	-724.7	-81.2	-216.5	-19.9	-30.3	0.0	-2102.6
Milk	1.09	0.62	7.12	-562.6	-164.5	-18.3	-63.6	-57.0	-37.8	-76.9	-600.2
Baking Goods 4.66	4.38	6.77	-3149.9	-2793.2	-323.7	-12.8	-4.4	-0.5	0.0	-35.5	
Total	-0.46			-10495.5	-4888.7	-558.0	-834.5	-112.7	-252.9	-411.7	-18792.1
Percent				-10.63	-10.26	-10.51	-5.49	-1.53	-7.56	-3.64	-10.03

Source: Nielsen Homescan data. Indirect utility costs are calculated using equation (1). We calculate V for each household under the old ar new price regimes; percent change in average indirect utility over all households and time periods are shown.

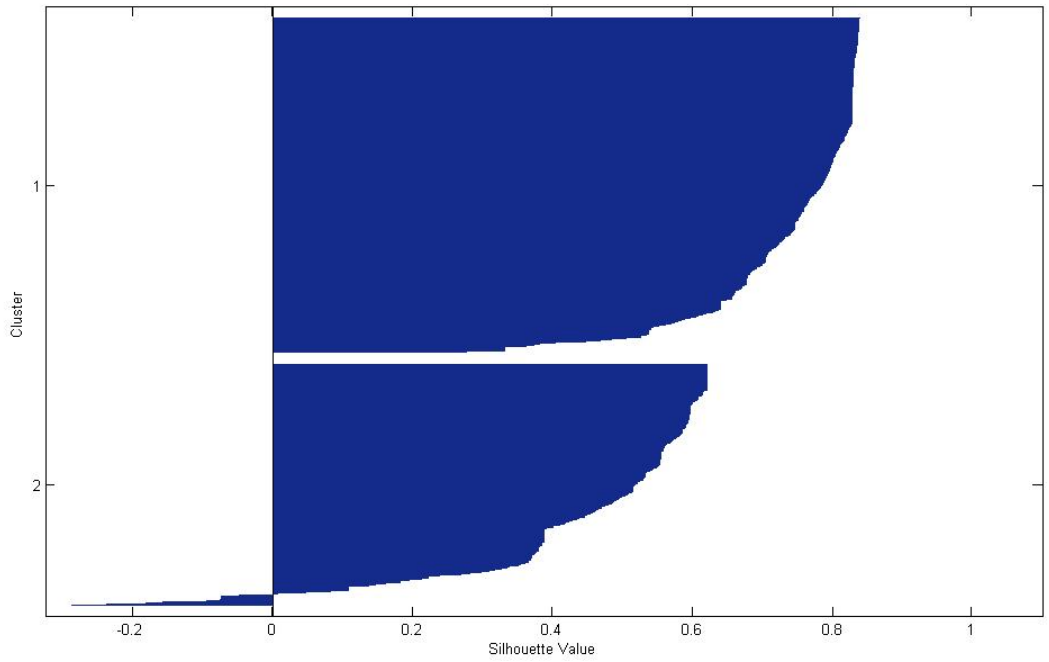

Figure 1: Silhouette values for soda products by UPC for $k = 2$ clusters.

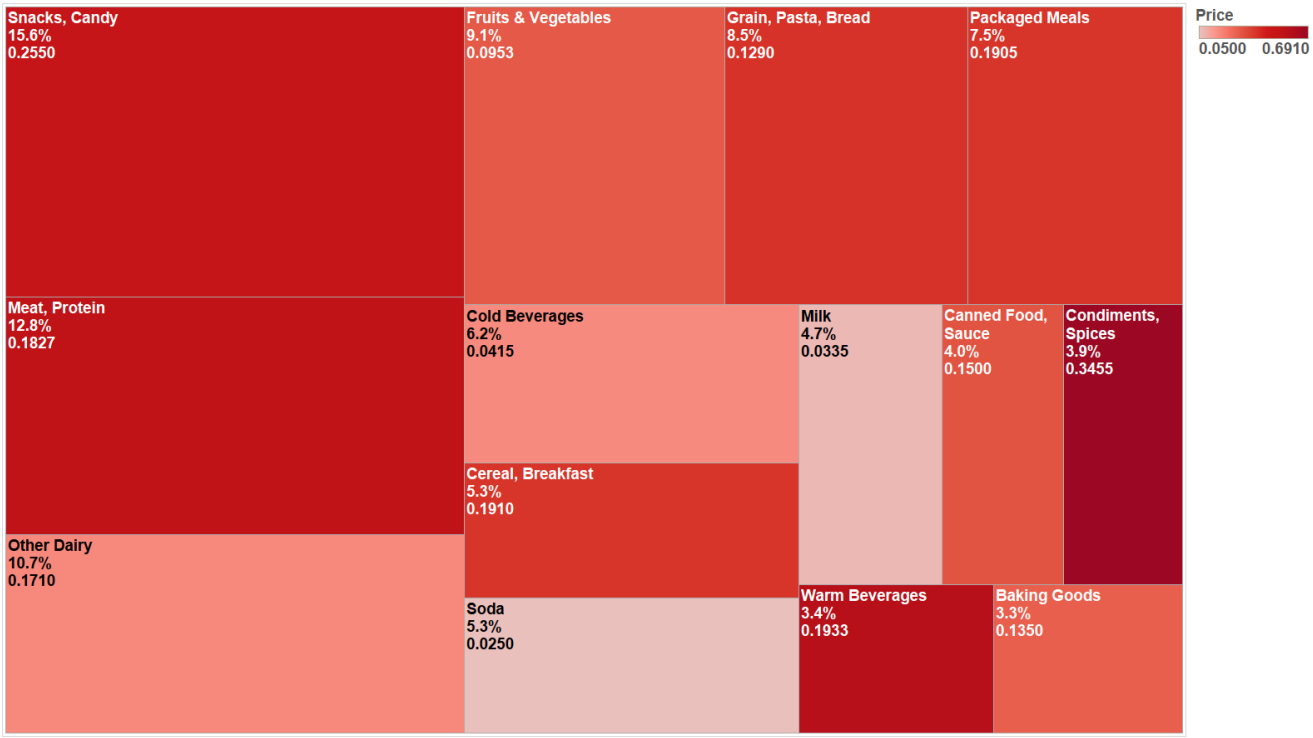

Figure 2: Graphical representation of food expenditures on the 14 major product categories in the sample. The size of the squares is proportional to the budget share of the corresponding product. The budget share is given in % under each product category name. The color shading of each rectangle corresponds to the price per ounce of products in each of the categories. The price per ounce in $ is also reported under the budget share for each category.

