8-31-2017

US Household Demand for Organic Fruit

Erik Nelson
Bowdoin College, enelson2@bowdoin.edu

John M. Fitzgerald
Bowdoin College, jfitzger@bowdoin.edu

Nathan Tefft
Bates College, ntefft@bates.edu

John L. Anderson
Bowdoin College, janderso@bowdoin.edu

US Household Demand for Organic Fruit

Erik Nelson
Department of Economics
Bowdoin College
enelson2@bowdoin.edu

John Fitzgerald
Department of Economics
Bowdoin College
jfitzger@bowdoin.edu

Nathan Tefft
Department of Economics
Bates College
ntefft@bates.edu

John L. Anderson
Department of Economics
Bowdoin College
janderso@bowdoin.edu

Abstract

We estimate US household monthly elasticities of demand for some of the more popular organic fruits. To our knowledge, this is the first US-wide, multi-year analysis of price and income elasticities for various organic fruits. We calculate elasticities of demand for low-income, middle class, and rich income bracket households using three estimation techniques: two econometric methods and one machine learning method (least absolute shrinkage and selection operator (LASSO)). Demand estimates are based on Nielsen scanner data from approximately 60,000 households collected from 2011 to 2013. Generally, we find that own-price conditional and unconditional elasticities of demand for organic fruits are negative. Unconditional elasticity magnitudes tend to be largest in the representative middle-class household. Income elasticities of demand measurements are inconsistent and often statistically insignificant. This finding is consistent with the survey literature finding that many consumers buy organic food for mostly moral or ethical reasons. We run two policy experiments: a 10% subsidy of organic fruits, and a 10% tax on conventional fruits. Our hypothetical policies engender a stronger reaction among the general public than habitual buyers of organic fruit; unconditional purchase and expenditure elasticities are generally larger than conditional purchase and expenditure elasticities. Finally, we find that elasticities measured with the LASSO technique are not radically different than those measured with econometric methods. The most noticeable difference between the two analytical techniques is that LASSO is more likely to find price and income elasticities of demand that indistinguishable from zero, both substantively and statistically.

JEL No. Q18, C01, C24, C34, C55, D10, D12

Keywords: organic fruit; elasticities of demand; econometrics; machine learning; Nielsen Consumer Panel data

1. Introduction

The consumption of organic food in the US has increased rapidly (USDA – ERS. 2016A). According to the US Organic Trade Association, organic food sales were $39.7 billion in 2015, a 165% increase since 2006.[1] As the Economic Research Service of the U.S. Department of Agriculture puts it: "Organic products have shifted from being a lifestyle choice for a small share of consumers to being consumed at least occasionally by a majority of Americans" (USDA – ERS. 2016A).

Given the increasing importance of organic food in US food markets, information on US price and income elasticities of demand for organic food is increasingly important. However, there are relatively few organic food elasticity estimates, especially in the US (e.g., Zhang et al. 2008).[2] Further, existing estimates are typically based on small consumer surveys. Here, we present US-wide estimates of price and income elasticities of demand for organic fruit for the years 2011 through 2013. We also estimate the impact that small subsidies for organic fruit or, alternatively, small taxes on conventional fruit, would have had on organic fruit purchases and expenditures during that time period. If organic fruit production, marketing, and consumption in lieu of conventional fruit production, marketing, and consumption create health and nonmarket environmental benefits, then subsidizing organic fruit purchases could enhance social welfare. We have chosen to focus on fruit because organic fruit is widely available in grocery stores across the US and is not limited to larger urban markets or farmers' markets.[3] In addition, organic produce, especially fruit, is the most popular organic food type. For example, produce accounted for 43% of U.S. organic food sales in 2012 (USDA – ERS 2016A; see Willer and Schaack (2015) for European data).

This paper contributes to the empirical food consumption literature in several important ways. First, to our knowledge, this is the first US-wide, multi-year analysis of price and income elasticities for various organic fruits. Second, we experiment with methods for estimating price and income elasticity of demands. Besides econometric techniques, we also use a machine learning (ML) technique known as least absolute shrinkage and selection operator (LASSO) to estimate elasticities. We use a large consumer dataset on household purchases for which the LASSO technique should be well suited. Many ML "methods are focused on finding patterns in data sets that are useful for prediction and classification" and are less concerned with "deriving asymptotic (large sample) results of the type that are common in econometrics" (p. 1, Athey and Imbens 2015). Therefore, the econometric and ML estimation techniques are analytically complementary as they have closely related, but distinct empirical goals. Similar estimates across all estimation techniques would suggest that our results are not a function of analytical choice or consumer theory fundamentals (e.g., symmetry in the price coefficients and homogeneity in prices and income). However, any differences in demand estimates across estimation methods can give us some insight into consumer behavior over organic fruit. For example, if the ML method estimates very different organic fruit demands than the econometric methods then the applicability of consumer theory fundamentals to organic fruit consumption behavior would not be supported by the data.

We have organized the rest of our paper as follows. In section 2 we provide additional background on household demand for organic food in the US. In section 3 we describe our data. In section 4 we describe our methods for estimating organic fruit price and income elasticities. In section 5 we describe our results, and in section 6 we present our conclusions.

2. Literature on household demand for organic food

2.1 Why do people pay the organic premium?

Organic foods command a price premium over their conventional counterparts (Fig. 1) (USDA – ERS 2016B).[4] Therefore, consumers that buy organics must believe that they provide benefits above and beyond their conventional analogs. Hughner et al. (2007) found that the majority of consumers believe that organic versions of food are more nutritious than their conventional analogs and that organics expose them to less pesticide residue (also see Huang 1996). In other words, many consumers are willing to pay the organic premium partly because they view it as an investment in their health (Yiridoe et al. 2005). However, organic certification codifies a production process and does not necessarily imply that organic products are safer or healthier than conventional products for consumers (Dimitri and Greene 2002, Winter and Davis 2006). Whether the organic process actually generates more nutritious food and reduces consumer exposure to dangerous chemicals is contested (Yirido et al. 2005, Pearson et al. 2011, Holzman 2012, Dangour et al. 2010, Smith-Spangler et al. 2012, Johansson et al. 2014, Lockeretz 2003, Crinnion 2010, Lairon 2010).

Beyond household health and wellness concerns, Hughner et al. (2007) also found that some consumers are willing to pay more for organics because they believe organic production is better for the environment than conventional production (Zepeda and Nie 2012, Jones et al. 2003, Hjelmar 2011, Dimitri and Lohr 2007). In this case, empirical evidence generally supports consumer belief (Reganold and Wachter 2016). For example, organic production is associated with lower irrigation requirements, better soil carbon sequestration rates (Gattinger et al. 2012), less energy use (El-Hage Scialabba 2013), and more biodiversity (Tuck et al. 2014) than conventional production.

Finally, some consumers are willing to pay more for organics because they believe their purchases support the creation of a healthier, more equitable, and more vibrant rural economy (Hughner et al. 2007). There is evidence that organic farming has some of these effects. For example, limited use of pesticides in organic farming significantly reduces illnesses and injuries obtained by agricultural workers (Winter and Davis 2006). In addition, organic farms in the US tend to be smaller and less capitalized, and therefore are a countervailing force against the overall trend of farm consolidation across the rural US. For example, in 2014, 12,595 farms that covered 3,642,933 acres were certified organic. This gives an average certified farm size of 289 acres. In contrast, there were 2,085,000 farms of all types in the US in 2014, with an average size of 438 acres (USDA-NASS 2015, USDA-NASS 2016). Finally, organic price premiums could translate into higher incomes for organic farmers despite the higher costs of production. For example, farm gate prices for organic carrot and broccoli were 21 to 489 percent greater than their conventional counterparts in the US from 1999 to 2007 (USDA-ERS 2016B). Further, in a survey of 44 studies, representing 55 crops grown in 14 countries on five continents, Crowder and Reganold (2015) found that gross returns, benefit/cost ratios, and net present values to organic crop production were 21%, 24%, and 35%, respectively, greater than the gross returns, benefit/cost ratios, and net present values to crop production.

