A TWO-STAGE ESTIMATION OF ELASTICITIES FOR DISAGGREGATED SALAD PRODUCTS

by

Andrew Lobo

A Thesis Submitted to the Faculty of the

Department of Agricultural and Resource Economics

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2018

STATEMENT BY AUTHOR

The thesis titled ***A Two-Stage Estimation of Elasticities for Disaggregated Salad Products*** prepared by ***Andrew Lobo*** has been submitted in partial fulfillment of requirements for a master's degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: ANDREW LOBO

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

5-11-18

Gary Thompson — Date

Professor and Department Head, Agricultural and Resource Economics

Table of Contents **Page**

List of Tables

Abstract

Demand elasticities are estimated for seven lettuce and leafy green products through two-stage estimation using data from the 2015 National Consumer Panel. Products are aggregated into categories by the amount of convenience they offer the consumer. The two least-convenient good categories—unprocessed lettuce and fresh-processed lettuce—are found to be inferior goods, while more convenient goods are found to be normal or even luxury goods. All seven categories are found to be own-price elastic.

Problem Statement and Literature Review

Deficiencies exist in Americans' diets. According to the Centers for Disease Control and Prevention's 2013 Behavioral Risk Factor Surveillance System (BRFSS), only 9.3% of adults in the U.S. eat the USDA-recommended 2-3 cups of vegetables per day (Lee-Kwan et. al., 2017). Low consumption of vegetables can lead to deficiencies in micronutrients like potassium, dietary fiber, folate, vitamin A, and vitamin C, which in turn can increase risks for cardiovascular disease, cancer, and all-cause mortality (Aune et. al., 2017). Eating more vegetables, which have high micronutrient density, can reduce these risks. The reduction in risk is non-linear, though: there are diminishing marginal returns to consuming vegetables, and much of the benefits from them come at the lower levels of consumption (Aune et. al., 2017). These intertwined facts suggest that, for the vast majority of Americans, increasing the consumption of vegetables would result in significant health benefits. This study looks to determine what economic factors influence consumers' decisions regarding whether to buy salad and what kinds of salad to buy, with the goal of informing policy makers about potential levers that can be used to improve nutrition.

One important category of vegetables to eat to maintain overall nutrition is leafy greens. Leafy greens have high levels of calcium, fiber, folate, and vitamin C. Consumption of these vitamins and minerals can reduce the risk of type 2 diabetes (Slavin and Lloyd, 2012) and cancer (Fardet et. al., 2017), which can in turn offset healthcare costs and improve quality of life. Good nutrition generally, and leafy greens in particular, promotes optimal functioning, both physical (Amarantos, Martinez, and Dwyer, 2001) and mental (Morris et. al., 2018). Though other foods can be fortified with some of the vitamins and minerals that greens offer, our bodies do not absorb added nutrients as well as naturally-occuring nutrients (see, e.g., Slavin and Lloyd, 2012, and Caleja et. al., 2016). Leafy greens are particularly important because they are, in aggregate, the second-most consumed fresh vegetable in the U.S. and third-most consumed overall, behind only potatoes and tomatoes ("Ag and Food Statistics," 2017). Within the lettuce category, iceberg and other lighter-colored varietals are often less nutrient-dense than other, darker-colored lettuces (Kim et. al., 2016), but they all contain important nutrients that have salubrious impacts on overall health and well-being.

There is a lot of discussion about the underlying causes of why most Americans do not eat enough vegetables. Some of the suggested reasons include personal preference, price, and access (Jahns et. al., 2015). There are also factors that are correlated with whether or not someone will meet the USDA vegetable consumption recommendations, like demographic variables. For example, only 7.6% of men, 6.7% of young adults, and 5.8% of West Virginians eat the recommended amount of vegetables—all lower than the 9.3% U.S. adult average (Lee-Kwan et. al., 2017). Understanding how price, income, demographics, and food access impact consumption is critical in order to address the issue of under-consumption of vegetables.

There has been a push from both the public and private sectors to improve the diets of the U.S. population. For example, the Partnership for a Healthier America (PHA) has worked to provide more access to affordable, fresh produce across the United States by getting retail partners to build locations in areas with low food access (Simon, Kocot and Dietz, 2017). Other initiatives that have been undertaken have tried to induce consumers to purchase or consume more vegetables by changing the purchasing power of the consumer. The Supplemental Nutrition Assistance Program (SNAP, formerly Food Stamps) has incentivized fruit and vegetable consumption by providing "double dollars," dollar-for-dollar benefit matching on purchases of fruits and vegetables (Polacsek et. al., 2017); school lunches, which are free for those who receive Federal benefits, have been made healthier (Simon, Kocot and Dietz, 2017); and taxes on unhealthy foods (e.g., foods low in micronutrients and high in sugar, fat, and calories) can make healthier foods seem more attractive (Peñalvo et. al., 2017).

Though the foregoing programs are all different, their efficacy can be judged by examining the economic levers they use to achieve the goal of improving nutrition. Some of the economic levers that can be used to affect consumer demand are changing relative prices, augmenting consumer income, and increasing consumer access to products. Policies that look to make healthier goods more affordable, like the double dollars program, and policies that look to make unhealthy goods more expensive, like the Danish "fat taxes" on soft drinks and sweets (see Smed, 2012), try to induce changes in behavior through relative price changes. Other policies, such as SNAP, increases a consumer's purchasing power by providing them with money that must be used to purchase food. Economic theory suggests this will lead to higher consumption of healthy products if those healthy products are normal (and not inferior) goods. Increased access to healthy foods can also be seen in an economic light: it can reduce travel costs to purchase healthy goods, which can change the relative prices of healthy and unhealthy goods, and it can expand the total feasible set of goods that can be consumed.

In order to better understand both how consumers will react to current initiatives and to decide how to structure future initiatives, we must first understand the ways consumers respond to changes in relative prices and income. Numerous authors have used demand elasticities to understand consumer behavior. Generally, they find that produce items are own-price inelastic (for a discussion of the literature, see Dong and Lin, 2009) and have expenditure elasticities that are positive and small (see Blisard, Variyam, and Cromartie, 2003). Looking specifically at lettuce, own-price elasticities have been calculated at -0.09 (Huang, 1993), -0.12 (You et. al., 1997), and -0.14 (George and King, 1971). These elasticities were calculated using aggregate data, but consumers are heterogeneous and may have within-group elasticities that differ from overall elasticities. For example, elasticities for consumers have been shown to differ based on income level (see, e.g., Jones, 1997 and Davis et. al., 2011). However, Weatherspoon et. al. (2015) found lettuce elasticities for a poor Detroit neighborhood matched the national elasticities calculated by You (1997). Another demographic variable that has been shown to affect elasticities is age of the consumer. Gustavsen (2014) found that younger groups have lower expenditure elasticities for meats than older groups, though he did not find age differences in the expenditure elasticities for vegetables. Understanding how these demographic variables affect demand is important, as BRFSS surveys indicate that different groups fall short of eating the USDA-recommended 2-3 cups of vegetables at different rates. Therefore, understanding how policies designed to increase vegetable consumption will impact certain groups can ensure that the policies reach their goals.

One of the major drivers in the market for lettuce and leafy greens is demand for convenience. Bagged salad entered the U.S. market in 1986, according to Earthbound Farms (Hesser, 2003), and today constitutes 49% of the sales of value-added produce in U.S. retailers (Cook, 2014). Bagged salad offers advantages over head lettuce, since it does not require washing or chopping, but that convenience often comes at a higher price. Time-constrained households or households that value time more highly may opt to pay the premium for fresh-processed goods. Higher income households may do the same; Kuchler (2011) estimated an income "cutpoint" above which a consumer would switch from head lettuce to bagged lettuce. Ready-to-eat salad kits, which are even more convenient than bagged salads and often cost more, may also be substituted for less convenient goods. The separate but related drivers of head lettuce, bag salad, and kits consumption are not well understood, as lettuce is often simply considered as an aggregate category in the economics literature. This paper will look to fill this gap.