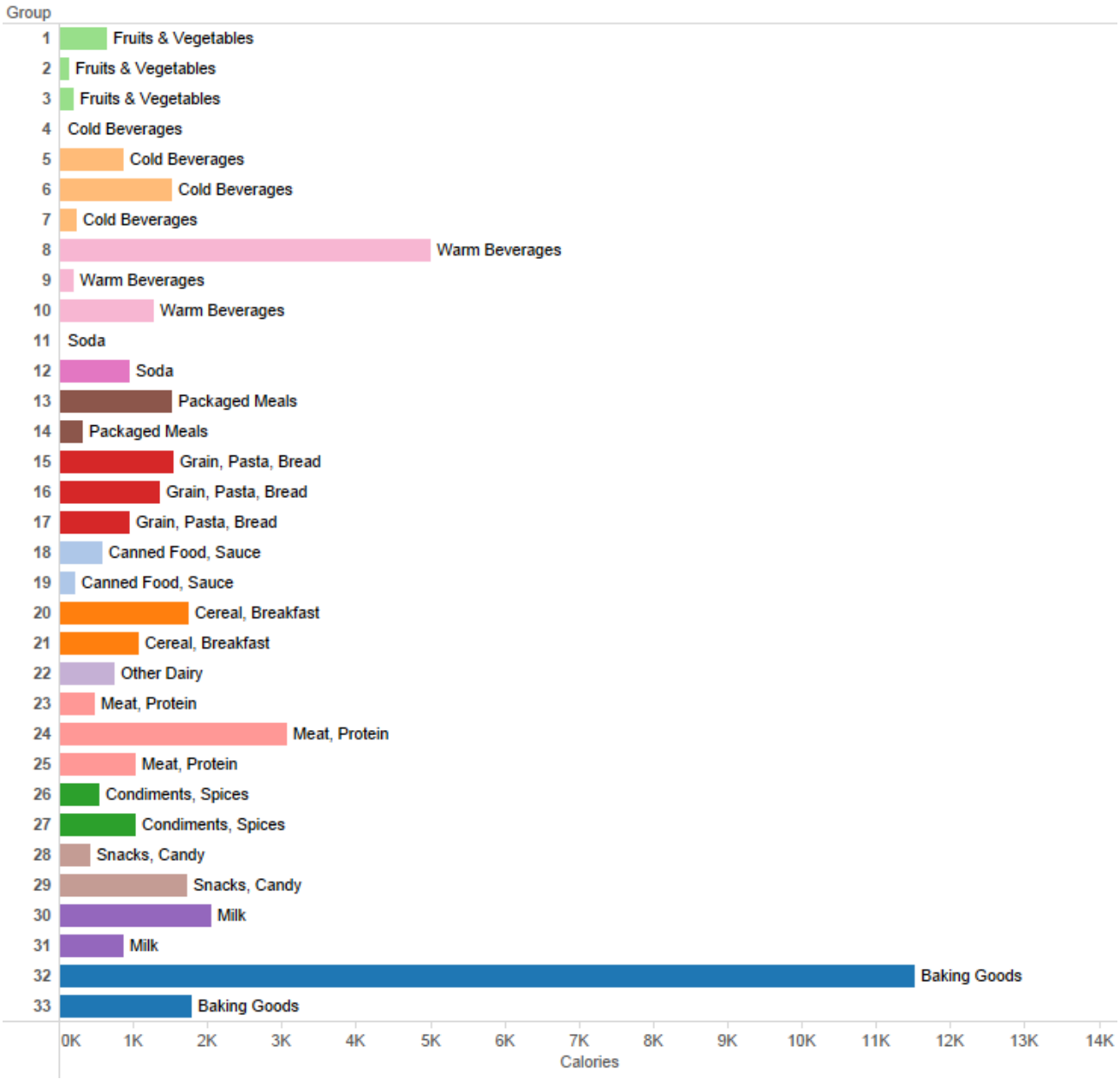

Figure 3: Number of calories per serving for each of the 33 product groups resulting from the partition of 14 major food categories into nutrition based groups. Colors and labels denote the 14 product categories. For example, groups 11 and 12 correspond to the 2 partitions of the soda category. The groups within each product category are distinguished by their nutritional composition and are not separately named.