2.2. Which households buy organics?

Consumers are not only heterogeneous in their motivation(s) for buying organic, but they also differ in how frequently they buy organic products. Smith et al. (2009) differentiates between devoted, casual, and nonuser organic consumers. Other classification systems include frequent versus occasional buyers (Janssen and Hamm 2012), market participants versus nonparticipants (Smed 2012), committed versus mainstream consumers (Hallam 2003), and light versus heavy users (Wier et al. 2008, Stevens-Garmon et al. 2007). Stevens-Garmon et al. (2007) found that light or casual buyers are most likely to purchase organic vegetables, while committed or devoted users are most likely to purchase organic fruits. Further, Zhang et al. (2008) found that in the US the conditional income elasticity of demand (an elasticity that only considers the behavior of organic market participants) for fresh organic produce was lower than the unconditional income elasticity of demand for the same food group. Zhang et al. (2008)'s findings suggest that organic produce is often treated as more of a necessity by US organic market participants or frequent buyers and as more of a luxury good by the representative US household.

Researchers have also identified household characteristics that are correlated with a higher likelihood of buying organic food. Urban households with more educated, older, and married heads of house and at least one child at home are more likely to buy organic produce (Kasteridis and Yen 2012, Smith et al. 2009, Zhang et al. 2008, Hughner et al. 2007, Thompson and Kidwell 1998, Dimitri and Greene, 2002, Monier et al. 2009).[5] Of particular interest to this study, Lin et al. (2009) found that a household is more likely to buy organic *fruit* if the heads of household are college educated and married. Further, households with higher incomes, all else equal, are more likely to buy organic products (Dimitri and Greene 2002, Zhang et al. 2008; Smith et al. 2009; Lin et al. 2009). Finally, Hispanic and black households are more likely to buy organic produce than white households, all else equal (Smith et al., 2009; Zhang et al., 2008; Stevens-Garmon et al., 2007).

Organic consumption is also explained by household location. First, the relative share of organic produce at farmer's markets is much larger than that of organic produce at grocery stores (USDA-ERS 2016B). Accordingly, farmer's market presence in a household's community tends to increase the probability that the household consumes organic produce (Zepeda and

Nie 2012). Several studies have concluded that western U.S. households are more likely to buy organic produce than households anywhere else, all else equal (Smith et al. 2009, Lin et al. 2009, Kasteridis and Yen 2012, Stevens-Garmon et al. 2007). A cursory analysis of our data on organic fruit expenditures supports this conclusion: on a per capita basis, from 2011 to 2013, five of the top six organic fruit expenditure markets were the west coast markets of San Francisco, Denver, Seattle, Sacramento, and Portland, Oregon (see Fig. 2 for gross expenditures by market). Whether this spatial pattern is due to the western population's greater exposure to the organic industry (the western US hosts the most organic produce packers, shippers, manufacturers, and distributors in the US (USDA-ERS 2016A)) or is due to lifestyle sorting is unclear.[6]

While research on which households are more or less likely to but organic produce is abundant, there is less research on how much people are winning to pay for organic products. Yiridoe et al. (2005) found that households were willing to pay more for organic produce with shorter shelf life, such as fruits and vegetables, than other organic foods. In addition, several researchers have found that more educated households and households with children are willing to pay more for organic produce, all else equal (Loureiro and Hine 2002, Batte et al. 2007, Rousseau and Vranken 2013). Of course, willingness to pay the organic premium only goes so far. For example, Yiridoe et al. (2005) found demand for organic food declined quickly once premiums reached a 20% threshold.

2.3. Previous price elasticity estimates

Previous efforts to measure the price elasticity of demand for organic produce have produced inconsistent results. A London-area study found own-price elasticities for organic fruit and vegetables were approximately three times higher than for conventional ones (Fourmouzi et al. 2012; also see Kasteridis and Yen 2012 for similar US results from 2006). An estimate of Dutch consumers' price elasticity of demand for several vegetables found the expected negative signs but never found elastic demand (Bunte et al. 2010). Using data from a large retail chain in the northeastern US, Bezawada and Pauwels (2013) also found that organic products have greater own-price elasticity than conventional products. Not all studies agree on this last point, however. For example, a study based on US scanner data from 1999 to 2003 found own-price elasticities were not always higher for organic vegetables than for conventional ones (Zhang et al. 2011).

As to cross-price elasticities between organic and conventional produce, Fourmouzi et al. (2012) found they were low, especially for devoted or frequent organic buyers. Bezawada and Pauwels (2013) found that reductions in organic product prices hurt conventional product sales more than vice versa. Further, the "results of Zhang et al. (2011, p. 453) regarding cross-price elasticities indicated that a decrease in organic price premiums would lead to a strong increase in the purchase of organic vegetables" (p. 14, Rödiger and Hamm 2015).

When it comes to income elasticities, Zhang et al. (2011) found positive income elasticities for organic potatoes, onions, tomatoes, and lettuce among US households, at least between 1993 and 2003. Organic potatoes and onions had income elasticity estimates greater than one. Finally, the four vegetable organic income elasticities were greater than the income elasticities for their conventional analogs.

3. Data

3.1. Consumer panel data

We use US data on organic and conventional fruit purchases and the households that made these purchases from 2011 through 2013 to estimate organic fruit demand functions. The demand functions are explained by variables associated with the propensity to buy and the willingness to pay for organic produce as reviewed in section 2. From these demand functions, we estimate organic fruit price and income elasticities of demand.

Almost all the data comes from the Nielsen Corporation's Consumer Panel Data. Each year Nielsen recruits approximately 60,000 US households to record, if possible, each purchase they make.[7] Using an in-home scanner, Nielsen households record the place and date of each item they purchase. Each purchased item is either assigned the store's listed price for that item the week it was purchased (if the store participates in the Nielsen program) or the participating households record the prices manually. Purchase data from sampled households, as well as their demographic information, can be projected to market and national levels using projection factors assigned to each household (Kilts Nielsen Center 2014, Cotti et al. 2016).

We use only a subset of the data from the annual Consumer Panel datasets. First, we flagged all household shopping trips that included a purchase of fruit in 2011, 2012, and 2013 datasets where i indexes fruit type × variety (read as fruit type by variety, i.e. organic apple, conventional apple, etc.), k indexes households, and d indicates the date of the incident. Fruit type is coded as organic if it had the U.S. Department of Agricultural organic label; otherwise fruit type is coded as conventional. For each fruit purchase incident in the set of flagged trips we recorded the dollar expenditure on fruit type × variety i, given by e_{ikd}, and ounces or number of i bought, given by o_{ikd} or t_{ikd}, respectively.[8] We converted all fruit purchase incidents recorded in number of items bought to ounces bought – in other words, we converted each t_{ikd} to o_{ikd} – using representative fruit weights. For example, if household k bought 2 organic apples on a day d trip (t_{ikd} = 2 where i indicates organic apples) then o_{ikd} = 2 x 6.42 ounces = 12.84 ounces of organic apples where 6.42 ounces is the weight of a typical apple (see Appendix Table A for a list of assumed fruit weights).