Data

The data used in this analysis are from the 2015 National Consumer Panel (NCP), which is a joint venture owned by Information Resources, Inc. (IRI) and the Nielsen Company. The data were purchased by the Economic Research Service to facilitate analysis by government and academic researchers. The panel, which is nationally representative, has transactional and demographic data on 127,484 households. Households are recruited to the panel through third-party vendors and register through NCP's online recruitment site (Muth et. al., 2016). Participating households are instructed to scan the barcodes of all food and alcohol products they purchase on each trip to grocery stores, dollar stores, pharmacies, superstores, and club stores. The household does the scanning either with a home scanner that is supplied to them by IRI/Nielsen or one that is downloaded onto their smartphone. The data that are recorded are the date of the purchase, the store where the purchase was made, each product purchased, and whether the good was on sale or if a coupon was used. A subset of households is also instructed to enter spending information on random weight purchases (i.e., purchases of items that do not have a barcode) under a generic product description.

The 2015 panel has 127,484 households in it, but only 62,004 of them are in the "static panel." To be in the static panel, a household must meet two criteria. The first criterion is that the household must have purchases in at least 11 of the 13 28-day periods, called "quads," that make up the year. The second is that average weekly spending must exceed $25 for a single-person household, $35 for a two-person household, or $45 for a three-person or larger household. Of the aforementioned 62,004-person static panel, 54,158 of those households had purchases of random weight goods over the course of 2015, which led us to believe they had the ability to key in random-weight purchases and hence were part of what is referred to as the "random weight panel." Of these remaining 54,158 households, 3,232 had no purchases of salad or leafy greens over the course of the year, so they were excluded from the analysis. Finally, because quantity data for random-weight sales are not collected, we could not use those data for our demand system. Hence, the 4,202 households who only had random-weight salad purchases were also removed from our data set. That left 46,724 households in the final panel we used for our analysis.

Table 1: number of households from the full panel that met each successive criterion for being in our panel

Households in the….	# of households
…full panel	127,484
…and the static panel	62,004
…and the random weight panel	54,158
…and made a salad purchase	46,724

Upon being recruited for participation in the National Consumer Panel, households are instructed to fill out a survey that asks for demographic information. These variables include: household size; household annual income (broken into twelve different levels, the lowest being less than $9,999 and the highest being greater than $100,000); race; marital status; whether the household owns or rents; region (northeast, midwest, south, or west); the education level and occupation of the heads of household; the size of the county the household is in (broken into four different levels); and if the household has a cat or a dog. A household can have one or two heads, but if it has two heads, the heads are of different genders. Of our 46,724 households, 35,477 (75.9%) have a male head, 42,939 (92.0%) have a female head, and 31,692 (67.8%) have both a male and a female head. The races choices are white, black, Asian, and other, and each household must select exactly one of those choice, regardless of the number of heads of household. The household then is asked if they are Hispanic or non-Hispanic, which they also must answer. The demographic data are time-invariant for 2015. Our subset of the data set resembles the full panel, which itself is nationally representative, in most cases, but with the Hispanic variable, only 6% of the subset is Hispanic, while 9% of the full panel and 17% of the U.S. is Hispanic. All of the variables are binary, and the means of these variables for our data set can be found in table 2 below.

Table 2: definitions and sample statistics of demographic variables

Variable Name	Definition	Mean
Household Size		
One	Household has one person in it	.20
Two	Household has two people in it	.44
Three	Household has three people in it	.15
Four	Household has four people in it	.13
Five	Household has five people in it	.05
Six	Household has six people in it	.02
Seven	Household has seven people in it	.01
Eight or More*	Household has eight or more people in it	<.01
Annual Household Income		
Income < $10,000	Annual household income < $10,000	.02
Income $10,000 - $11,999*	Annual household income $10,000 - $ 11,999	.01
Income $12,000 - $14,999	Annual household income $12,000 - $14,999	.01
Income $15,000 - $19,999	Annual household income $15,000 - $19,999	.03
Income $20,000 - $24,999	Annual household income $20,000 - $24,999	.05
Income $25,000 - $34,999	Annual household income $25,000 - $34,999	.11
Income $35,000 - $44,999	Annual household income $35,000 - $44,999	.11
Income $45,000 - $49,999	Annual household income $45,000 - $49,999	.05
Income $50,000 - $59,999	Annual household income $50,000 - $59,999	.10
Income $60,000 - $69,999	Annual household income $60,000 - $69,999	.08
Income $70,000 - $99,999	Annual household income $70,000 - $99,999	.22
Income > $100,000	Annual household income > $100,000	.18
Region		
West	Household resides in the Western region of the U.S.	.20
South	Household resides in the Southern region of the U.S.	.37
Midwest	Household resides in the Midwestern region of the U.S.	.18
Northeast*	Household resides in the Northeastern region of the U.S.	.25
Race		
White	Head of household is Caucasian	.81
Black	Head of household is African-American	.11
Asian	Head of household is Asian-American	.03
Other*	Head of household is Other American	.05
Educational level, female head		
Post grad	Female head has a postgraduate degree	.12
College degree	Female head has a college education	.28
Some college	Female head has some college education	.28
High school diploma	Female head has a high school diploma	.22
Less than high school	Female head has no high school diploma	.02

Table 2 continued

Variable Name	Definition	Mean
Educational level, male head		
Post grad	Male head has a postgraduate degree	.10
College degree	Male head has a college education	.21
Some college	Male head has some college education	.22
High school diploma	Male head has a high school diploma	.20
Less than high school	Male head has no high school diploma	.03
Presence of children in household		
Young children	A child age < 6 is present in the household	.08
Older children	A child age 6-12 is present in the household	.12
Teenagers	A child age 13-17 is present in the household	.14
County size		
A	County in one of 25 largest cities	.38
B	County with > 150,000 people but not an A county	.32
C	County with between 40,000 and 150,000 people	.16
D*	County with < 40,000 people	.14
Marital status		
Single	Head of household is single	.11
Married	Head of household is married	.69
Separated / divorced	Head of household is separated / divorced	.14
Widowed*	Head of household is widowed	.06
Homeowner status		
Owner	Household owns home	.79
Renter	Household rents home	.19
Other*	Household neither rents nor owns	.02
Dog owner	Household has a dog	.45
Cat owner	Household has a cat	.35
Hispanic	Head of household is Hispanic	.06
Employment status, female head		
> 35 hours	Female head of household works > 35 hours	.35
< 35 hours	Female head of household works < 35 hours	.18
Homemaker / student	Female head of household is a homemaker / student	.39
Employment status, male head		
> 35 hours	Male head of household works > 35 hours	.45
< 35 hours	Male head of household works < 35 hours	.07
Homemaker / student	Male head of household is a homemaker / student	.23

Note: reference groups are denoted with an asterisk.

Though the dataset is extensive, there are some noteworthy data integrity issues. As noted by Einav, Leibtag, and Nevo (2008), the National Consumer Panel asks its participants to input a lot of data, an onerous, time-consuming task that may lead to errors in recording the data. Recording errors occur when households underreport their consumption (such as when they purchase a good but do not scan it) or when they misreport their consumption (such as when they input the quantity of a good purchased as “11” instead of “1”). Sweitzer et. al. (2017) found that expenditures on fruits and vegetables in the IRI/Nielsen panel are lower than those in the Consumer Expenditure Survey and the National Food Acquisition and Purchase Survey, two nationally-representative government surveys that collect data on food expenditures. These findings held across all demographics and food groups, though the difference was largest in random-weight goods, such as fresh fruits and vegetables, where the IRI/Nielsen panel reporting was usually half of the other two panels.

Errors in prices may occur due to issues with how price are imputed by IRI/Nielsen and how quantities are coded. In order to reduce the amount of information that households have to input, IRI/Nielsen uses their weekly store-level data to assign a price to a product that a household reports buying. Error can be introduced here because prices may change over the course of a week, sales may occur, or consumers may use coupons. Within the panel, 69% of transactions and 65% of sales have prices that are imputed (“IRI Household and Retail Scanner Data User Guide,” 2017). Households may also improperly record the store they went to, which would result in the wrong price being imputed for the goods they purchased. There are also concerns relating to the random-weight goods in the dataset. Because retailers may vary in how they price random weight goods (e.g., by pound or by count), quantity data are, according to the documentation for the data, not collected. There does exists a random-weight quantity variable in the data set, though. For random-weight lettuce, for example, we see that 95.9% of purchases have a quantity of 1, 3.8% have a quantity of 2, .2% have a quantity of 3, and .1% have the quantities 4, 5, 6, or 10. The interpretation of this variable is difficult, though, because it may be weight or count, so we have excluded the random-weight goods from our analysis.