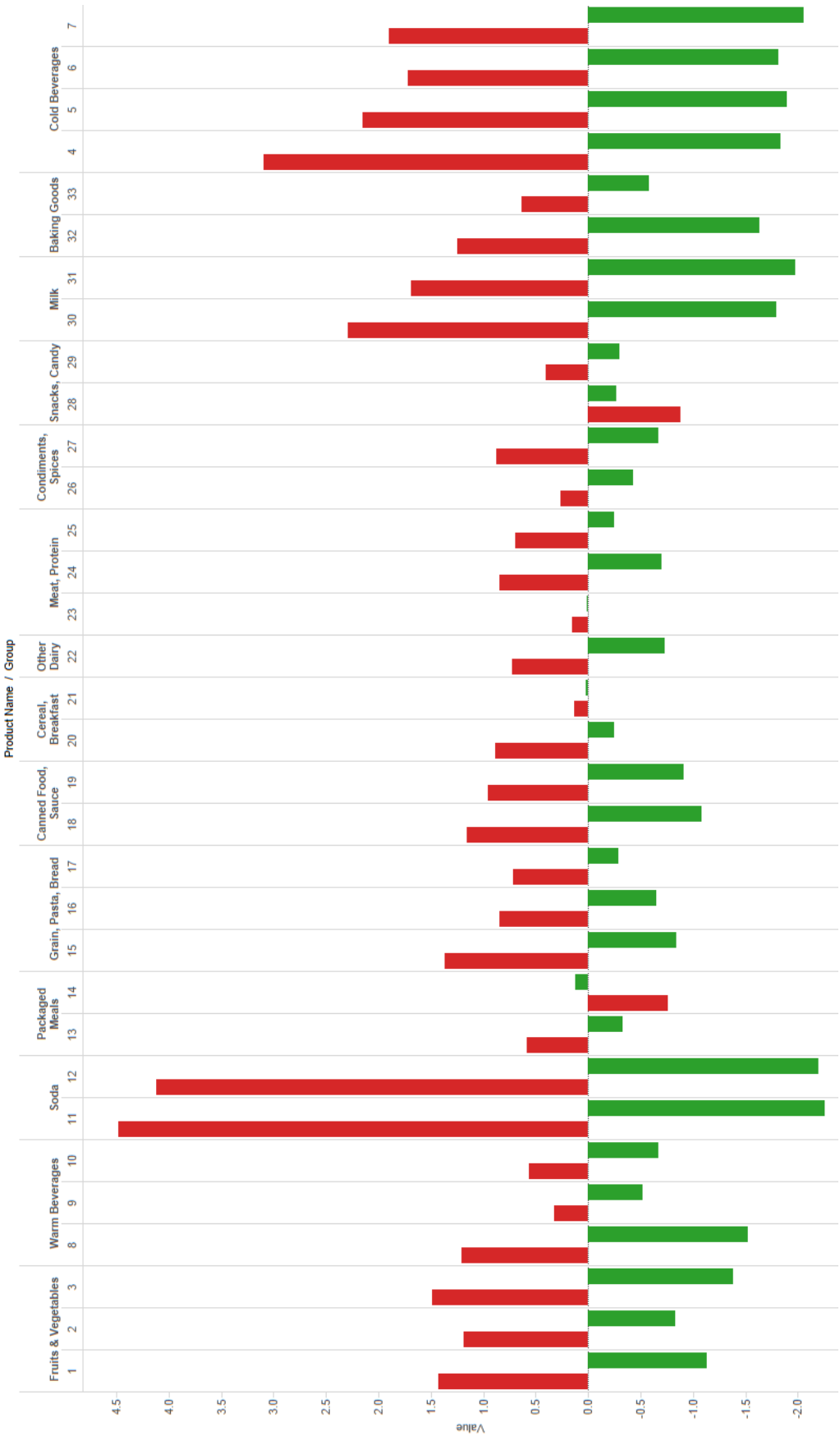

Figure 4: Own Price and Expenditure Elasticities by Product Groups and Aggregate Categories. For each group we report both elasticities. The top bar in each group corresponds to the own price elasticity.

Table A-1: Descriptive Statistics of Analysis Variables

Variable	Mean	SD
Household Income		
<$30,000	0.247	0.431
$30,000-$45,000	0.207	0.405
$45,000-$70,000	0.261	0.439
≥$70,000	0.285	0.451
Head Education		
Less Than HS	0.033	0.178
HS Graduate	0.254	0.436
Some College	0.320	0.466
BA+	0.393	0.488
Female Head Age		
<35	0.073	0.260
35-49	0.323	0.268
50-64	0.371	0.483
≥65	0.233	0.423
Male Head Age		
<35	0.062	0.242
35-49	0.325	0.469
50-64	0.366	0.481
≥65	0.246	0.431
Race		
White	0.811	0.391
Hispanic	0.062	0.240
Black	0.103	0.304
Asian	0.025	0.155
Kids Over 12?	0.153	0.360
Kids Under 12?	0.167	0.373
Married	0.579	0.494
Female Labor Supply		
Part Time	0.163	0.370
Full Time	0.424	0.494
Not Employed	0.368	0.482
Male Labor Supply		
Part Time	0.068	0.252
Full Time	0.614	0.487
Not Employed	0.318	0.466
Male Household Head	0.278	0.448
Female Household Head	0.107	0.309
Number of Household Members		
1	0.268	0.443
2	0.394	0.489
3	0.144	0.351
4	0.120	0.325
5	0.049	0.215
6+	0.025	0.157
Number of Transactions	89,627,027	
Number of Household-Years	162,195	
Number of Aggregate Obs.	21,145,678	

2002-2007 Nielsen Homescan data. Head Education refers to female household head if a female household head is present. In the cases in which no female household head is present, this variable refers to the male household head.

Table A-2: Transaction-nutrition Matching Statistics by Year

Year	Direct	Match 1	Match 2	Match 3	Match 4	Match 5	Unmatched
2002	68.04%	13.92%	11.15%	4.05%	1.33%	1.53%	4.49%
2003	68.92%	13.42%	10.60%	4.24%	1.24%	1.59%	4.34%
2004	64.02%	15.67%	12.32%	4.97%	1.27%	1.76%	2.28%
2005	64.79%	15.42%	11.56%	5.18%	1.26%	1.78%	2.19%
2006	65.77%	15.41%	10.70%	5.14%	1.18%	1.80%	3.05%
2007	64.97%	15.83%	10.85%	5.24%	1.29%	1.81%	2.33%

The table shows the percent of the transaction sample that is matched to the nutrition data under each of the matching sets.
Direct matches are products that are matched directly based on UPC or based on product name for "random weight" products such as fruits, vegetables and meats.
Match 1 = Assign Nielsen products the average per-serving nutrient characteristics among products with exact matches based on "product module," size, type, brand, product, flavor, variety and formula.
Match 2 = Assign Nielsen products the average per-serving nutrient characteristics among products with exact matches based on "product module," size, type, and product.
Match 3 = Assign Nielsen products the average per-serving nutrient characteristics among products with exact matches based on "product module" and size.
Match 4 = Assign Nielsen products the average per-serving nutrient characteristics among products with exact matches based on "product module," type, brand, flavor, product, formula, and variety.
Match 5 = Assign Nielsen products the average per-serving nutrient characteristics among products with exact matches based on "product module," type, product, and formula.
Unmatched = Percent of transactions from raw data within the 52 markets that cannot be matched to nutritional information.