Next, we summed e_{ikd} and o_{ikd} across all organic i other than organic apples, blueberries, oranges, and strawberries to create e_{ikd} and o_{ikd} for i = organic × other. Similarly, we summed e_{ikd} and o_{ikd} across all conventional i other than conventional apples, blueberries, oranges, and strawberries to create e_{ikd} and o_{ikd} for conventional × other. After this aggregation, our database contained fruit purchase incident data for ten type × variety fruits, five organic and five conventional.

Then we aggregated fruit purchase incidents within months to create a database of households' purchases of and expenditures on all ten type × variety fruits in each month of 2011 through 2013. Specifically, household k's ounces bought and expenditures on type × variety fruit i in month m is given by $o_{ikm} = \sum_{d \in m} o_{ikd}$ and $e_{ikm} = \sum_{d \in m} e_{ikd}$, respectively. If household k did not purchase type × variety i in month m then we set o_{ikm} = 0 and e_{ikm} = 0.

Then we calculated $p_{ikm} = e_{ikm} / o_{ikm}$, the price of i faced by household k in month m, for each unique ikm combination. We imputed p_{ikm} if e_{ikm} = 0 and o_{ikm} = 0. See Appendix Sections 1 and 2 for details on the calculation of e_{ikm}, o_{ikm}, and price imputation. Let $\mathbf{P}_{km}$ indicate the vector of ten real fruit prices (Dec, 2013 $) that household *k* faced in month *m*.

Let $\mathbf{X}_{km}$ be household *k*'s vector of characteristics in month *m*. The vector contains the household variables that previous research, as described in section 2, has flagged as affecting propensity to buy and overall demand for organic produce, including the household's monthly real income (Dec, 2013 $), household size, whether or not the household contains one or more children under 18, whether at least one head of household has a college degree, whether the household is headed by a married couple, and the racial makeup of the household (see Appendix Table B for a complete list of variables in $\mathbf{X}_{km}$). In a separate vector of independent variables, given by $\mathbf{C}_{km}$, we control for season × year interactions (e.g., winter 2011, spring 2011, etc.), whether or not the household is in a metro or non-metro county, and the Nielsen Scantrack market the household is in (see Appendix Table C for a complete list of variables in $\mathbf{C}_{km}$). For estimation purposes, we also define $\mathbf{c}_{km} \subset \mathbf{C}_{km}$ where $\mathbf{c}_{km}$ only includes the season × year interaction dummy variables (see Appendix Table D for a complete list of variables in in $\mathbf{c}_{km}$).

Household information for the year y Consumer Panel is current as of late year y – 1 (year y Consumer Panel household data is collected from October through December of year y – 1). For household income, panelists are asked to report their combined total household annual income at the end of the previous calendar year. For example, households that are part of the 2011 panel are surveyed in the fall of 2010 about their total annual income at the end of 2009. Nielsen believes panelists are actually reporting their "annualized" estimated income as of late 2010 and are not referring to previous year tax returns (Kilts Nielsen Center 2014). In any case, a household's actual income in year y may be a bit different than its recorded year y income value.

Household information for year y is not updated in year y's panel if household status changes at some point in year y. Therefore, household *k*'s values in $\mathbf{X}_{km}$ and market identification dummy variables in $\mathbf{C}_{km}$ are the same across all months of year y.[9] Households that are in the Consumer Panel for multiple years have the opportunity to update their characteristics before the beginning of each new year. Because household circumstances can change before they are updated in the panel, some household fruit purchases will be associated with an “outdated” version of the household. In other words, we expect our results are affected by some unavoidable measurement error. See Appendix Section 6 for a few more observations on the dataset we use to estimate organic fruit demand.

3.2. Trends in 2011-2013 US household organic fruit purchases

Only considering the organic fruit purchases made through traditional retail outlets (e.g., grocery stores, Walmart, etc.), US household expenditures on organic fruits increased by 46.1%, from $144.82 million in 2011 to $211.53 million in 2013 (December, 2013 dollars). Over the same time period, US household expenditures on conventional fruits only at traditional retail outlets increased by 6.9% (see the “Total Expenditures (M $)-All” row in Table 1). In total, households in the low income bracket (a household income 130% or less of the poverty line conditional on year and household size) increased their expenditures on organic fruits the most during this time period, 68.7%, compared to 34.5% (middle class) and 51.6% (rich). However, on a per household level, rich households (household income greater than 500% of the poverty line conditional on year and household size) not only bought more organic fruit than typical low income and middle class households, they also experienced the greatest growth in organic fruit purchases between 2011 and 2013, 62.5% versus 50.8% (poor) and 31.9% (middle class) (see the “Expenditures / Household” rows in Table 1).

The 2011-2013 growth in household-level organic fruit purchases from traditional retail outlets occurred at both the extensive and intensive margins. Higher purchases on the extensive margin occurred across all three income brackets (Table 2). Among all US households, the number of “organic fruit-only” households[10] and “both varieties of fruit” households[11] increased by 80.0% and 35.7%, respectively, between 2011 and 2013. Conversely, “conventional fruit-only” households fell by 4.5%. Of the three income brackets, the percentage of middle class households participating in the organic fruit market via traditional retail outlets, either as “organic fruit-only” or “both varieties of fruit” households, expanded the most between 2011 and 2013 (Table 2). As to the intensive margin, of the US households that were represented in the Consumer Panels in 2011 and 2013, real expenditures on organic fruit from traditional retail outlets was 47.1% higher in 2013 compared to 2011.

The spatial concentration of organic fruit purchase was more intense than that of conventional fruit. Consumers in the top six markets (out of 52) for organic fruit expenditures at traditional retail outlets during the 2011 to 2013 period, San Francisco, Los Angeles, Seattle, Boston, Denver, and Chicago in descending order, were responsible for a third of all organic fruit purchases (Fig. 2). Further, consumers in the top eleven markets for organic fruit expenditures at traditional retail outlets during the 2011 to 2013 period were responsible for a half of all organic fruit expenditures. Conversely, households in the top eight markets for *conventional* fruit purchases at traditional retail outlets during the 2011 to 2013 period were responsible for a third of all conventional fruit purchases. Further, households in the top fourteen markets for conventional fruit purchases during the 2011 to 2013 period were responsible for a half of all conventional fruit purchases (Appendix Fig. A)

The pattern of organic fruit purchases across households ordered by income *within* markets was also skewed. In the most Scantrack markets, during the 2011 to 2013 period, households in the lower 50th percentile of income spent more than their proportionate share on organic fruit from traditional retail outlets relative to households in the upper 50th percentile (Fig. 3). Interestingly, the top five organic fruit markets in terms of gross expenditures during the 2011 to 2013 period, San Francisco, Los Angeles, Seattle, Boston, and Denver, were outliers when it came to within market expenditure patterns. These five markets had nearly proportional spending on organic fruits across their income spectrums. The markets with the most uneven distribution of organic fruit expenditures across household income tended to be the smaller markets for organic fruit. For example, some of the smaller organic fruit markets (as measured by gross expenditures), including Miami, Minneapolis, St. Louis, Syracuse, Richmond, Oklahoma City-Tulsa, Little Rock, Jacksonville, and Omaha, had some of the most uneven distributions of organic fruit expenditures, with households in the lower income percentiles spending much more on organic fruit than their proportional share. It is likely that these uneven expenditure patterns were explained by the well-to-do's tendency to spend more at restaurants and less on nondurable goods like groceries relative to other household types (The

JPMorgan Chase Institute 2016), meaning households from the upper 50th percentile had less opportunity to buy fruit, conventional or organic, from stores covered in our data. This hunch is supported by the fact that even conventional fruit expenditures in *all* 52 markets during this period were skewed towards lower income percentile households (Appendix Fig. A).