There are also some aspects of consumer behavior and expenditure that are not included within the dataset. One of the gaps in the dataset is that food away from home is not included. This may be problematic, since households, on average, spend 42% of their food dollars on food eaten away from home (Kuhns and Saksena, 2017). Additionally, there may be a sample selection problem when it comes to the panel. Since panel participants self-select, they may be different than the overall population. Lusk and Brooks (2011) found that, after controlling for demographics, the IRI/Nielsen panelists were more price-sensitive than participants in a random sample, suggesting sample selection and participation bias. Finally, because price data are only recorded when a household makes a purchase, we do not know the price a household faced for a product they did not buy. To solve this issue, we imputed missing prices by taking each mean product group price and assigning that as the price the household faced if they did not make a purchase. We could have used an auxiliary regression to condition estimated product prices on factors like geography (see, e.g., Park et. al., 1996), but the level of aggregation within product groups was already high, so we believed using mean price was sufficient.

Demand systems generally require assumptions to be made the separation of individual food items into groups. As discussed, we have removed random-weight lettuce from our dataset because of a lack of quantity information. We then made decisions about what remaining products could go into what groups by using information linked to each product’s Universal Product Code (UPC). UPCs are the bar codes on fixed-weight products that distinguish different products from one another. Within our data set, each UPC has information associated with it, including the name of the product, the variety of the product (e.g. romaine, iceberg), what kind of packaging it is in, and a brief description of the product. With this information, we were able to separate the UPCs in to seven distinct categories, examples of which can be seen in Appendix A.

The first category we used is unprocessed lettuce, which are products such as heads of lettuce in shrink-wrap with a bar code. This can be contrasted with the next category, fresh-processed lettuce, which are the same types of lettuce, but washed, chopped, and bagged. The next category, garden, is salads that have not just greens but other vegetables, such as shredded carrots and red cabbage. An even more processed category, kits, are salads that are ready-to-eat and include ingredients like croutons, dressing, and even animal proteins. We also included a separate category for spinach, since it is a leafy green but is used in many ways that lettuce is not, and a category for slaw—coleslaws and broccoli slaws, which may or may not have dressing—as it is made of greens and often sold in the same section as salads. Finally, we included an aggregate category for the remaining greens, which include arugula, cabbage, watercress, endive, bok choy, and others.

The number of UPCs that pertain to each of the seven categories is shown in table 3. The number of UPCs in each category is determined by a few different factors. One is how many types of products are in the category. For example, the other green category includes many types of greens, so it has more UPCs than a category like spinach, which has one type of green. Another determinant of number of UPCs in a category, though, is how differentiated the products within the category are. Kit salads can be differentiated in ways that unprocessed and fresh-processed greens cannot be, such as by varying what dressing or what protein is included. Finally, more popular types of products will often have more UPCs as more producers sell them.

Table 3: number of UPCs in each product category

Product category	Number of UPCs
Unprocessed lettuce	137
Fresh-processed lettuce	135
Garden salad	463
Kits	683
Spinach	180
Coleslaw	73
Other greens	235
Total	1,906

Table 4 displays sample statistics on quantity, expenditure, and price for the seven categories of goods over the course of the year 2015 for our panelists. One takeaway from this table is how infrequent purchase within the categories are. The mean units purchased in the category with the most purchases, garden salads, is 4.5, meaning that, on average, households are buying a garden salad every 2.5 months. Households are buying kits, on average, twice a year. The average price paid for a unit of each of the categories runs from a low of $2.06 for coleslaw to a high of $4.19 for a kit salad. Generally speaking, the more-processed, more-differentiated goods have higher prices, often in unit cost but almost always in price per ounce. Of note is the maximum column for price, as it is unlikely anyone paid any of those very large prices of a single unit of any of those goods. These large prices are likely errors, either from the household inputting the wrong price or from IRI/Nielsen imputing the wrong price. We have trimmed the highest 1% of prices in each category and replaced them all with the corresponding category's 99th percentile price. The minimum prices also look suspiciously low, but they remain in our data set.

Table 4: sample statistics on quantity, expenditure, and price (n=46,724 households), 2015

Price (average price paid in $ per household per ounce)	Mean	Median	Min	Max	99th Percentile
Unprocessed lettuce	0.21	0.21	0.03	10.67	0.56
Fresh-processed lettuce	0.32	0.32	0.04	8.4	0.8
Garden salad	0.31	0.31	0.03	5.35	0.87
Kits	0.39	0.39	0.05	8.49	0.99
Spinach	0.37	0.37	0.02	3.75	0.8
Coleslaw	0.14	0.14	0.03	1.79	0.28
Other greens	0.39	0.39	<0.00	6.84	0.78
Price (average price paid in $ per household per unit)	Mean	Median	Min	Max	99th Percentile
Unprocessed lettuce	2.83	2.85	0.5	32	5.64
Fresh-processed lettuce	2.86	2.73	0.54	42	5.98
Garden salad	2.96	2.81	0.41	59.94	6.56
Kits	4.19	3.8	0.49	119.88	8.89
Spinach	3.15	2.89	0.42	30	5.99
Coleslaw	2.06	1.8	0.48	25	3.96
Other greens	3.27	2.98	0.01	35.76	5.6

Table 4 continued

Average quantity (ounces) purchased over 12 months per household	Mean	Median	Min	Max
Unprocessed lettuce	27.94	0	0	1414
Fresh-processed lettuce	18.37	0	0	1120
Garden salad	59.89	24	0	3652
Kits	25.75	0	0	1856
Spinach	17.96	0	0	1416
Coleslaw	11.12	0	0	1358
Other greens	3.91	0	0	1392
Average quantity (units) purchased over 12 months per household	Mean	Median	Min	Max
Unprocessed lettuce	1.65	0	0	72
Fresh-processed lettuce	1.69	0	0	106
Garden salad	4.56	2	0	151
Kits	2.05	0	0	182
Spinach	1.73	0	0	98
Coleslaw	0.67	0	0	94
Other greens	0.34	0	0	45

Table 4 continued

Expenditure (average $ spent over 12 months per household)	Mean	Median	Min	Max
Unprocessed lettuce	4.98	0	0	938.78
Fresh-processed lettuce	5.01	0	0	582.21
Garden salad	13.62	4.79	0	1044.63
Kits	9.12	0	0	1258.49
Spinach	5.5	0	0	346.75
Coleslaw	1.21	0	0	189.94
Other greens	1.09	0	0	211.87

Our analysis uses price per ounce for each product in order to allow for price comparisons between goods with different weights. For example, the per-unit price for both unprocessed and fresh-processed lettuce are similar, but, as unprocessed lettuce usually weighs more than fresh-processed lettuce, the average per-ounce cost is lower (21 cents to 31 cents). Ounce weight data were missing for 354 observations out of the 926,974 salad purchases made over the course of the year in our panel, so those observations were excluded from the analysis. Some authors have refined the per-ounce analysis by choosing to include only the edible portion of that weight for unprocessed lettuce (see Kuchler, 2011), but that is outside the scope of this paper.

In table 5, we show, for each household, how many of the seven categories they have made purchases in. For example, if a household has purchased coleslaw and spinach over the course of the year but has not made purchases in any of the other categories, they would be under

“2” in the table. This would be true if they purchased both coleslaw and spinach multiple times or just once, so long as they had zero purchases of goods in all of the other five categories for the whole year. This table shows that there are many consumers who only buy a few of the types of products over the course of the year, perhaps due to prices or perhaps due to preference. The high numbers of households that buy few products suggest that there might not be broad substitutability between the categories.