Table A-3: First Stage Log Price Estimates

Ind. Var.	Dependent Variable Group										
Group	1	2	3	4	5	6	7	8	9	10	11
1	0.999	0.006	0.004	0.001	0.002	0.000	0.000	0.000	0.002	0.001	0.002
2	0.008	0.998	0.009	0.003	0.005	0.002	0.003	0.000	0.007	0.004	0.004
3	0.006	0.007	0.994	0.002	0.002	0.001	0.002	0.001	0.003	0.003	0.001
4	-0.001	-0.001	0.000	1.015	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001
5	0.001	0.002	0.002	0.002	1.013	0.001	0.003	0.000	0.002	0.002	0.002
6	0.001	0.001	0.003	0.001	0.001	1.014	0.001	0.001	0.001	0.002	0.000
7	0.000	0.000	0.001	-0.001	0.001	0.000	1.008	0.000	0.001	0.001	0.000
8	-0.003	-0.006	-0.004	-0.004	-0.005	-0.006	-0.005	1.001	-0.004	-0.001	-0.004
9	-0.001	-0.002	-0.001	0.000	0.000	0.000	0.000	0.000	0.996	0.000	0.000
10	0.001	0.000	0.000	-0.001	0.000	0.000	0.001	0.000	0.004	0.994	0.000
11	0.001	0.000	0.002	0.001	0.001	0.000	0.001	0.000	0.002	0.001	1.015
12	-0.001	-0.001	-0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.001	0.002
13	-0.001	-0.002	-0.003	-0.002	-0.002	-0.002	-0.002	0.000	-0.003	-0.001	-0.002
14	0.009	0.011	0.007	0.002	0.003	0.002	0.001	0.000	0.004	0.002	0.003
15	0.007	0.009	0.005	0.001	0.003	0.002	0.002	0.000	0.005	0.002	0.003
16	0.001	0.001	0.003	0.000	0.001	0.000	0.001	0.000	0.000	0.001	0.001
17	0.006	0.007	0.003	0.000	0.002	0.001	0.001	0.000	0.002	0.001	0.000
18	0.000	-0.001	0.001	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	-0.001
19	0.002	0.003	0.003	0.000	0.001	0.000	0.000	0.000	0.002	0.001	0.000
20	0.002	0.002	-0.001	-0.001	-0.002	-0.001	-0.001	0.000	-0.001	-0.001	-0.001
21	0.000	0.000	-0.002	0.000	-0.001	0.000	-0.002	0.000	-0.001	-0.002	0.000
22	0.011	0.017	0.010	0.003	0.008	0.003	0.004	0.000	0.009	0.004	0.005
23	0.000	0.005	0.004	0.001	0.002	0.001	0.001	0.000	0.002	0.001	0.002
24	-0.009	-0.008	-0.007	-0.004	-0.006	-0.004	-0.004	0.000	-0.008	-0.003	-0.004
25	0.001	0.003	0.002	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000
26	0.002	0.003	0.002	0.000	0.001	0.000	0.001	0.000	0.002	0.001	0.000
27	0.002	0.003	0.002	0.000	0.001	0.000	0.000	0.000	0.001	0.000	0.001
28	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000	0.000	0.000	0.000
29	0.016	0.020	0.011	0.003	0.008	0.004	0.004	0.001	0.009	0.004	0.006
30	0.003	0.004	0.005	0.002	0.004	0.001	0.002	0.000	0.003	0.002	0.002
31	-0.001	-0.001	-0.001	0.001	0.000	0.000	-0.001	0.000	0.001	0.001	0.001
32	0.000	-0.001	0.001	0.000	-0.001	0.000	0.001	0.001	0.001	0.002	-0.001
33	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.001

Source: Nielsen Homescan data. Each column represents a separate regression with the log price of the product category as the dependent variable. All estimates include month, year and market fixed effects. Standard errors are available upon request.

Table A-4: First Stage Log Price Estimates (continued)

Ind. Var.	Dependent Variable Group											
Group	12	13	14	15	16	17	18	19	20	21	22	23
1	0.001	0.001	0.003	0.003	0.002	0.003	0.001	0.003	0.003	0.002	0.003	0.004
2	0.003	0.003	0.008	0.010	0.008	0.011	0.005	0.011	0.007	0.006	0.010	0.010
3	0.001	0.002	0.004	0.006	0.005	0.006	0.003	0.006	0.003	0.003	0.003	0.005
4	0.000	0.000	-0.001	-0.001	0.000	-0.001	0.001	0.000	-0.001	0.000	-0.001	-0.001
5	0.002	0.001	0.003	0.002	0.003	0.002	0.001	0.003	0.003	0.003	0.002	0.001
6	0.001	0.001	0.000	0.003	0.002	0.002	0.003	0.002	0.001	0.001	0.002	0.002
7	0.000	0.001	0.001	0.000	0.001	0.000	0.001	0.001	0.001	0.001	0.000	-0.001
8	-0.004	0.002	-0.004	-0.007	-0.005	-0.006	-0.002	-0.005	-0.002	-0.001	-0.005	-0.008
9	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	-0.001	0.000	0.000	-0.001	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.001
11	0.002	0.002	0.002	0.001	0.001	0.000	0.002	0.002	0.002	0.002	0.001	0.001
12	1.013	0.000	-0.001	0.000	0.001	0.000	0.000	-0.001	-0.002	-0.001	0.000	-0.001
13	-0.002	0.984	0.000	-0.002	-0.003	-0.001	-0.003	-0.003	-0.002	-0.001	-0.003	-0.004
14	0.003	0.003	0.986	0.011	0.007	0.011	0.004	0.008	0.005	0.004	0.009	0.013
15	0.003	0.003	0.006	0.978	0.004	0.006	0.003	0.006	0.005	0.004	0.007	0.008
16	0.000	0.001	0.003	0.000	0.994	0.002	0.001	0.003	0.002	0.002	0.001	0.002
17	0.001	0.001	0.005	0.006	0.004	0.975	0.002	0.004	0.005	0.003	0.005	0.006
18	-0.001	0.001	0.000	0.000	0.000	0.000	0.998	0.002	0.000	0.000	0.000	0.001
19	0.001	0.001	0.002	0.003	0.002	0.004	0.001	0.991	0.003	0.002	0.003	0.003
20	-0.002	0.001	0.000	0.001	-0.001	0.001	-0.001	-0.002	0.983	0.001	-0.001	0.000
21	0.000	0.000	0.000	-0.001	-0.001	0.000	-0.001	-0.002	0.000	0.983	-0.001	0.001
22	0.006	0.003	0.012	0.017	0.011	0.018	0.005	0.013	0.012	0.008	0.997	0.021
23	0.002	0.001	0.005	0.007	0.005	0.007	0.003	0.006	0.004	0.003	0.007	0.979
24	-0.005	-0.001	-0.007	-0.007	-0.007	-0.008	-0.006	-0.009	-0.005	-0.004	-0.008	-0.007
25	0.001	0.002	0.002	0.006	0.001	0.005	0.001	0.002	0.003	0.001	0.003	0.007
26	0.001	0.001	0.001	0.003	0.002	0.003	0.001	0.002	0.002	0.002	0.003	0.003
27	0.001	0.001	0.001	0.003	0.002	0.002	0.002	0.002	0.001	0.001	0.002	0.002
28	0.001	-0.001	0.000	0.000	0.002	0.001	0.000	0.000	-0.001	0.000	0.002	0.002
29	0.005	0.005	0.014	0.021	0.013	0.020	0.006	0.013	0.013	0.010	0.017	0.019
30	0.002	0.002	0.006	0.005	0.006	0.006	0.004	0.006	0.007	0.006	0.006	0.006
31	0.000	0.000	0.001	-0.002	-0.002	-0.002	-0.001	-0.001	0.000	0.001	-0.002	-0.004
32	-0.001	0.001	-0.001	-0.001	0.001	0.000	0.003	0.000	-0.001	0.000	-0.001	0.001
33	0.000	0.000	0.000	0.001	0.001	0.001	0.000	0.001	0.000	0.000	0.001	0.002

Source: Nielsen Homescan data. Each column represents a separate regression with the log price of the product category as the dependent variable. All estimates include month, year and market fixed effects. Standard errors are available upon request.