4. Price and income elasticity of demand estimation

In this paper, we use three techniques, two econometric-based and one machine learning-based, to estimate price and income elasticities for organic fruit during the 2011 to 2013 period. We also use the Consumer Panel data to predict how organic fruit expenditures would have changed if organic fruit had been subsidized or conventional analogs had been taxed. We do all of this for three household income groups: households that have an annual income 130% or less of the poverty line (low-income households); households with annual incomes between 130% and 500% of the poverty line (middle income households); and households with annual incomes greater than 500% of the poverty line (rich households).[12]

We estimate demand for organic fruit for several reasons. First, as noted in the introduction, the US market for organic fruit is growing rapidly (see Tables 1 and 2), yet recent demand elasticities for such fruit are not known. Second, there have been calls to subsidize relatively expensive organic produce 1) to reduce the impact of agriculture on the environment and promote sustainable agriculture, 2) to increase access to healthy produce (under the assumption that organic food is healthier than conventional analogs), and 3) to make organic produce more affordable for low-income Americans.[13] A typical low income household purchases much less organic fruit, for example, than their middle- and rich-income counterparts (see Tables 1 and 2). Whatever the reason for an organic subsidy and wherever the subsidy's incidence, the policy is likely to reduce the market price for organics. We determine whether low income households would benefit the most from a subsidy of organic fruit or if the bulk of the consumer surplus created by lower organic fruit prices would go to middle class and rich households. For illustrative purposes, we use an organic fruit subsidy of 10% and alternatively, a conventional fruit tax rate of 10%.[14]

4.1 Elasticity Estimation Assumptions

We assume that changes in fruit prices during 2011 to 2013 as observed in the data were mostly driven by changes in seasonal and annual market supplies and not changes in annual and seasonal market demands. We have several reasons to believe this. First, the real incomes of most households that remained in the Consumer Panel all three years barely changed between the 2011 and 2013 panels (although recall household income data is lagged in the dataset) (Fig. 4). Further, U.S. Census Bureau data also indicates zero growth in median household income during this period; real median household income in the US only began to increase steadily again *after* 2013 (U.S. Bureau of the Census 2017). In other words, one of the major determinants of market demand in our data did not change much, at least in the aggregate or representative level, from 2011 through 2013.

Second, we cannot recall any major changes in organic food policy or the organic market between 2011 and 2013 that would have engendered a shock in demand for organic produce. In 2014 Walmart and Target announced that they would greatly expand their organic offerings. In all likelihood, these changes induced greater organic demand among their many customers. Again, however, this change occurred after the period reflected in our data.

In contrast, there were major increases in organic fruit supply from 2011 through 2013 (Table 3). For example, compared to 2011, 2014 US production of organic apples, blueberries, and strawberries was 231%, 74%, and 49% greater, respectively (organic orange production fell by 2% over this time period).[15] Further, compared to 2011, in the 2014 the value of organic apple, blueberry, and strawberry imports was 419%, 114%, and 212% greater, respectively (organic orange import data are not available).

If organic fruit was sold in competitive markets between 2011 and 2013, our assumptions would imply that organic fruit market prices decreased between 2011 and 2013. In other words, if seasonal market demand stayed fixed during the 2011 to 2013 period but seasonal market supplies increased over these years then seasonal prices would have to have fallen in order to clear markets. However, organic products are not traded in competitive markets. Substantial price premiums for organic produce provide evidence of imperfect competition (Crowder and Reganold 2015). Given imperfect competition, increases in organic fruit supplies and steady demand during the 2011 to 2013 period, holding seasonal effects fixed, may or may not have led to decreases in prices. In other words, constant or even small increases in prices, holding seasonal effects fixed, over time would not necessarily be inconsistent with our market demand-market supply assumptions. Interestingly, we do find many instances where organic fruit prices were lower in 2012 and 2013 compared to 2011 (Table 4). In the case of strawberries and blueberries, 2012 and 2013 seasonal prices were markedly lower than same-season 2011 prices in the seasons when these two berries were most heavily consumed. Lower prices concurrent with a surge in quantity purchased are particularly suggestive of an increase in market supply and relatively steady demand.

Given our market demand and supply assumptions, it is unnecessary to instrument for fruit prices in our organic fruit demand models, thus simplifying the analysis. This means avoiding the complications of combining instrumental variable and demand system estimation (our second estimation method). Further, our working assumption means we do not need to derive a whole new technique that modifies LASSO to estimate organic fruit demand (our third estimation method).

4.2. Estimation method 1: Individual fruit Heckman models of consumption

In our first estimation method, we econometrically parameterize unconditional and conditional demand for each fruit type × variety *i* and household income class *z* combination *separately*. Conditional demands for each *i*,*z* combination are estimated over all household-month observation where $e_{ikm} > 0$ and household-month *km* belongs to class *z*. Unconditional demands for each *i*,*z* combination are estimated over all household-month observations that belong to class *z*. In this estimation method, household-month's demand for fruit type × variety *i* is comprised of a selection equation (eq. 1) and a latent demand equation (eq. 2)[16]:

$$o_{ikm} = \begin{cases} o_{ikm}^{*} & \text{if } \mathbf{X}_{km}\boldsymbol{\alpha}_i + \mathbf{C}_{km}\boldsymbol{\mu}_i + \mathbf{P}_{km}\boldsymbol{\omega}_i + v_{ikm} > 0 \\ 0 & \text{if } \mathbf{X}_{km}\boldsymbol{\alpha}_i + \mathbf{C}_{km}\boldsymbol{\mu}_i + \mathbf{P}_{km}\boldsymbol{\omega}_i + v_{ikm} \leq 0 \end{cases} \tag{1}$$

and

$$o^{*}_{ikm} = \mathbf{X}_{km}\boldsymbol{\beta}_i + \mathbf{c}_{km}\boldsymbol{\sigma}_i + \mathbf{P}_{km}\boldsymbol{\theta}_i + \varepsilon_{ikm} \quad (2)$$

where o_{ikm}, $\mathbf{X}_{km}$, $\mathbf{c}_{km}$, $\mathbf{C}_{km}$, and $\mathbf{P}_{km}$ are defined above, $\boldsymbol{\alpha}_i$, $\boldsymbol{\mu}_i$, $\boldsymbol{\omega}_i$, $\boldsymbol{\beta}_i$, $\boldsymbol{\sigma}_i$,and $\boldsymbol{\theta}_i$ are fruit-specific model coefficients to be estimated, and v and ε are error terms.[17] In this estimation method it is as if the fruit buying-household chooses how much fruit to buy in a month, one by one, with no explicit budget constraint and no recognition of its monthly demand for other fruit. In other words, no consistent system of consumer behavior across goods is assumed in this estimation method. According to the system of equations (1) and (2) the market type and location affects the probability of purchase, but not the quantity purchased for those who buy, conditional on price, income and demographics. We assume that geographic location affects seasonal availability but then quantity demanded is only determined by price and household characteristics.