Table 5: number of products purchased by each household

Number of products	Households that consume that number of products	% of households
7	452	1.0%
6	1,968	4.2%
5	4,781	10.2%
4	7,755	16.6%
3	10,253	21.9%
2	11,180	23.9%
1	10,335	22.1%

Methods

Demand system estimation in data sets with disaggregated panel microdata often encounters the problem of censoring. Censoring, for us, is caused by households who do not purchase a product from one or more of our product groups. There are many reasons why a household might have zero purchase in a group in our study. They might never buy it, perhaps due to an aversion to the product, like an allergy, or they might simply buy it infrequently, to the point where we do not see an instance of their purchase in our data. The price responses to these two kinds of non-purchase are different, though, as the consumer with an aversion to the product will not respond to price at all while the infrequent consumer may.

Our data set was initially both time-series and cross-sectional, with data on each shopping trip a household made over the course of the year. We aggregated over the full year of 2015 to remove the time-series component, since temporal aggregation often reduces censoring because infrequent purchasers have a longer period over which to buy the product. Even still, censoring remains a problem. We see censoring in all seven of our categories of products (see table 6), with the heaviest censoring occurring in the other greens category (14.5% of households made a purchase), the coleslaw category (26.0% made a purchase), and the kits category (38.1% made a purchase).

Table 6: percentage of households with zero annual purchase in each category

Product category	% of household with non-zero purchase
Unprocessed lettuce	45.8
Fresh-processed lettuce	46.1
Garden salad	74.2
Kits	38.1
Spinach	40.7
Coleslaw	26.0
Other greens	14.5

When estimating demand systems, censoring needs to be accounted for, as parameter estimates that do not consider censoring can be inconsistent, while simply removing censored data can introduce selection bias if the censoring has not occurred randomly. In estimating censored demand systems, there exist two general approaches: maximum likelihood and two-step procedures. We have chosen the two-step procedure because of its computational ease. A further discussion of the maximum likelihood approach is beyond the scope of this paper, but for reference, see Wales and Woodland (1983) and Chiang and Lee (1992). The two-step method, which is a generalization of the Heckman (1973) procedure found in Park et. al. (1996), addresses the problem of zero expenditures by first running a probit regression using all observations to determine the probability that a given household purchases the good in question. Incorporating this first-stage estimate solves the issue of censoring because, if non-purchase occurs due to preference or relative prices (which is a corner optimum), that information can be incorporated into the demand system estimation.

In our first stage, the dependent variable is y_{hi}, a binary variable for the *h*th household for the *i*th commodity that takes a value of 1 if the household makes a purchase and 0 if not. A probit model is specified such that:

$$y_{hi} = 1 \text{ if } W_{hi}\delta_i + \nu_{hi} > 0 \tag{1}$$

$$y_{hi} = 0 \text{ if } W_{hi}\delta_i + \nu_{hi} \leq 0$$

$$h = 1, \dots, H; i = 1, \dots, n$$

Where W_{hi} is a vector of regressors related to the purchase decision, δ_i is the coefficient vector for those regressors, and ν_{hi} is the random error term. The regressors used here to predict the purchase decision are the prices of the seven goods and the demographic variables discussed in the data section. From Park et. al. (1996), we see that:

$$prob[y_{hi} = 1] = \Phi(W_{hi}\delta_i) \tag{2}$$

$$prob[y_{hi} = 0] = 1 - \Phi(W_{hi}\delta_i)$$

Where $\Phi(W_{hi}\delta_i)$ indicates the standard normal cumulative density function of $(W_{hi}\delta_i)$. From here, we can derive estimates for δ_i, which will be used in the second stage.

For the second stage, we chose the Almost Ideal Demand System (Deaton and Muellbauer, 1980) to model demand. For the purposes of estimation, the seven goods in the system are considered to be weakly separable from other foods. In the AIDS model, w_i is the expenditure share on commodity *i* for household *h* (though, following convention, we have dropped the household subscripts), *x* is total salad expenditure, and p_i is the price of the *i*th commodity.

$$w_i = \alpha_i + \beta_i \left(log \frac{x}{P}\right) + \sum_{k=1}^{n} \gamma_{ik} log p_k + \varepsilon_i \tag{3}$$

$$i = 1, 2, \dots, n$$

and where $log\ P = \alpha_0 + \sum_{l=1}^{n} \alpha_l log p_l + .5 \sum_{l=1}^{l} \sum_{k=1}^{n} \gamma_{lk} log p_l log p_k$

We use the linear approximation (LA-AIDS) suggested by Deaton and Muellbauer, though, where:

$$log\ P^* = \sum_{l=1}^{n} w_i log p_l \tag{4}$$

The demand system is made to fit the following theoretical restrictions:

$$\sum_{i=1}^{n} \alpha_i = 1,\ \ \sum_{i=1}^{n} \beta_i = 0,\ \sum_{i=1}^{n} \gamma_{ik} = 0 \text{ (adding up)} \tag{5}$$

$$\sum_{k=1}^{n} \gamma_{ik} = 0 \text{ (homogeneity)}$$

$$\gamma_{ik} = \gamma_{ki} \text{ (symmetry)}$$

The demographic variable vector d_j is incorporated in the demand equation by parameterizing vector a_i such that:

$$a_i = \sum_{j=1}^{j} a_i d_j \tag{6}$$

This guarantees adding-up of the deterministic system (see Yen, Lin, and Smallwood, 2003 and Davis et. al., 2011). Finally, the results of the first stage are incorporated into the demand system. The censoring gives us a model that looks like this:

$$w_i = \sum_{j=1}^{j} a_i d_j + \beta_i \left(log \frac{x}{P^*}\right) + \sum_{k=1}^{n} \gamma_{ik} logp_k + \varepsilon_i \text{ if } W_i \delta_i + \nu_i > 0 \tag{7}$$

$$w_i = 0 \text{ if } W_i \delta_i + \nu_i \leq 0$$

From Yen, Lin, and Smallwood (2003), using the normality of the marginal distribution of [ν_i, ε_i] $\forall$ i, the expectation of the price share is then:

$$E[w_i] = \Phi(W\delta_i) * \left[\sum_{j=1}^{j} a_i d_j + \beta_i \left(log \frac{x}{P^*}\right) + \sum_{k=1}^{n} \gamma_{ik} logp_k \right] + \tau_i \phi(W\delta_i) \tag{8}$$

Where $\phi(W\delta_i)$ indicates the standard normal probability density function of $(W\delta_i)$. After we find estimates for δ_i in the probit regression, we can rewrite the above equation as:

$$w_i = \Phi\left(W\hat{\delta}_i\right) * \left[\sum_{j=1}^{j} a_i d_j + \beta_i \left(log \frac{x}{P^*}\right) + \sum_{k=1}^{n} \gamma_{ik} logp_k \right] + \tau_i \phi\left(W\hat{\delta}_i\right) + \xi_i \tag{9}$$

Where $\widehat{\delta_\iota}$ is the estimated value of the parameter vector from the first stage, τ_i is a scalar parameter for the standard normal PDF of the first stage (which, in the second stage, is a regressor), and ξ_i is the error term vector. (For a discussion of the expectation and variance of the error term, see Shonkwiler and Yen (1999).) We then calculate and report elasticities using the methodology from Sam and Zheng (2010), where e_{ik}, the uncompensated price elasticity, is calculated as:

$$e_{ik} = -\theta_{ik} + \left[\frac{\partial E(w_i)}{\partial \log(p_k)}\right]\frac{1}{E(w_i)} = \tag{10}$$

$$-\theta_{ik} + \left\{\left[\phi(W\delta_i)\delta_{ik}(\sum_{j=1}^{j} a_i d_j + \beta_i\left(log\frac{x}{P^*}\right) + \sum_{k=1}^{n}\gamma_{ik} logp_k)\right]\right.$$

$$+\left[\Phi(W\delta_i)(\gamma_{ik} - \beta_i(\sum_{j=1}^{j} a_k d_j + \sum_{k=1}^{n}\gamma_{ik} logp_i)\right]$$

$$\left. + [\tau_i\delta_{ik}(W\delta_i)\phi(W\delta_i)]\right\} * \frac{1}{E(w_i)}$$