Table A-5: First Stage Log Price Estimates (continued)

Ind. Var.	Dependent Variable Group									
Group	24	25	26	27	28	29	30	31	32	33
1	0.003	0.003	0.002	0.003	0.003	0.003	0.002	0.001	0.000	0.002
2	0.000	0.007	0.012	0.009	0.010	0.008	0.002	0.005	0.002	0.009
3	0.003	0.005	0.007	0.006	0.004	0.004	0.000	0.002	0.001	0.005
4	0.000	0.000	0.000	0.000	-0.001	-0.001	0.000	-0.001	0.000	0.000
5	0.001	0.001	0.003	0.003	0.003	0.002	0.001	0.002	0.000	0.003
6	0.002	0.003	0.004	0.003	0.001	0.001	0.002	0.001	0.001	0.001
7	0.000	0.000	0.002	0.001	0.000	0.000	0.001	0.000	0.000	0.001
8	0.005	-0.006	-0.005	-0.004	-0.005	-0.004	0.000	-0.005	-0.001	-0.003
9	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000	0.000
10	0.001	0.001	0.002	0.000	0.001	0.000	0.000	-0.001	0.000	0.000
11	0.001	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.000	0.001
12	0.001	0.000	0.001	0.001	0.000	-0.001	-0.001	0.000	0.000	0.000
13	-0.001	-0.001	-0.003	-0.002	-0.003	-0.002	-0.002	-0.002	-0.001	-0.003
14	0.006	0.009	0.009	0.007	0.008	0.008	0.003	0.003	0.001	0.005
15	0.002	0.008	0.007	0.007	0.005	0.007	0.002	0.001	0.001	0.005
16	0.001	0.001	0.003	0.002	0.003	0.002	0.001	0.000	0.000	0.003
17	0.002	0.006	0.006	0.004	0.005	0.005	0.001	0.001	0.000	0.003
18	0.001	0.002	0.004	0.003	0.000	0.000	-0.001	0.000	0.000	0.001
19	0.001	0.003	0.003	0.002	0.002	0.002	0.001	0.001	0.000	0.002
20	0.000	0.001	0.000	-0.001	0.000	0.000	0.001	-0.002	-0.001	-0.001
21	0.002	0.000	-0.001	-0.001	0.000	-0.001	0.000	-0.002	0.000	-0.002
22	0.007	0.016	0.016	0.012	0.017	0.016	0.006	0.007	0.002	0.010
23	0.001	0.008	0.007	0.004	0.007	0.005	0.002	0.003	0.001	0.004
24	0.980	-0.005	-0.008	-0.008	-0.007	-0.008	-0.002	-0.005	-0.001	-0.008
25	0.001	0.977	0.004	0.003	0.002	0.004	0.001	0.000	0.000	0.002
26	0.001	0.002	0.955	0.002	0.003	0.002	0.001	0.001	0.000	0.002
27	0.001	0.002	0.003	0.976	0.002	0.002	0.000	0.001	0.001	0.002
28	0.001	0.000	0.002	-0.001	0.986	0.004	0.000	0.000	0.000	0.001
29	0.007	0.018	0.017	0.013	0.020	0.995	0.004	0.006	0.002	0.010
30	0.002	0.003	0.005	0.004	0.005	0.005	1.003	0.008	0.001	0.005
31	-0.002	-0.003	-0.002	-0.001	-0.001	-0.002	-0.002	1.014	0.000	-0.001
32	0.004	0.001	0.003	0.003	-0.001	-0.001	-0.001	-0.001	1.003	0.001
33	0.001	0.001	0.003	0.002	0.001	0.001	0.000	0.000	0.000	0.985

Source: Nielsen Homescan data. Each column represents a separate regression with the log price of the product category as the dependent variable. All estimates include month, year and market fixed effects. Standard errors are available upon request.

Table A-6: First Stage Expenditure Estimates

Dep. Var: Ln(Expenditure)	
Independent Variable	Estimates
Ln(Market Income)	13.394**
	(0.203)
Ln(Market Income)^2	-0.541**
	(0.012)
Ln(Market HPI)	0.584**
	(0.101)
Ln(Market HPI)^2	0.178**
	(0.004)
Ln(Market Unemployment)	4.665**
	(0.075)
Ln(Market Unemployment)^2	0.020**
	(0.003)
Ln(Income)*Ln(HPI)	-0.239**
	(0.011)
Ln(Unemployment)*Ln(HPI)	-0.038
	(0.005)
Ln(Unemployment)*Ln(Income)	-0.427**
	(0.007)
F-Statistic of Instruments:	24.92

"Market Income" is MSA-by-year real income per capita, "HPI" is the MSA-by-quarter housing price index from the Federal Housing Finance Agency Repeat Sales Index, and "Market Unemployment" is the MSA-by-month unemployment rate. All estimates include market, month and year fixed effects as well as the demographic controls included in the demand model. Standard errors clustered at the Census tract level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level. The F-statistic shows the test for joint significance of the instruments.

Table A-7: Own and Cross-Price Elasticities of Demand (ϵ_{ij})