To estimate (1)-(2) for household income class z we first use a probit to parameterize a discrete choice equation where $Y_{ikm} = 1$ if household-month $km \in z$ bought fruit type × variety i (i.e., if $o_{ikm} > 0$) and equals 0 otherwise,

$$\text{Prob}(Y_{ikm} = 1) = \text{Prob}\left(\mathbf{X}_{km}\boldsymbol{\alpha}_i + \mathbf{C}_{km}\boldsymbol{\mu}_i + \mathbf{P}_{km}\boldsymbol{\omega}_i + \varepsilon_{ikm} > 0\right) =$$

$$\Phi\left(\mathbf{X}_{km}\boldsymbol{\alpha}_i + \mathbf{C}_{km}\boldsymbol{\mu}_i + \mathbf{P}_{km}\boldsymbol{\omega}_i\right) \quad (3)$$

Let $\hat{\lambda}_{ikm}$ be the estimated inverse Mills ratio for fruit type × variety i in household-month $km \in z$:

$$\hat{\lambda}_{ikm} = \phi\left(\mathbf{X}_{km}\hat{\boldsymbol{\alpha}}_i + \mathbf{C}_{km}\hat{\boldsymbol{\mu}}_i + \mathbf{P}_{km}\hat{\boldsymbol{\omega}}_i\right) / \Phi\left(\mathbf{X}_{km}\hat{\boldsymbol{\alpha}}_i + \mathbf{C}_{km}\hat{\boldsymbol{\mu}}_i + \mathbf{P}_{km}\hat{\boldsymbol{\omega}}_i\right) \quad (4)$$

where ϕ and Φ indicate the standard normal probability and cumulative density functions, respectively and $\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\mu}}_i,$ and $\hat{\boldsymbol{\omega}}$ are estimated coefficient vectors. The conditional expectation for ounces of fruit type × variety i bought by household-month $km \in z$ is given by,

$$E\left[o_{ikm} \mid Y_{ikm} = 1\right] = \mathbf{X}_{km}\boldsymbol{\beta}_i + \mathbf{c}_{km}\boldsymbol{\sigma}_i + \mathbf{P}_{km}\boldsymbol{\theta}_i + \gamma_i \hat{\lambda}_{ikm} \tag{5}$$

where $\boldsymbol{\beta}_i, \boldsymbol{\sigma}_i$, and $\boldsymbol{\theta}_i$ are estimated with OLS over the subset of household-month $km \in z$ where $Y_{ikm} = 1$. Finally, the unconditional expectation for ounces of fruit type × variety *i* bought by household-month $km \in z$ is given by,

$$E\left[o_{ikm}\right] = P(Y_{ikm} > 0)E\left[o_{ikm} \mid Y_{ikm} = 1\right] + P(Y_{ikm} = 0)0 \tag{6}$$

$$= \Phi\left(\mathbf{X}_{km}\boldsymbol{\alpha}_i + \mathbf{C}_{km}\boldsymbol{\mu}_i + \mathbf{P}_{km}\boldsymbol{\omega}_i\right)\left(\mathbf{X}_{km}\boldsymbol{\beta}_i + \mathbf{c}_{km}\boldsymbol{\sigma}_i + \mathbf{P}_{km}\boldsymbol{\theta}_i\right) + \gamma_i \phi\left(\mathbf{X}_{km}\hat{\boldsymbol{\alpha}}_i + \mathbf{C}_{km}\hat{\boldsymbol{\mu}}_i + \mathbf{P}_{km}\hat{\boldsymbol{\omega}}_i\right) \tag{7}$$

where $\boldsymbol{\beta}_i, \boldsymbol{\sigma}_i$, and $\boldsymbol{\theta}_i$ are estimated with OLS over *all* household-months $km \in z$ and the term $\phi\left(\mathbf{X}_{km}\hat{\boldsymbol{\alpha}}_i + \mathbf{C}_{km}\hat{\boldsymbol{\mu}}_i + \mathbf{P}_{km}\hat{\boldsymbol{\omega}}_i\right)$ is the same as the numerator of equation (4). Because we are primarily interested in organic fruit demand we skip estimating (5)-(7) for the conventional varieties of *i*.

Finally, from equations (5) and (7) we derive the formulas for organic fruit elasticities allowing for sample selection as given in Byrne, Capps, and Saha (1996) and Saha, Capps and Byrne (1997). We compute 1) conditional and unconditional price elasticity of demand ($CPED_{ij}$ and $UPED_{ij}$, respectively), 2) conditional and unconditional income elasticity of demand ($CIED_i$ and $UIED_i$, respectively), 3) purchase probability elasticity with respect to price ($PPEP_{ij}$), and 4) purchase probability elasticity with respect to income ($PPEI_i$) for each organic fruit type *i* and household income class *z* combination. These elasticities are calculated at each z's mean $\mathbf{X}_{km}$, $\mathbf{C}_{km}$, $\mathbf{c}_{km}$, and $\mathbf{P}_{km}$ values (Appendix Table E). Further, the elasticity measures are generated assuming the Boston market intercept. Please note that we report elasticity standard errors without adjusting for the use of predicted parameters in equations (5) and (7). See Appendix Section 3 for more on the estimation of equations (3)-(7) and the derivation of elasticities.

4.3. Estimation method 2: incomplete demand system of consumption

In our first approach, we assumed that expenditures on fruit type $\times$ variety *i* was exogenous to other fruit purchases and household budget. However, this many not be the case. Instead, consumers may allocate a portion of their income over a joint purchase of several fruit types and varieties. If this latter narrative better represents actual consumer behavior then we should estimate fruit purchases with a demand system where all fruit expenditures are determined jointly. Convention would suggest the use of the Almost Ideal Demand System (AIDS) or some similar demand system structure (e.g., QUAIDS) to estimate price, income, and purchase probability elasticities (e.g., Zhang et al. 2011). However, two features of monthly-level fruit consumption make demand system estimation with AIDS or similar demand system structures difficult. First, fruit consumption represents a small fraction of a household's monthly expenditures at traditional retail outlets, and second, in any given month, a typical household will not buy any organic fruit from a conventional retail outlet (i.e., $o_{ikm} = 0$ for many observations). AIDS and other similar demand systems require data on shares of expenditures across all categories of consumer goods (we only calculated household expenditures on fruit) and work best with few zeros observations.

To overcome the incomplete expenditure shares data problem we adopt a demand system technique developed by LaFrance (1990) and LaFrance and Hanemann (1989). This demand system is designed to work with incomplete expenditure shares data. This system develops demand from a well-defined expenditure function and imposes several consumer theory restrictions, including homogeneity in prices and Slutsky substitution matrix symmetry.[18]

To avoid the bias introduced by frequent dependent variable observations of zero we use a first stage selection and associated latent demand equation combination similar to the one we used in estimation method one (eqs. 1 and 2). However, in this case the consumer demand restrictions suggested by consumer theory are imposed on the latent demand parameters (i.e., the parameters Shonkwiler and Yen 1999).[19] In other words, this estimation method's latent demand functional form is different than the functional form of the latent demand equation from estimation method 1. Therefore, the differences in elasticity estimates between this econometric approach and the one described in section 4.2 will be due to the behavioral structure imposed by the demand system's consumer theory.