Where θ_{ik} is the Kronecker delta (θ_{ik}=1 for i=k; θ_{ik}=0 otherwise). e_i, the expenditure elasticity, is calculated as:

$$e_i = 1 + \left[\frac{\partial E(w_i)}{\partial \log(x)}\right]\frac{1}{E(w_i)} = 1 + \Phi(W\delta_i) * \frac{\beta_i}{E(w_i)} \tag{11}$$

The elasticities for the residual commodity can be found through Engel ($\sum_{i=1}^{n} w_i e_i = 1$), Cournot ($\sum_{i=1}^{n} w_i e_{ik} + w_k = 0$), and Euler ($\sum_{k=1}^{n} e_{ik} + e_i = 0$) aggregation, and the compensated elasticities can be found using the Slutsky equation ($e_{ik}^* = e_{ik} + w_k e_i$). Engel aggregation runs into an issue, though, when w_r, the share for the residual good, is zero, since, rewritten, the formula for e_r, the residual good's expenditure elastictity, is:

$$e_r = \frac{\sum_{i=1}^{n-1} w_i e_i}{w_r} \tag{12}$$

Hence, when using Engel aggregation, we derived E(w_i) for each good and used that instead of w_i.

Confidence intervals are generated through bootstrapping. Two hundred (200) samples of 46,724 households are drawn with replacement from the set of households with salad consumption and all elasticities are calculated for each new sample. We then sort each elasticity's 200 observations by value and report the intervals between the 6th and 195th observations for each elasticity, which correspond to the 95% confidence intervals.

Results

Results from the first stage probit regressions are reported in Appendix B, but some results are discussed below. The higher levels of income were associated with statistically significant positive effects on the purchasing decisions of all products (with the exception of garden salad, where the positive effects were not statistically significant), with the magnitude of the effect being larger at the higher levels. The majority of the household dummies were not statistically significant, though there was some evidence that single-person households are less likely to buy some types of lettuce. The presence of children, both younger and older, dogs, and cats in the household all also suggested generally lower rates of purchase of salad products.

The race variables also revealed information regarding preferences. African-Americans had a large, significant, positive effect for consuming other greens and garden salads and a smaller, negative effect toward unprocessed lettuce and kits. Asians had a negative and significant result toward each category except spinach. Whites were more likely to opt for garden salads and unprocessed and fresh-processed lettuce, while buying less of other greens, kits, and spinach. Also significant were higher purchasing rates across the board for households in cities with more than 140,000 people. Having a female head of household who works also suggested less salad purchase in some categories, though so did having a female head of household who was a homemaker or student. Higher levels of educational attainment were associated with higher levels of salad consumption.

The elasticities generated from the demand system are reported below in table 7. The expenditure elasticities are noteworthy. We see elasticities greater than one for kits, garden salads, slaw, and spinach, indicating that they are luxury goods. Kits and garden salads are two of the most processed goods in our study, so seeing households increase their consumption of those goods as their income increases can be understood in the context of them choosing to consume more of more convenient goods. Other greens has an elasticity between zero and one, showing consumers increasing their consumption of them as their income rises, but proportionally less with respect to income. Most interestingly, we see unprocessed lettuce and fresh-processed lettuce, the two least-processed goods in our study, with negative income elasticities, showing them to be an inferior goods that consumers make fewer overall purchases of as their income rises. These negative expenditure elasticities can also be seen in the context of convenience: as incomes rise, demand for more convenient products rises.

Table 7: expenditure and own-price elasticities

	Unprocessed lettuce	Fresh-processed lettuce	Garden salad	Kits	Spinach	Slaw	Other greens	Expenditure
Unprocessed lettuce	-1.79	0.27	0.22	0.01	0.06	1.44	0.62	-0.82
Fresh-processed lettuce		-1.83	0.12	0.04	0.11	1.29	0.21	-0.21
Garden salad			-1.24	-0.06	0.19	-1.64	-0.30	2.70
Kits				-1.86	0.09	-1.40	0.25	1.18
Spinach					-1.93	0.36	0.88	2.00
Slaw						-2.25	1.94	1.12
Other greens							-4.04	0.44

All of the own price elasticities are negative, as expected. Garden salads are close to unit elastic, while the other categories all are elastic. This result is surprising, as the literature suggests that lettuce is typically own-price inelastic. The good bundle in our study that consumers are most price-sensitive to is the other green category, with an own-price elasticity of –4.04. That category includes many specialty greens that households might be unfamiliar with, which could explain the high price sensitivity. (Glaser and Thompson, 1998, comment that newly-introduced and novelty goods tend to have higher own-price elasticities.) The unprocessed and fresh-processed salads, kit salads, and spinach all have elasticities that are tightly grouped between -1.79 and -1.93. This broadly suggests that consumers are price-sensitive within all of our categories.

Finally, the cross-price elasticities, for the most part, are close to zero, but many are statistically significant. Some of the noteworthy cross-price elasticities that are significant are the ones between unprocessed lettuce, processed lettuce and garden salads. These three goods are shown as substitutes, despite the differences in convenience levels between them. We also see kits being substitutable with garden salads and fresh-processed lettuce, but not with unprocessed lettuce. Spinach is substitutable with unprocessed and fresh-processed salads, as well as with garden salads and kits. Few of the slaw and other greens cross-price elasticities are significant.

Policy Implications

The fact that many of the expenditure elasticities are greater than one suggests that, for those products, there are increasing nutritional returns to increasing the income of a consumer. Typically, aggregated vegetable expenditure elasticities are shown to be less than one and vegetables are thought of as necessity goods, not luxuries. Our disaggregation shows great variability in the expenditure elasticities for the different product categories, though, with the more-convenient, more-processed goods having larger elasticities while the less-convenient, less-processed goods have smaller, or even negative, elasticities. Policies like SNAP, then, are likely to improve nutrition, as people will buy more of the majority of these categories of healthy goods when their purchasing power increases.

The fact that the expenditure elasticities for unprocessed lettuce and fresh-processed lettuce are negative is an interesting finding. Though households who have more purchasing power may purchase larger amounts of some salad products, they will reduce their consumption of the less-convenient forms of salad. This switching suggests that some of the nutritional gains from increasing a household’s income may be lost as consumers stop consuming other healthy but less convenient types of salad. This also raises the question of the relative nutrition of the different types of salad: if consumers switch from more-healthy to less-healthy types of salad, then the policy intervention of increasing income may in fact have negative consequences. It is unclear from our data set which salads are healthy or unhealthy, both because we do not have nutritional data and because we do not know the other products that households consume with different types of lettuce, but it is a potential avenue for future work.

The evidence also provides indirect support for the argument that the double dollars program in SNAP could be effective at improving consumers' nutrition. We have shown all groups of salad to be elastic, which means that lowering the relative prices of salads will lead to an increase in the consumption of salad. Previous studies that found salad to be inelastic suggest this is not the case. In those studies, changing relative prices had a small impact, though this may be because food consumption is inelastic at the aggregate level but elastic at the consumer level (Durham and Eales, 2011). One reason for this discrepancy could be that, with disaggregate data, there is more variability and thus a larger range of price-quantity combinations over which consumer behavior can be observed. Our finding regarding the elasticities of salad should be heartening: it suggests that finding ways to make salad cheaper or make unhealthy food more expensive will result in consumers eating healthier foods.

Future Work

Our demand system currently only looks at seven goods, but these categories could potentially be disaggregated further. For instance, we have one category for spinach, but spinach can come in both unprocessed and fresh-processed forms. Future work could center on seeing if consumer behavior regarding the disaggregated types of spinach mirrors our findings on unprocessed and fresh-processed lettuce. Disaggregating the other greens category further might also be interesting, as many of the goods in it are reasonably different (i.e., kale and cabbage).