						i					
j	1	2	3	4	5	6	7	8	9	10	11
1	-1.128	-0.149	-0.058	0.046	0.049	0.012	0.041	0.001	0.016	0.016	0.033
2	0.023	-0.830	-0.025	0.059	0.036	0.008	0.061	-0.014	0.013	0.031	-0.014
3	-0.035	-0.105	-1.379	0.072	0.042	0.034	0.053	0.004	0.026	0.026	0.081
4	0.029	0.032	0.030	-1.835	0.037	0.027	0.019	0.016	0.005	0.000	0.071
5	0.004	0.023	0.037	0.048	-1.895	0.031	0.031	0.017	0.057	0.023	0.125
6	0.013	0.051	0.012	0.079	0.046	-1.812	0.032	0.017	0.034	-0.002	0.135
7	0.022	0.046	0.035	-0.035	0.023	0.027	-2.055	0.032	0.021	0.007	0.031
8	-0.082	-0.091	-0.013	-0.085	-0.033	0.139	0.146	-1.521	-0.057	-0.092	-0.125
9	-0.002	0.010	-0.004	-0.011	-0.021	-0.006	-0.016	0.019	-0.518	0.010	-0.048
10	0.000	-0.005	-0.002	-0.005	0.002	0.017	0.016	0.043	-0.097	-0.665	-0.015
11	0.091	0.076	0.049	0.083	0.113	0.057	0.087	0.029	0.054	0.037	-2.260
12	0.094	0.122	0.040	0.109	0.036	0.049	0.065	0.013	0.046	0.002	0.201
13	-0.056	-0.104	-0.003	-0.080	-0.063	0.030	-0.006	0.060	-0.010	-0.012	-0.153
14	-0.079	-0.021	-0.025	-0.035	0.009	-0.044	0.008	-0.016	-0.028	-0.013	-0.123
15	-0.030	-0.011	0.071	-0.015	0.013	-0.006	-0.004	-0.015	0.012	0.015	-0.088
16	-0.046	-0.119	-0.085	0.035	-0.015	-0.001	0.030	0.005	0.004	-0.009	0.061
17	-0.055	-0.052	0.041	-0.080	-0.005	-0.033	-0.043	-0.012	0.005	0.013	-0.170
18	0.031	0.044	-0.005	0.007	0.016	0.030	0.031	0.025	0.022	0.017	-0.037
19	-0.034	0.009	0.057	-0.044	-0.019	-0.008	-0.018	-0.003	0.005	0.002	-0.045
20	-0.065	-0.092	-0.061	-0.133	0.044	-0.011	-0.038	0.004	0.011	-0.002	-0.283
21	-0.042	-0.085	-0.057	0.033	0.018	-0.014	0.044	0.017	-0.021	-0.016	-0.054
22	-0.105	0.017	0.004	0.050	-0.012	-0.006	0.010	-0.029	-0.033	-0.013	0.009
23	-0.002	0.126	-0.058	-0.024	-0.020	-0.032	-0.013	-0.030	-0.048	-0.029	-0.130
24	0.010	0.032	-0.054	-0.176	-0.095	0.039	-0.009	0.208	0.004	-0.044	-0.288
25	-0.057	-0.022	-0.051	-0.127	-0.052	-0.016	-0.083	-0.014	-0.040	-0.018	-0.178
26	0.027	0.086	0.052	-0.007	0.011	0.002	0.003	0.002	-0.006	-0.002	-0.044
27	-0.001	0.017	0.036	-0.015	0.000	0.005	-0.004	0.005	-0.005	0.006	-0.032
28	-0.032	0.011	-0.065	0.024	-0.079	-0.046	0.016	-0.014	0.001	-0.022	0.066
29	-0.104	-0.103	-0.081	0.045	0.000	-0.007	0.025	-0.034	-0.060	-0.017	0.036
30	0.011	0.066	0.066	0.127	0.023	0.036	0.100	0.022	0.058	0.025	0.191
31	0.027	-0.021	0.032	0.009	-0.043	0.026	0.059	0.041	0.011	0.015	0.016
32	0.006	0.004	-0.033	0.074	0.047	0.059	0.075	0.055	-0.023	-0.017	0.042
33	-0.005	0.018	0.006	0.012	-0.023	-0.003	-0.006	-0.001	-0.011	0.007	0.004

Source: Authors' calculations from 2002-2007 Nielsen Homescan data. The sample includes only consumers in the 52 identified Nielsen markets. The table shows average own and cross-price elasticities of demand. The i good represents the quantity good and the j good is the price good.

Table A-8: Own and Cross-Price Elasticities of Demand (ϵ_{ij}) - continued

					i					
j	12	13	14	15	16	17	18	19	20	21
1	0.030	0.003	0.074	-0.018	-0.011	0.000	-0.020	-0.006	0.049	0.049
2	-0.093	0.010	0.159	-0.087	-0.030	-0.056	-0.020	0.009	-0.003	0.050
3	0.062	0.037	0.160	-0.007	-0.117	-0.031	-0.039	-0.114	0.030	0.011
4	0.050	0.004	-0.004	0.018	0.007	0.003	0.015	0.022	0.005	-0.017
5	0.070	0.037	0.004	0.003	-0.003	-0.033	0.041	-0.012	-0.031	-0.026
6	0.087	-0.039	0.021	0.004	0.011	0.013	-0.017	0.016	-0.016	0.007
7	0.050	-0.001	-0.018	0.036	0.010	0.011	0.026	0.016	0.000	-0.045
8	-0.049	-0.129	0.007	0.023	0.043	0.074	0.039	0.013	-0.033	-0.116
9	-0.053	-0.030	-0.058	-0.008	-0.008	-0.018	0.002	0.001	-0.003	-0.002
10	-0.028	-0.028	-0.046	-0.002	-0.011	-0.021	0.010	0.001	0.004	-0.025
11	0.171	0.010	-0.006	0.049	0.031	0.008	0.018	0.042	0.044	0.017
12	-2.197	-0.001	0.020	-0.017	-0.028	-0.050	0.003	0.015	0.049	0.073
13	-0.098	-0.332	0.031	-0.011	-0.038	-0.028	0.011	-0.034	-0.026	-0.057
14	-0.114	-0.020	0.120	-0.071	-0.023	-0.012	-0.069	0.015	-0.044	0.001
15	-0.033	0.012	0.068	-0.845	-0.001	0.110	0.021	0.118	-0.015	0.052
16	0.065	0.015	0.152	-0.019	-0.647	0.055	-0.062	-0.139	-0.016	0.004
17	-0.080	-0.002	0.013	0.111	0.094	-0.292	-0.005	0.090	-0.031	-0.001
18	-0.022	-0.041	-0.002	-0.017	0.008	-0.033	-1.083	0.024	-0.004	-0.001
19	-0.009	-0.008	-0.047	0.018	0.002	0.021	0.064	-0.907	-0.042	-0.029
20	-0.179	-0.039	0.031	-0.024	-0.026	-0.018	-0.034	-0.002	-0.244	-0.017
21	-0.120	-0.046	0.061	-0.057	-0.021	-0.007	-0.057	-0.022	0.017	0.022
22	0.000	-0.046	-0.085	-0.031	0.059	0.026	-0.001	0.054	-0.164	-0.090
23	-0.102	-0.069	-0.013	-0.067	-0.041	-0.028	-0.077	-0.073	-0.083	0.017
24	-0.177	-0.064	-0.082	-0.029	-0.035	-0.017	0.003	-0.059	-0.052	-0.087
25	-0.098	-0.043	-0.067	0.000	-0.052	-0.009	-0.039	-0.054	-0.064	-0.036
26	-0.029	-0.059	-0.132	-0.035	0.017	-0.033	0.041	0.045	-0.073	-0.044
27	-0.019	-0.023	-0.027	-0.020	-0.002	-0.012	0.023	0.041	-0.017	-0.008
28	-0.028	-0.087	-0.073	-0.107	-0.013	-0.045	-0.038	-0.051	-0.079	0.012
29	0.066	-0.074	-0.094	-0.117	-0.114	-0.077	-0.057	-0.061	-0.122	-0.024
30	0.130	0.009	0.075	-0.037	0.035	-0.009	0.037	0.060	-0.194	-0.042
31	-0.007	0.056	0.002	-0.015	-0.013	-0.060	0.058	0.016	-0.020	-0.025
32	0.022	0.008	0.177	-0.016	-0.019	0.026	-0.056	0.030	0.048	0.053
33	-0.069	-0.005	0.004	-0.032	-0.031	-0.020	0.003	0.012	-0.024	0.008