The derivation of an incomplete demand system we use with this estimation method is known as the LinQuad system (Fabiosa and Jensen 2003). The latent demand for o_{ikm} within the LinQuad system is,

$$o_{ikm} = \delta_i + \mathbf{X}_{km}\boldsymbol{\beta}_i + \mathbf{c}_{km}\boldsymbol{\sigma}_i + \mathbf{P}_{km}\boldsymbol{\theta}_i +$$

$$\boldsymbol{\Psi}_i[s_{km} - \mathbf{P}_{km}\boldsymbol{\delta} - \mathbf{X}_{km}\boldsymbol{\beta}\mathbf{P}'_{km} - \mathbf{c}_{km}\boldsymbol{\sigma}\mathbf{P}'_{km} - 0.5\mathbf{P}_{km}\boldsymbol{\theta}\mathbf{P}'_{km}] + \varepsilon_{km} \qquad (8)$$

where, as before, o_{ikm} is the ounces of fruit type × variety *i* bought in month *m* by household *k* in income class *z*, $\mathbf{X}_{km}$ is a vector of household *k*'s characteristics in month *m*, $\mathbf{c}_{km}$ is vector of year × season interaction dummy variables, $\mathbf{P}_{km}$ is a vector of fruit prices faced by household *k* in month *m*, and, in this case, s_{km} refers to household monthly income. The budget constraint, in the bracket of equation (8), limits expenditure on fruit type × variety *i* in month *m* by household *k* to be less than or equal to its monthly income s_{km}.

We estimate equation (8) for each fruit type × variety *i* (less organic-other and conventional-other) jointly.[20] In other words, unlike estimation method 1, where we estimate the i^{th} vector of $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$, and $\boldsymbol{\theta}$ one organic fruit at a time, here we estimate the matrices $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$, and $\boldsymbol{\theta}$ in one fell swoop, for organics and conventionals alike. Specifically, the LinQuad system use a seemingly unrelated regression approach (we allow for correlation in the regression errors across equations) to estimate $\boldsymbol{\delta}$, $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$, $\boldsymbol{\theta}$, and $\boldsymbol{\Psi}$ with imposed symmetry in the price coefficients[21] and homogeneity in prices and income.[22] Note that this second econometric method has more parameter matrices than the first method, $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ for price/demographic interactions and $\boldsymbol{\theta}$ for quadratic price and cross-price effects .

As before (eqs. 6-7), the conditional and unconditional expectations for ounces of fruit type × variety *i* bought by a household over the course of a month is given, respectively, by,

$$E\left[o_{ikm} \mid Y_{ikm} = 1\right] = \mathbf{X}_{km}\boldsymbol{\beta}_i + \mathbf{c}_{km}\boldsymbol{\sigma}_i + \mathbf{P}_{km}\boldsymbol{\theta}_i + \gamma_i \hat{\lambda}_{ikm} \qquad (9)$$

$$E\left[o_{ikm}\right] = \hat{\Phi} f\left(\mathbf{X}_{km}\boldsymbol{\beta}_i, \mathbf{c}_{km}\boldsymbol{\sigma}_i, \mathbf{P}_{km}\boldsymbol{\theta}_i\right) + \gamma_i \hat{\phi} \qquad (10)$$

where $\hat{\Phi}$ is estimated probability (3), $\hat{\phi}$ is the estimated numerator of equation (4) (Shonkwiler and Yen 1999), and $f\left(\mathbf{X}_{km}\boldsymbol{\beta}_i, \mathbf{c}_{km}\boldsymbol{\sigma}_i, \mathbf{P}_{km}\boldsymbol{\theta}_i\right)$ is short-hand for equation (8).

We calculate income class *z*'s *CPEDs*, *UPEDs*, *CIEDs*, *UIEDs*, *PPEPs*, and *PPEIs* at their mean $\mathbf{X}_{km}$, $\mathbf{C}_{km}$, $\mathbf{c}_{km}$, and $\mathbf{P}_{km}$ values (Appendix Table E). Further, the elasticity measures are generated assuming the Boston market intercept. Finally, again, just as in the first econometric approach, we report standard errors without adjusting for the use of predicted parameters for the CDF and PDF in the system estimation. See Appendix Section 4 for more on the estimation of the LinQuad elasticities.

4.4. Estimation method 3: Inferring consumer demand from Machine Learning

Unlike the econometric approaches we employ, machine learning (ML) techniques are not based on any economic theory. Instead these techniques use various heuristic algorithms to find the subset of independent variables that best predict the outcome of interest. Therefore, ML offers comparatively unstructured estimates where the "data" determines the final set of predictors (the universe of *possible* explanatory variables is the only place for economic intuition in these methods). Not only are ML heuristics not beholden to any economic theory but they do not make any of the data generating assumptions used in typical least squares estimation. Like other recent work on ML in the field of economics, we are interested to find out if an estimation process focused on model fit by way of careful variable and interaction selection (i.e., ML approaches) performs equally well to, if not better than, structural models, e.g. other models described in this paper, that could be mis-specified,.

To perform this comparison, we estimate a LASSO model over our data. Among the universe of ML models, the LASSO is one of the closest in structure to the econometric models we use. Because LASSO assumes no statistical structure we expect that it will most accurately represent the overall relationships found in the data at the cost of (a) some potentially biased coefficients,[23] (b) some less precise estimates, and (c) the inability to conduct welfare analysis made possible by counterfactual manipulation in a model based on a utility function. Further, our LASSO results could be ill-suited for policy analysis. For example, if the LASSO algorithm assigns coefficient values of zero to some organic and conventional price variables then analyses of the impact of organic price subsidies on organic fruit consumption is likely to be unreliable unless the prices of fruits with coefficients of zero are truly irrelevant to average consumer decision-making.

4.4.1. LASSO estimation

LASSO is a linear model of predictors estimated by minimizing the sum of squares plus a shrinkage penalty based on the sum of the absolute values of the model coefficients (Efron et al. 2004). The shrinkage penalty is multiplied by a tuning parameter, which adjusts the severity of the shrinkage penalty, and this parameter is set to minimize the above objective function using 10-fold cross validation. When estimating probabilities, the LASSO implements a log-odds regression model, and when estimating quantities, the LASSO implements a linear regression model.

The LASSO models are estimated using the **glmnet** package in R (Friedman et al. 2010). It takes two steps to estimate demand for fruit type × variety *i* in income group z using the LASSO model. In the first step, we find the LASSO coefficients that best explain whether or not a household-month *km*∈*z* purchased fruit type × variety *i*. This is done by maximizing the log likelihood function of a linear logistic regression model,

$$\max_{\boldsymbol{\pi}_i} \frac{1}{N_z}\left[\sum_{km=1}^{N_z}\left\{I(g_{km}=0)\log p(\mathbf{W}_{km}) + I(g_{km}=1)\log\left(1-p(\mathbf{W}_{km})\right)\right\}\right.$$

$$\left. -\lambda_i \sum_{j=1}^{p}\left|\pi_{ij}\right|\right] \qquad (11)$$

where N_z is the set of all *km*∈*z*, $\mathbf{W}_{km} = [\mathbf{X}'_{km}\mathbf{C}'_{km}\mathbf{P}_{km}]$ is a vector of *p* standardized[24] candidate predictors for the binary consumption of fruit type × variety *i,* the vector $\boldsymbol{\pi}_i$ contains independent variable coefficients for fruit type × variety *i* purchases by household-months *km*∈*z*, and λ_i is the tuning parameter, as described above. Further, $I(g_{km}=0)$ indicates a *km*∈*z* observation where fruit was not purchased, $I(g_{km}=1)$ indicates a *km*∈*z* observation where fruit was purchased, $p(\mathbf{W}_{km}) = \frac{1}{1+\exp(-\mathbf{W}_{km}\boldsymbol{\pi}_i)}$ is the probability that fruit type × variety *i* is not purchased by $km \in z$, and $1 - p(\mathbf{W}_{km})$ is the probability that fruit type × variety *i* is purchased $km \in z$. The shrinkage penalty in equation (11) is a kinked function of $\boldsymbol{\pi}_i$. Therefore, the LASSO tends to set some model coefficients to zero. The standard errors for estimates of π_{ij} are bootstrapped by replicating the joint solution to equations (11) 100 times. In each replicate, the sample is randomly drawn with replacement, so while the sample size is the same each time, the households represented in the dataset are different in each replication.