Future work could also attempt to solve the problem of random-weight goods. Since we only have the amount a household paid for a random-weight good and not what quantity was purchased, we were not able to incorporate random-weight goods into our demand system. This is problematic, though, as random-weight lettuce is likely a substitute for other types of unprocessed lettuce, meaning it should be included in our demand system. Understanding the relationship between random-weight and fixed-weight lettuce has important policy implications—perhaps consumers with rising incomes also switch away from random-weight lettuce, offsetting nutritional gains even more—so finding a way to use the random-weight data would shed more light on the dynamics of these products.

Our analysis used the LA-AIDS model, but another AIDS models may have been more appropriate. In one study, Durham and Eales (2010) used AIDS, LA-AIDS, QUAIDS, and log-log models to derive demand elasticities for produce in two supermarkets in Portland over 80 weeks. They then used the elasticities to predict consumer behavior in the subsequent 60 weeks in order to see which model's results were the closest. As our data are panel in nature, we could also do this to see which model would give us the most predictive results. The fact that the data are panel also lends itself to adding a time-series component to our analysis, which is another avenue for future work.

One piece of this project that could be improved upon is the imputation of prices for households who did not purchase the good over the course of the year. We used product group mean price as the price households faced if they did not make a purchase in that product group. Future work could focus on using regression analysis to augment those mean prices based on, for instance, the region of the country the household is in. This would be particularly important if the demand system were further disaggregated, as the ranges in prices would likely decrease and smaller differences in price would be more important.

One important piece that was left out of our analysis was the nutrient value of the different types of salad. Generally speaking, the more-processed goods have fewer ounces of salad greens, so any switching to them could potentially mean a reduction in nutrient consumption; our analysis did not consider that. Different types of lettuces have different amounts of nutrients as well—for instance, if consumers switch from kale to iceberg, they are likely to be consuming fewer nutrients. The more-processed salads in our study are also likely to be less healthy, as they frequently contain meats, cheese, and dressings. Because we do not know how people prepare their unprocessed and fresh-processed salads, though, it is hard to say which type of salad provides the most nutrients. Future work could look to see the relative nutrition content of the goods when consumed to understand how potential policy would practically affect nutrition.

Finally, our analysis does not include food away from home. As discussed above, food away from home is an important and growing piece of the food consumption puzzle, but our data set did not have data on it. Using other data sources, such as the Consumer Expenditure Survey or the National Food Acquisition and Purchase Survey, could allow for the building of demand systems that include food away from home, which could provide a fuller picture of what salad consumption looks like in the U.S.

Bibliography

Ag and Food Statistics: Charting the Essentials (2017). United States Department of Agriculture, Economic Research Service Administrative Publication 075.

Amarantos, E., Martinez, A., & Dwyer, J. (2001). Nutrition and quality of life in older adults. *The Journals of Gerontology series A: Biological sciences and Medical sciences*, *56*(suppl_2), 54-64.

Aune, D., Giovannucci, E., Boffetta, P., Fadnes, L. T., Keum, N., Norat, T., Greenwood, D., Rioli, E., Vatten, L., & Tonstad, S. (2017). Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality—a systematic review and dose-response meta-analysis of prospective studies. *International Journal of Epidemiology*, *46*(3), 1029-1056.

Blisard, N., Variyam, J. N., & Cromartie, J. (2003). Food expenditures by US households: looking ahead to 2020. United States Department of Agriculture, Economic Research Service Agricultural Economics Reports 34045.

Caleja, C., Barros, L., Antonio, A. L., Carocho, M., Oliveira, M. B. P., & Ferreira, I. C. (2016). Fortification of yogurts with different antioxidant preservatives: A comparative study between natural and synthetic additives. *Food chemistry*, *210*, 262-268.

Chiang, J., & Lee, L. F. (1992). Discrete/continuous models of consumer demand with binding nonnegativity constraints. *Journal of Econometrics*, *54*(1-3), 79-93.

Cook, R. (2014). Trends in the marketing of fresh produce and fresh-cut/value-added produce. *Articles and presentations, Roberta Cook, Faculty, UC Davis Agriculture and Resource Economics.* Retrieved from: http://agecon ucdavis.edu/ (Website accessed: February 2018).

Davis, C. G., Yen, S. T., Dong, D., & Blayney, D. P. (2011). Assessing economic and demographic factors that influence United States dairy demand. *Journal of dairy science*, *94*(7), 3715-3723.

Deaton, A., & Muellbauer, J. (1980). *Economics and consumer behavior*. Cambridge university press.

Dong, D., & Lin, B. H. (2009). Fruit and vegetable consumption by low-income Americans. United States Department of Agriculture, Economic Research Report 70.

Durham, C., & Eales, J. (2010). Demand elasticities for fresh fruit at the retail level. *Applied Economics*, *42*(11), 1345-1354.

Einav, L., Leibtag, E., & Nevo, A. (2008). On the accuracy of Nielsen Homescan data. United States Department of Agriculture, Economic Research Report 69.

Fardet, A., Druesne-Pecollo, N., Touvier, M., & Latino-Martel, P. (2017). Do alcoholic beverages, obesity and other nutritional factors modify the risk of familial colorectal cancer? A systematic review. *Critical reviews in oncology/hematology*, *119*, 94-112.

George, P. S., & King, G. A. (1971). Consumer demand for food commodities in the United States with projections for 1980 (Vol. 26). California Agricultural Experiment Station.

Glaser, L. K., & Thompson, G. D. (1998, August). Demand for organic and conventional frozen vegetables. Presented at the American Agricultural Economics Association Annual Meeting, Nashville, TN.

Gustavsen, G. W. (2014). Consumer cohorts and demand elasticities. *European Review of Agricultural Economics*, *42*(2), 217-237.

Gustavsen, G. W., & Rickertsen, K. (2006). A censored quantile regression analysis of vegetable demand: the effects of changes in prices and total expenditure. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie, 54*(4), 631-645.

Heckman, J.J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and A Simple Estimator for Such Models. *Annals Econ. and Soc. Meas.* 5:475-492.

Hesser, Amanda (2003, January 14). Salad in Sealed Bags Isn't So Simple, It Seems. *The New York Times*. Retrieved from https://www.nytimes.com/2003/01/14/us/salad-in-sealed-bags-isn-t-so-simple-it-seems.html (Website accessed: April 2018).

Huang, K. S. (1993). *A complete system of US demand for food*. Washington, DC: United States Department of Agriculture, Economic Research Service Technical Bulletin No. 1821.

IRI Household and Retail Scanner Data User Guide. (2017). NORC Data Enclave Documentation.

Jahns, L., McDonald, L., Wadsworth, A., Morin, C., Liu, Y., & Nicklas, T. (2015). Barriers and facilitators to following the Dietary Guidelines for Americans reported by rural, Northern Plains American-Indian children. *Public health nutrition*, *18*(3), 482-489.

Jones, E. (1997). An analysis of consumer food shopping behavior using supermarket scanner data: differences by income and location. *American Journal of Agricultural Economics*, *79*(5), 1437-1443.

Kim, M. J., Moon, Y., Kopsell, D. A., Park, S., Tou, J. C., & Waterland, N. L. (2016). Nutritional value of crisphead 'Iceberg'and romaine lettuces (Lactuca sativa L.). *Journal of Agricultural Science*, *8*(11), 1.

Kuchler, F. (2011). Is it food quality or quantity that responds to changing income? *Applied Economic Perspectives and Policy*, *33*(2), 205-221.

Kuhns, A., & Saksena, M. (2017). Food Purchase Decisions of Millennial Households Compared to Other Generations. United States Department of Agriculture, Economic Research Service Economic Information Bulletin Number 186.

Lee-Kwan, S. H., Moore, L. V., Blanck, H. M., Harris, D. M., & Galuska, D. (2017). Disparities in State-Specific Adult Fruit and Vegetable Consumption—United States, 2015. *MMWR. Morbidity and mortality weekly report*, *66*(45), 1241.

Lusk, J. L., & Brooks, K. (2011). Who participates in household scanning panels? *American Journal of Agricultural Economic, 93*(1), 226-240.

Morris, M. C., Wang, Y., Barnes, L. L., Bennett, D. A., Dawson-Hughes, B., & Booth, S. L. (2018). Nutrients and bioactives in green leafy vegetables and cognitive decline: Prospective study. *Neurology*, *90*(3), e214-e222.