Source: Authors' calculations from 2002-2007 Nielsen Homescan data. The sample includes only consumers in the 52 identified Nielsen markets. The table shows average own and cross-price elasticities of demand. The i good represents the quantity good and the j good is the price good.

Table A-9: Own and Cross-Price Elasticities of Demand (ϵ_{ij}) - continued

						i						
j	22	23	24	25	26	27	28	29	30	31	32	33
1	-0.006	-0.074	-0.043	-0.050	0.001	-0.029	0.116	0.059	-0.006	0.013	-0.004	-0.009
2	0.013	-0.049	-0.004	-0.084	0.030	-0.034	0.027	-0.031	-0.029	0.055	-0.013	-0.085
3	0.025	-0.112	-0.010	-0.047	-0.038	-0.021	0.095	0.065	0.013	0.025	0.009	-0.076
4	0.023	0.028	0.008	0.026	0.014	0.018	-0.032	0.012	0.031	0.018	0.015	0.009
5	0.026	0.049	0.017	0.084	0.003	0.055	0.036	0.040	0.014	-0.031	0.036	0.035
6	0.032	0.015	0.004	-0.028	0.015	-0.005	0.017	-0.007	0.031	0.039	0.018	0.035
7	0.037	0.052	0.013	0.060	0.023	0.036	-0.035	0.016	0.029	0.038	0.040	0.058
8	0.033	0.068	-0.116	0.106	0.044	0.025	0.165	0.082	0.003	0.047	0.188	0.111
9	-0.008	-0.017	-0.013	-0.010	0.012	0.003	-0.027	-0.028	-0.005	0.024	0.011	0.027
10	0.001	-0.016	-0.015	-0.007	-0.002	0.000	-0.010	-0.004	0.018	0.042	0.028	-0.016
11	0.012	0.069	0.012	0.016	0.034	0.004	-0.024	-0.011	0.059	0.065	0.047	0.054
12	0.050	0.035	-0.003	-0.004	0.047	0.032	-0.026	-0.028	0.031	0.035	0.031	0.025
13	-0.029	-0.029	-0.034	-0.018	-0.012	-0.004	0.014	0.027	-0.036	-0.088	0.022	-0.001
14	-0.032	-0.075	-0.051	-0.137	-0.002	-0.093	-0.038	-0.093	-0.074	0.036	-0.057	-0.088
15	0.000	-0.050	-0.008	-0.115	-0.007	-0.020	0.020	-0.038	-0.045	-0.024	-0.010	0.010
16	-0.029	-0.072	-0.022	-0.048	-0.040	-0.071	0.156	0.027	-0.001	0.011	-0.021	-0.080
17	-0.018	-0.077	-0.017	-0.135	0.001	-0.020	0.025	-0.044	-0.060	-0.007	-0.018	0.010
18	0.017	-0.014	-0.004	-0.059	-0.012	-0.055	0.027	0.003	0.008	0.000	0.034	0.020
19	0.003	0.019	0.011	0.023	0.014	0.035	-0.011	-0.003	-0.021	-0.016	0.003	0.029
20	-0.021	-0.047	-0.018	-0.062	-0.002	-0.031	0.053	-0.005	0.048	-0.018	-0.019	0.003
21	-0.029	-0.093	-0.044	-0.098	-0.033	-0.062	0.099	-0.013	-0.027	0.010	-0.033	-0.065
22	-0.731	0.023	0.016	0.006	0.006	0.007	-0.069	-0.076	-0.055	0.019	-0.013	-0.112
23	-0.041	0.011	0.079	-0.149	-0.029	-0.073	-0.062	-0.126	-0.042	0.018	-0.036	-0.129
24	-0.029	0.175	-0.696	0.044	-0.006	-0.012	0.036	0.004	-0.115	-0.027	0.098	0.055
25	0.011	0.028	0.019	-0.253	-0.019	-0.023	-0.041	-0.023	-0.037	-0.070	-0.005	-0.042
26	0.011	-0.008	-0.007	-0.026	-0.431	0.073	-0.090	-0.075	-0.007	0.022	0.037	0.140
27	-0.009	-0.017	-0.001	-0.024	0.030	-0.669	-0.021	-0.024	-0.011	-0.014	0.018	0.036
28	-0.012	-0.070	-0.030	-0.093	0.004	-0.036	-0.270	-0.060	-0.038	0.013	-0.028	0.028
29	-0.044	-0.186	-0.045	-0.163	-0.036	-0.061	0.080	-0.295	-0.080	-0.006	-0.035	-0.101
30	-0.007	0.077	0.014	0.032	0.036	0.037	0.077	0.030	-1.793	0.255	0.036	-0.005
31	-0.012	0.051	0.010	0.090	0.011	0.062	0.059	0.071	0.107	-1.972	0.040	0.009
32	0.020	-0.027	-0.034	-0.070	-0.021	-0.057	0.079	0.049	0.016	0.024	-1.631	-0.268
33	-0.001	-0.026	-0.013	-0.033	0.003	0.015	-0.008	-0.032	-0.013	-0.009	-0.003	-0.575

Source: Authors' calculations from 2002-2007 Nielsen Homescan data. The sample includes only consumers in the 52 identified Nielsen markets. The table shows average own and cross-price elasticities of demand. The i good represents the quantity good and the j good is the price good.