After estimating equation (11) for each fruit type × variety *i* and income class combination, we calculate $\hat{\rho}_{ikm} = \frac{1}{1+\exp(-\mathbf{W}_{km}\hat{\boldsymbol{\pi}}_i)}$ for each $km \in z$ across all fruit type × variety and income class combinations. Then we calculate $\hat{\rho}_{ikm}$ for each $km \in z$ across all fruit type × variety and income class combinations again given a 10% increase in the price of fruit type × variety *j*, holding all other variables constant. We calculate LASSO purchase probability elasticities with respect to price ($PPEP_{ij}$) for each fruit type × variety *i* and income class combination by comparing mean $\hat{\rho}_{ikm}$ values derived with the observed price data to mean $\hat{\rho}_{ikm}$ values derived with 10% increase in the price of fruit type × variety *j*, all else equal. We calculate LASSO purchase probability elasticities with respect to income ($PPEI_{ij}$) for each fruit type × variety *i* and income class combination similarly except we use a 10% increase in km's household income to derive the alternative set of $\hat{\rho}_{ikm}$ values.

Please note the primes on the vectors $\mathbf{X}_{km}$ and $\mathbf{C}_{km}$ in the candidate predictor vector $\mathbf{W}_{km}$ from equation (11) ($\mathbf{P}_{km}$ is as before). The primes on $\mathbf{X}_{km}$ and $\mathbf{C}_{km}$ indicate that these household-month variable vectors are different than the $\mathbf{X}_{km}$ and $\mathbf{C}_{km}$ variable vectors used with estimation methods 1 and 2 in two ways. First, the $\mathbf{X}_{km}$ and $\mathbf{C}_{km}$ vectors used in estimation methods 1 and 2 are comprised of a set of author-selected household-month and market variables. In contrast, $\mathbf{X}'_{km}$ and $\mathbf{C}'_{km}$ include all of the household-month variables in the Consumer Panel dataset. Second, the household-month variables in $\mathbf{X}_{km}$ and $\mathbf{C}_{km}$ are simplified representations of more complex raw data; in $\mathbf{X}'_{km}$ and $\mathbf{C}'_{km}$ we use the more complex data as is. For example, in the Consumer Panel dataset a household-month in year y is placed into one of nine categories regarding the number and mix of children in the household. In $\mathbf{X}_{km}$ this variable was reduced to a binary variable that indicated whether the household had one or more children or not. In

$\mathbf{X}'_{km}$ all nine children categories are potential predictors. Further, in the $\mathbf{C}_{km}$ vector used in estimation methods 1 and 2 the rural-urban continuum code (RUCC) categories are reduced to a dummy representation.[25] In $\mathbf{C}'_{km}$ all 7 RUCC categories are used. See Appendix Tables F and G for descriptions of the candidate predictors in matrix **W**.

The second step in estimating demand for fruit type × variety *i* in income group *z* with the LASSO model involves minimizing the following log-likelihood function,

$$\min_{\boldsymbol{\vartheta}_i} \left[\frac{1}{T_z}\sum_{km=1}^{T_z}(o_{ikm} - \mathbf{w}_{km}\boldsymbol{\vartheta}_i)^2 + \lambda_i \sum_{j=1}^{p}\left|\vartheta_{ij}\right|\right] \tag{12}$$

where $\mathbf{w}_{km} = [\mathbf{X}'_{km}\mathbf{P}_{km}\ \hat{\rho}_{ikm}]$ is a matrix of *p* standardized candidate predictors for the consumption of fruit type × variety *i*, vector $\boldsymbol{\vartheta}_i$ contains linear model coefficients for fruit type × variety *i*, and λ_i is the tuning parameter, as described above. Note that we use *km*∈*z*'s predicted probability (propensity score) of purchasing the relevant fruit type × variety *i* as a control in equation (12). The standard errors for estimates of ϑ_{ij} (and the demand elasticities discussed below) are bootstrapped by replicating the joint solution to equations (12) 100 times. In each replicate, the sample is randomly drawn with replacement, so while the sample size is the same each time, the households represented in the dataset are different in each replication. See Appendix Section 5 for more on the LASSO estimations.

Solving equation (12) for fruit type × variety *i* over the set of *km*∈*z* where $o_{ikm} > 0$ (i.e., the set $T_z = n_z$), given by $\widetilde{\boldsymbol{\vartheta}}_i$, allows us to produce the conditional demand for *i* among *km*∈*z*, given by $\hat{o}_{ikm} = \mathbf{w}_{km}\widetilde{\boldsymbol{\vartheta}}_i$. Solving equation (12) for fruit type × variety *i* over all observed *km*∈*z* (i.e., $T_z = N_z$), given by $\widehat{\boldsymbol{\vartheta}}_i$, allows us to produce the unconditional demand for *i* among *km*∈*z*, given by $\hat{o}_{ikm} = \mathbf{w}_{km}\widehat{\boldsymbol{\vartheta}}_i$. We calculate LASSO conditional and unconditional demand elasticities for each fruit type × variety *i* and income class combination by comparing means of $\hat{o}_{ikm}$ calculated with observed prices (household incomes) to means of $\hat{o}_{ikm}$ calculated with a 10% increase in price of fruit type × variety *j* (household incomes), all else equal (we adjust the propensity scores $\hat{\rho}_{ikm}$ appropriately in all of these elasticity calculations).

4.5. Comparing the three estimation approaches to glean information on consumer behavior over organic fruits

As behavioral economists have shown again and again, human behavior is much more inconsistent and capricious than neoclassical theory allows. Therefore, are agnostic ML approaches like LASSO more appropriate than econometric approaches, especially those structured by consumer theory, structure when it comes to inferring inconsistent human behavior? Further, using econometrics to estimate consumer behavior over relatively rare occurrences, such as buying organic fruit, has been shown to be problematic (Meyerhoefer et al. 2005).If LASSO, which by construction predicts fruit purchases better than the econometric approaches, produces elasticity estimates that are markedly different from the econometrically derived estimates, then what does that say about econometric methods for estimating organic fruit demand? Data scientists may look at this as confirmation of the superiority of data mining methods over econometric methods when it comes to modeling human behavior.

A comparison of our two sets of econometric results also will allow us to say something about the validity of consumer theory, at least when it comes to organic fruit buying behavior. Recall the second econometric estimation method imposes symmetry in the price coefficients and homogeneity in prices and income while the first econometric estimation method does not. If the two econometric methods produce elasticity measures that are markedly different then it would appear that consumers, at least when it comes to organic fruit, do not act in a way that is consistent with, for example, the long-held assumption regarding the symmetry of the partial derivatives of the compensated demand functions.

Alternatively, what if LASSO yields less precise estimates or determines that prices do not affect purchasing behavior (i.e., the coefficients on the price variables are set equal to 0)? Econometricians may then conclude that economic theory provides a structure that is better at producing empirical estimates of interest and is the preferred method for policy analysis.