Muth, M. K., Sweitzer, M., Brown, D., Capogrossi, K., Karns, S., Levin, D., Okrent, A., Siegel, P., & Zhen, C. (2016). Understanding IRI household-based and store-based scanner data. United States Department of Agriculture, Economic Research Service Technical Bulletin 1942.

Okrent, A. M., & Alston, J. M. (2011). Demand for food in the United States: a review of literature, evaluation of previous estimates, and presentation of new estimates of demand (No. 48). Giannini Foundation of Agricultural Economics, University of California.

Park, J. L., Holcomb, R. B., Raper, K. C., & Capps Jr, O. (1996). A demand systems analysis of food commodities by US households segmented by income. *American Journal of Agricultural Economics*, *78*(2), 290-300.

Peñalvo, J. L., Cudhea, F., Micha, R., Rehm, C. D., Afshin, A., Whitsel, L., Wilde, P., Gaziano, T., Pearson-Stuttard, J., O'Flaherty, M., & Capewell, S. (2017). "The potential impact of food taxes and subsidies on cardiovascular disease and diabetes burden and disparities in the United States." *BMC medicine,* 15(1), 208.

Polacsek, M., Moran, A., Thorndike, A. N., Boulos, R., Franckle, R. L., Greene, J. C., Blue, D. J., Block, J. P., & Rimm, E. B. (2017). A Supermarket Double-Dollar Incentive Program Increases Purchases of Fresh Fruits and Vegetables Among Low-Income Families With Children: The Healthy Double Study. *Journal of nutrition education and behavior*, *50*(3), 217-228.

Sam, A. G., & Zheng, Y. (2010). Semiparametric estimation of consumer demand systems with micro data. *American Journal of Agricultural Economics*, *92*(1), 246-257.

Shonkwiler, J. S., & Yen, S. T. (1999). Two-step estimation of a censored system of equations. *American Journal of Agricultural Economics*, *81*(4), 972-982.

Simon, C., Kocot, S. L., & Dietz, W. H. (2017). Partnership for a Healthier America: Creating Change Through Private Sector Partnerships. *Current obesity reports*, *6*(2), 108-115.

Slavin, J. L., & Lloyd, B. (2012). Health benefits of fruits and vegetables. *Advances in nutrition*, *3*(4), 506-516.

Smed, S. (2012). Financial penalties on foods: the fat tax in Denmark. *Nutrition Bulletin*, *37*(2), 142-147.

Sweitzer, M., Brown, D., Karns, S., Muth, M. K., Siegel, P., & Zhen, C. (2017). Food-at-Home Expenditures: Comparing Commercial Household Scanner Data From IRI and

Government Survey Data. United States Department of Agriculture, Economic Research Service Technical Bulletin 1946.

Wales, T. J., & Woodland, A. D. (1983). Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics*, *21*(3), 263-285.

Weatherspoon, D., Oehmke, J., Dembele, A., & Weatherspoon, L. (2015). Fresh vegetable demand behaviour in an urban food desert. *Urban Studies*, *52*(5), 960-979.

Yen, S. T., Lin, B. H., & Smallwood, D. M. (2003). Quasi-and simulated-likelihood approaches to censored demand systems: food consumption by food stamp recipients in the United States. *American Journal of Agricultural Economics*, *85*(2), 458-478.

You, Z., Huang, C. L., & Epperson, J. E. (1997). Demand elasticities for fresh vegetables in the United States. *Journal of International Food & Agribusiness Marketing*, *9*(2), 57-71.

ppendix A: categories of salads

Clockwise from top left: unprocessed lettuce, fresh-processed lettuce, garden salad, salad kit, other greens, spinach, and coleslaw.

Photo credits:

Unprocessed lettuce: https://www.walmart.com/ip/Head-Lettuce-each/10402650

Fresh-processed lettuce: https://www.readypac.com/product/shredded-iceberg-lettuce/

Garden salad: https://www.freshexpress.com/products/salad-greens/iceberg-garden

Salad kit: https://www.freshexpress.com/products/kits/gourmet-cafe-salads-chicken-caesar-salad-kit

Other greens: http://www.taylorfarms.com/products/power-greens/chopped-kale/

Spinach: http://www.taylorfarms.com/products/classic-bag-salads/baby-spinach/

Coleslaw: http://www.dolesalads.ca/products/colourful-coleslaw/

Appendix B: results from the first stage probit regressions

	Unprocessed		Fresh-processed		Garden		Kit	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Intercept	-0.58	<.0001	-0.60	<.0001	0.60	<.0001	-0.80	<.0001
Midwest region	0.10	<.0001	0.01	0.63	0.03	0.12	-0.01	0.47
South region	0.07	<.0001	-0.02	0.17	-0.04	0.01	0.17	<.0001
West region	-0.01	0.43	-0.05	0.01	0.02	0.26	0.52	<.0001
Income < $10,000	0.13	0.08	0.11	0.15	-0.08	0.27	0.10	0.18
Income $12,000 - $14,999	0.08	0.33	-0.02	0.84	-0.04	0.60	-0.03	0.66
Income $15,000 - $19,999	0.09	0.23	0.09	0.17	-0.07	0.33	0.04	0.53
Income $20,000 - $24,999	0.14	0.04	0.06	0.40	-0.08	0.24	0.09	0.19
Income $25,000 - $34,999	0.19	0.00	0.14	0.03	-0.03	0.67	0.12	0.07
Income $35,000 - $44,999	0.25	0.00	0.15	0.02	-0.04	0.59	0.12	0.07
Income $45,000 - $49,999	0.25	0.00	0.15	0.03	-0.02	0.72	0.15	0.03
Income $50,000 - $59,999	0.30	<.0001	0.18	0.01	0.04	0.60	0.15	0.02
Income $60,000 - $69,999	0.29	<.0001	0.24	0.00	0.05	0.44	0.19	0.00
Income $70,000 - $99,999	0.32	<.0001	0.24	0.00	0.07	0.30	0.23	0.00
Income > $100,000	0.36	<.0001	0.33	<.0001	0.08	0.23	0.29	<.0001
HH size one	-0.28	0.01	-0.02	0.82	-0.06	0.62	0.05	0.68
HH size two	-0.22	0.04	0.06	0.58	-0.01	0.93	0.07	0.54
HH size three	-0.18	0.08	0.10	0.36	0.01	0.96	0.10	0.36
HH size four	-0.14	0.17	0.11	0.29	0.06	0.60	0.13	0.21
HH size five	-0.10	0.36	0.08	0.43	0.04	0.73	0.11	0.32
HH size six	-0.07	0.54	0.08	0.44	0.04	0.71	0.08	0.46
HH size seven	-0.04	0.76	-0.04	0.78	0.03	0.81	0.16	0.21
Female head	0.04	0.04	0.04	0.02	0.05	0.01	0.07	0.00
Have kids 0-6	-0.02	0.49	0.05	0.05	0.00	0.99	-0.03	0.28
Have kids 7-12	-0.05	0.04	0.02	0.37	-0.06	0.01	-0.03	0.19
Have kids 13-18	0.02	0.30	0.08	0.00	-0.01	0.62	0.04	0.05
HH has dogs	-0.04	0.00	-0.02	0.09	0.01	0.44	0.03	0.04
HH has cats	-0.04	0.00	0.00	0.81	0.03	0.04	0.03	0.02
White	0.05	0.09	0.12	<.0001	0.04	0.22	-0.07	0.03
Black	-0.20	<.0001	-0.02	0.52	0.20	<.0001	-0.09	0.01
Asian	-0.13	0.00	-0.22	<.0001	-0.15	0.00	-0.20	<.0001
Divorced	-0.01	0.68	0.02	0.51	0.02	0.49	0.03	0.40
Single	-0.07	0.03	-0.01	0.66	-0.06	0.08	0.02	0.59
Married	0.08	0.03	0.07	0.06	0.00	0.98	0.00	0.99
Hispanic	-0.03	0.19	0.09	0.00	-0.04	0.22	-0.02	0.40
County size small	0.15	<.0001	0.00	0.87	0.00	0.86	0.18	<.0001
County size medium	0.14	<.0001	0.02	0.38	0.03	0.19	0.09	<.0001
County size large	0.10	<.0001	0.03	0.23	0.03	0.26	0.03	0.27
Female employed > 35 hours	-0.05	0.00	0.03	0.15	0.00	0.87	0.02	0.32
Female homemaker / student	0.00	0.82	-0.01	0.62	-0.03	0.10	-0.03	0.05
Male employed > 35 hours	-0.04	0.10	0.00	0.88	-0.01	0.75	-0.02	0.51
Male homemeaker / student	-0.01	0.81	-0.04	0.11	-0.04	0.15	-0.04	0.13
Female ed less than high school	0.07	0.26	0.07	0.20	-0.01	0.86	-0.01	0.85
Female ed high school	0.12	0.00	0.05	0.19	0.04	0.32	0.05	0.19
Female ed some college	0.19	<.0001	0.08	0.02	0.06	0.11	0.11	0.00