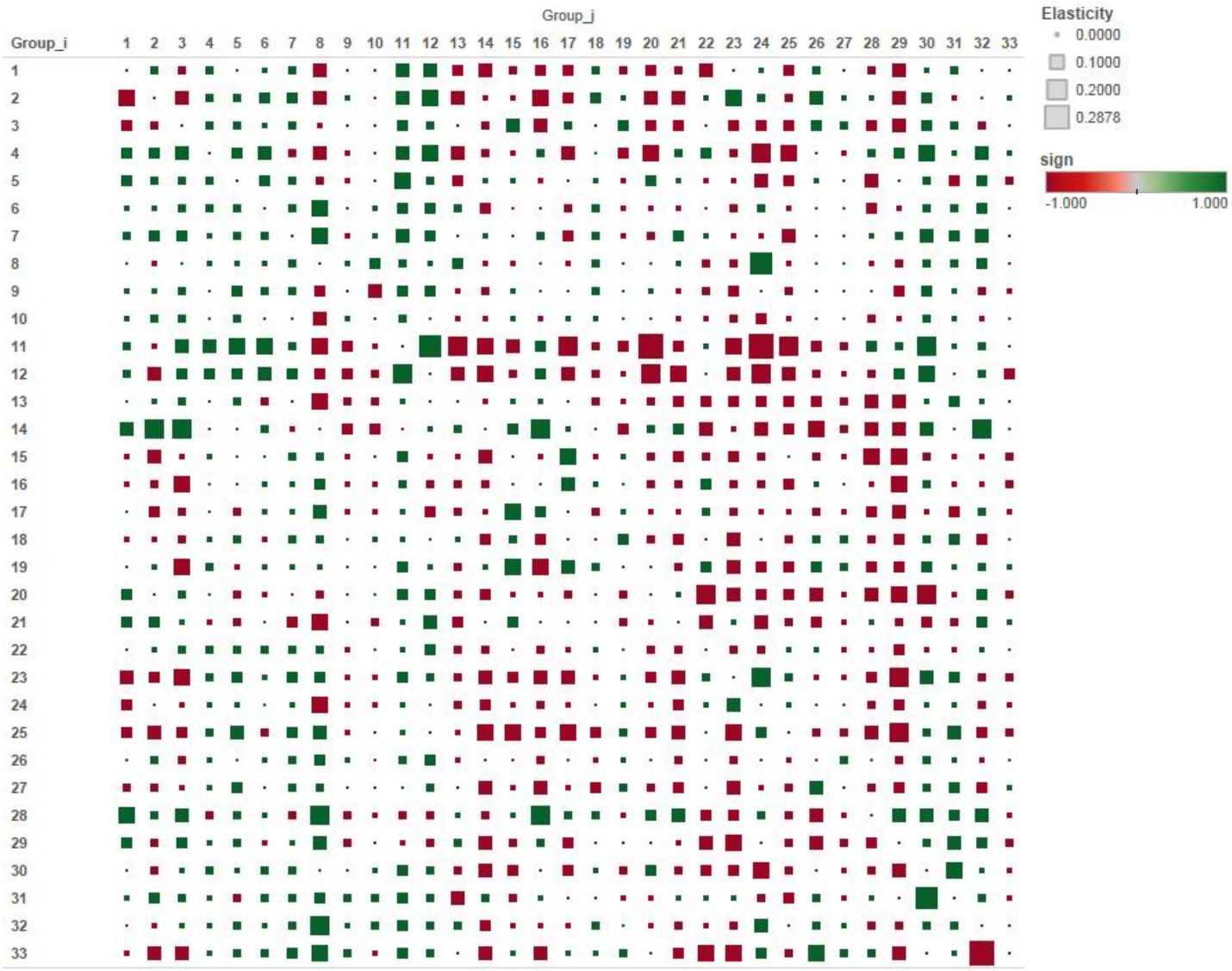

Figure A-1: Pattern of cross price elasticities. Rows correspond to product groups i and columns correspond to groups j. The size of the rectangles corresponds to the magnitude of the elasticity. The color corresponds to the sign. Own price elasticities are excluded from the plot.

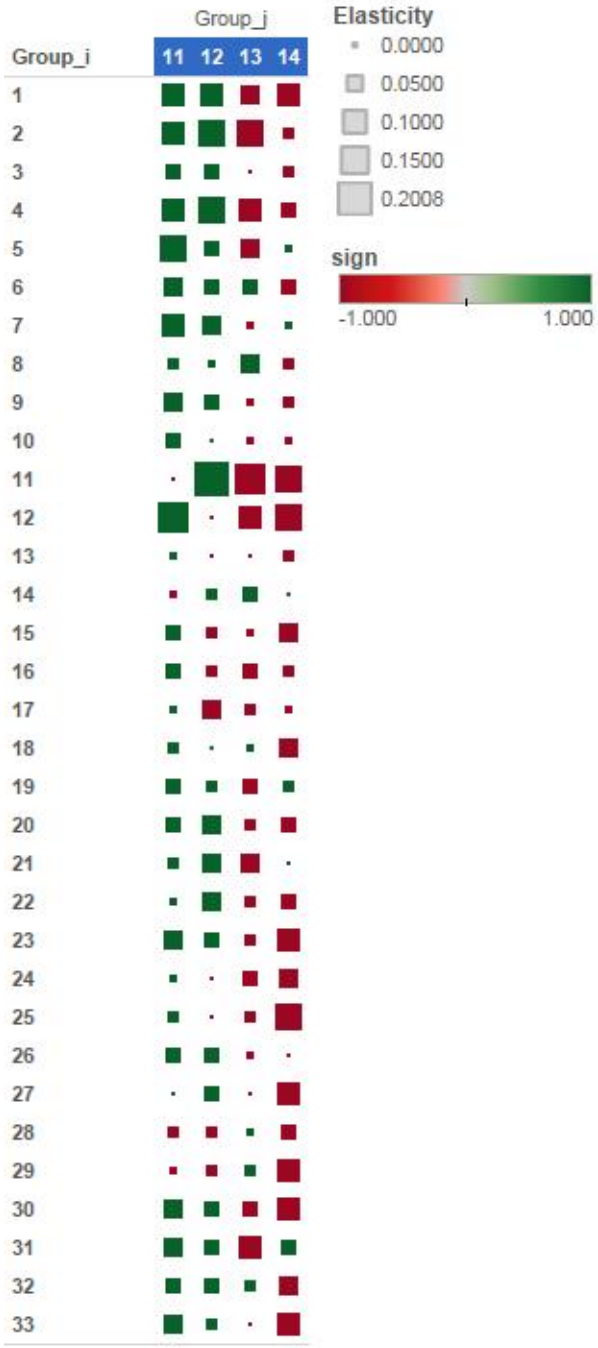

Figure A-2: Pattern of cross price elasticities for product groups 11-14, corresponding to soda and packaged meals. Rows correspond to product groups i and columns correspond to groups j. The size of the rectangles corresponds to the magnitude of the elasticity. The color corresponds to the sign. Own price elasticities are excluded from the plot.