Appendix B continued

	Unprocessed		Fresh-processed		Garden		Kit	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Female college graduate	0.19	<.0001	0.05	0.20	0.09	0.02	0.06	0.13
Female ed post graduate	0.16	<.0001	0.04	0.28	0.09	0.03	0.07	0.09
Male ed less than high school	-0.14	0.01	-0.10	0.04	-0.04	0.44	-0.08	0.10
Male ed high school	-0.05	0.18	-0.07	0.07	0.01	0.86	-0.06	0.14
Male ed some college	0.03	0.44	-0.05	0.16	0.03	0.48	-0.03	0.41
Male ed college graduate	0.05	0.16	-0.08	0.05	0.02	0.68	-0.04	0.36
Male ed post graduate	0.06	0.14	-0.12	0.00	0.00	0.92	-0.05	0.23
Owns home	0.09	0.04	-0.01	0.90	-0.09	0.07	-0.06	0.16
Rents home	0.01	0.77	0.06	0.20	-0.07	0.18	0.01	0.82
Log price kits	-0.16	<.0001	-0.11	<.0001	-0.20	<.0001	-0.85	<.0001
Log price slaw	-0.05	0.06	-0.01	0.66	-0.12	0.00	0.03	0.39
Log price other greens	-0.08	0.00	-0.02	0.27	0.07	0.00	-0.05	0.02
Log price spinach	-0.05	0.01	0.14	<.0001	0.12	<.0001	0.09	<.0001
Log price fresh processed	-0.38	<.0001	-0.90	<.0001	0.15	<.0001	-0.02	0.28
Log price garden	0.30	<.0001	0.19	<.0001	-0.51	<.0001	0.09	<.0001
Log price unprocessed	-0.63	<.0001	0.57	<.0001	0.34	<.0001	0.26	<.0001

	Spinach		Slaw		Other Greens	
	Estimate	p-value	Estimate	p-value	Estimate	p-value
Intercept	-0.63	<.0001	-0.91	<.0001	-1.34	<.0001
Midwest region	0.14	<.0001	-0.15	<.0001	0.26	<.0001
South region	0.08	<.0001	0.02	0.24	0.19	<.0001
West region	0.11	<.0001	-0.11	<.0001	0.23	<.0001
Income < $10,000	0.02	0.83	-0.01	0.86	0.07	0.45
Income $12,000 - $14,999	-0.11	0.17	-0.05	0.51	-0.07	0.52
Income $15,000 - $19,999	-0.10	0.18	0.04	0.61	0.04	0.68
Income $20,000 - $24,999	-0.02	0.72	0.04	0.58	0.03	0.74
Income $25,000 - $34,999	-0.01	0.84	0.07	0.33	0.04	0.68
Income $35,000 - $44,999	0.10	0.12	0.07	0.33	0.11	0.21
Income $45,000 - $49,999	0.13	0.05	0.11	0.11	0.14	0.13
Income $50,000 - $59,999	0.12	0.07	0.08	0.23	0.14	0.10
Income $60,000 - $69,999	0.13	0.04	0.10	0.15	0.21	0.02
Income $70,000 - $99,999	0.19	0.00	0.14	0.04	0.23	0.01
Income > $100,000	0.22	0.00	0.12	0.10	0.28	0.00
HH size one	-0.04	0.74	-0.17	0.16	-0.25	0.07
HH size two	0.02	0.85	-0.12	0.28	-0.15	0.26
HH size three	0.10	0.37	-0.13	0.24	-0.11	0.39
HH size four	0.11	0.30	-0.16	0.15	-0.16	0.23
HH size five	0.09	0.39	-0.12	0.27	-0.15	0.27
HH size six	0.13	0.24	-0.03	0.79	-0.16	0.26
HH size seven	0.07	0.59	-0.05	0.72	0.05	0.77
Female head	0.09	<.0001	0.05	0.00	0.04	0.07
Have kids 0-6	0.04	0.15	-0.18	<.0001	-0.05	0.14
Have kids 7-12	-0.08	0.00	-0.10	<.0001	-0.09	0.00

Appendix B continued

	Spinach		Slaw		Other Greens	
	Estimate	p-value	Estimate	p-value	Estimate	p-value
Have kids 13-18	-0.01	0.59	-0.08	0.00	-0.04	0.17
HH has dogs	-0.02	0.15	-0.03	0.02	-0.03	0.05
HH has cats	-0.02	0.10	-0.02	0.08	-0.02	0.25
White	-0.06	0.07	0.04	0.23	-0.12	0.00
Black	0.04	0.31	-0.06	0.11	0.42	<.0001
Asian	0.03	0.58	-0.36	<.0001	-0.02	0.73
Divorced	0.07	0.03	-0.02	0.44	0.09	0.03
Single	-0.01	0.78	-0.08	0.02	0.09	0.03
Married	0.12	0.00	0.06	0.15	0.08	0.09
Hispanic	-0.11	<.0001	0.17	<.0001	0.01	0.74
County size small	0.07	0.00	-0.10	<.0001	0.16	<.0001
County size medium	0.08	<.0001	-0.06	0.00	0.10	0.00
County size large	0.05	0.04	-0.06	0.01	0.03	0.29
Female employed > 35 hours	-0.04	0.02	-0.07	0.00	-0.09	0.00
Female homemaker / student	-0.03	0.10	0.03	0.09	-0.04	0.09
Male employed > 35 hours	0.03	0.29	-0.06	0.02	-0.01	0.74
Male homemeaker / student	-0.08	0.00	0.05	0.09	-0.03	0.31
Female ed less than high school	-0.05	0.38	0.15	0.02	0.05	0.51
Female ed high school	0.08	0.04	0.22	<.0001	0.00	0.95
Female ed some college	0.17	<.0001	0.22	<.0001	0.02	0.68
Female college graduate	0.23	<.0001	0.19	<.0001	0.08	0.07
Female ed post graduate	0.25	<.0001	0.21	<.0001	0.10	0.05
Male ed less than high school	-0.20	<.0001	-0.03	0.64	-0.12	0.06
Male ed high school	-0.16	<.0001	0.05	0.20	-0.15	0.00
Male ed some college	-0.03	0.43	0.10	0.02	-0.11	0.03
Male ed college graduate	0.00	0.90	0.07	0.10	-0.06	0.23
Male ed post graduate	0.06	0.17	0.03	0.47	0.03	0.53
Owns home	0.03	0.51	-0.01	0.79	0.00	0.94
Rents home	0.03	0.50	-0.13	0.01	0.05	0.40
Log price kits	-0.12	<.0001	-0.32	<.0001	-0.10	<.0001
Log price slaw	-0.11	0.00	-0.84	<.0001	-0.01	0.82
Log price other greens	-0.03	0.15	-0.16	<.0001	-0.99	<.0001
Log price spinach	-0.81	<.0001	-0.04	0.08	-0.04	0.15
Log price fresh processed	0.01	0.47	-0.04	0.04	0.11	<.0001
Log price garden	0.45	<.0001	0.04	0.00	0.46	<.0001
Log price unprocessed	0.01	0.46	0.02	0.23	0.07	0.00