5. Results

We estimate representative household monthly conditional and unconditional purchase probability, own-price, cross price, and income elasticities of demand for organic apples, blueberries, oranges, and strawberries during the 2011 to 2013 period with three methods: 1) individual Heckman models of consumption; 2) an incomplete demand system of consumption; and 3) the LASSO method. We focus on organic apples, blueberries, oranges, and strawberries because, based on expenditures, they are 4 of the 6 most popular organic fruit varieties in the US (Table 5).

We also predict the impact that two policies would have had on representative household monthly purchases of organic fruit. First we evaluate the expected purchasing and expenditure impacts of a joint 10% subsidy of all organic fruit for each representative household. Second, we evaluate the expected purchasing and expenditure impacts of a joint 10% tax on all conventional fruits for each representative household. The policy simulations generated with the individual Heckman models and the incomplete demand system of consumption assume the Boston market intercept. Because the elasticities and policy outcomes are relative, values that are derived with a specific intercept value will not affect our results.

The Stata code used to estimate the individual Heckman models of consumption and incomplete demand system of consumption and calculate the related elasticities and policy responses are available from the authors. The R code used to estimate the LASSO models, to iterate coefficient standard errors, and to calculate the related elasticities and policy responses are also available from the authors. Finally, all data used in these models are also available from the authors. See Appendix Section 7 for details.

5.1. Expected elasticities

5.1.1. Purchase probability elasticities with respect to own-prices (PPEP) and income (PPEI)

Purchase probability elasticities with respect to own-prices (*PPEP*) and income (*PPEI*) measure the relative change in the propensity of a household to buy a type of organic fruit in any given month for an own-price change or household income change, respectively. Overall we find that the decision to purchase organic fruit at the low-income household during the 2011 to 2013 period was not consistently caused / predicted by changes in price or income (Table 6). However, for the representative middle class and rich households, the decision to purchase organic fruit was mostly affected by changes in prices and income in ways consistent with theory. Finally, the middle class household had the most elastic *PPEPs* (either least positive or most negative) during this period.

Estimation method 1 and 2 produce the same *PPEPs* and *PPEIs* because they are estimated with the same exact model and data. *PPEPs* and *PPEIs* generated by estimation methods 1 / 2 have the same sign as *PPEPs* and *PPEIs* generated by estimation method 3 despite their methodological differences. However, in almost every case the *PPEPs* and *PPEIs* generated by estimation method 3 are *smaller* (less positive or more negative) than those found by estimation methods 1 / 2. Given that LASSO is a more accurate predictor than the econometric method, this suggests a persistent bias in the probit estimates of the decision to buy organic fruit or not. We also find that estimates of *PPEPs* and *PPEIs* are similarly precise across the econometric and ML estimations. When estimation method 1 / 2 produces a statistically significant *PPEP* or *PPEI* for a fruit type × variety *i* and income class combination, method 3 also tends to produce a statistically significant *PPEP* or *PPEI* for the same *i*,*z* combination.

Finally, LASSO estimates of organic fruit *PPEPs* and *PPEIs* are generally amenable to policy analysis. In most 100 iterative runs of the LASSO over each fruit type × variety *i* and income class combination the coefficients on fruit prices and household income in the organic fruit selection equation (eq. 11) were assigned nonzero values (Table 7). In other words, the LASSO technique finds that prices and income had affected binary organic fruit purchasing decisions by households, at least from 2011-2013.

5.1.2. Conditional and unconditional elasticities of demand

In general, the own-price conditional and unconditional elasticities of demand for organic fruit had the expected negative signs across all household income bracket types and estimation methods (Tables 8 and 9). For some organic fruit and household type combinations own-price *CPED* was larger than own-price *UPED*, and for other combinations the dominance relationship was reversed. This finding neither supports or contradicts Zhang et al. (2008)'s conclusion that organic produce is often treated as more of a necessity by the occasional to frequent buyer of organic food and often treated as a luxury good by the representative US household.[26]

Of the four fruits, demand for organic oranges tended to be the most inelastic, both conditionally and unconditionally, although the elasticities were typically not statistically different from zero. Unconditionally, strawberries and blueberries were the most own-price elastic. Conditionally, apples are most the own-price elastic where seven of the nine expected own-price CPEDs are less than –1.00. The middle class household tended to have the most elastic own-price *CPEDs* and *UPEDs*. Surprisingly, it is not clear that the rich household was the least own-price sensitive of the three household types. For example, the representative poor household had the lowest blueberry own-price *CPEDs* and *UPEDs*.

Several patterns emerge from the multitude of estimated cross-price *CPEDs* and *UPEDs*. First, if households act according to method 2's set of assumptions then we have strong evidence to suggest that 1) organic blueberries and strawberries and are complements and 2) organic apples and all other organic fruits are substitutes for each other when we consider US households at large (unconditional elasticities). Recall that estimation method 2 includes a budget constraint and enforces consistency in consumer behavior for substitution effects. Second, all estimation methods agree that the representative middle class and rich households treated organic apples and organic oranges as substitutes.

Our estimates of cross-price elasticities with respect to the prices of their *conventional* analogs highlight several other consumer habits. First, as expected, in most cases, organic and conventional fruits of the same type were treated as substitutes. However, expected same-fruit type *UPEDs* are systematically larger and more precise than expected same-fruit type *CPEDs*. This suggests that the buying habits of occasional to habitual organic fruit buying households were not as affected by a change in the price of the conventional analogs as the US public in general.

Finally, our estimates suggest that organic fruit purchases, particularly among occasional to habitual buyers of organic fruits (as measured by *CIEDs*), were driven by heterogeneity in household preferences and less by changes in their budgets (most income elasticities tend towards statistical zero). That we find price changes affect organic fruit purchases more than changes in household income is consistent with the Aschemann-Witzel and Zielke (2017)'s meta-study finding that "consumers report price as the primary barrier to the purchase of organic food" and that income has a more mixed impact on organic fruit purchases (p. 217).[27]

A comparison of estimated elasticities across methods 1 and 2 reveals some interesting patterns.[28] Method 2 generally produces more elastic own-price conditional elasticities than method 1 while method 1 generally produces more elastic own-price unconditional elasticities than method 2. Therefore, the imposition of consumer theory fundamentals (method 2) may overestimate how sensitive the occasional to frequent organic fruit buying household is to changes in organic fruit prices. Conversely, the imposition of consumer theory fundamentals may underestimate how sensitive the general US household is to changes in organic fruit prices.

Comparing the econometric estimates to LASSO estimates we find the econometric methods produce more elastic own-price *CPEDs* and *UPEDs* (the most negative or least positive elasticities) than the machine learning method. In fact LASSO is much more likely to find *CPEDs*, *UPEDs*, *CIEDs*, and *UIEDs* that are equal to 0 than the econometric methods. Interestingly, in the fruit type $\times$ variety i and income class combination cases where LASSO finds a 0 elasticity measure, the standard errors of the elasticity estimates across all 3 estimation methods are not that different. In other words, in these cases the LASSO estimates are just as precise but are biased towards 0.[29]

LASSO estimates of organic fruit *CPEDs*, *UPEDs*, *CIEDs*, and *UIEDs* are *not* as amenable to policy analysis as its estimates of organic fruit *PPEPs* and *PPEIs* are. To see this, compare the relative frequency values in the selection columns, especially for middle class (M) and rich (R) households, to the relative frequency values in the quantity columns of Table 7. As we mentioned above, this table shows that over most LASSO iterations, fruit prices and household income almost always explain the decision to consume organic fruits. In contrast fewer LASSO iterations assign a nonzero coefficient to price and income variables in in the quantity or organic fruit purchased columns. This finding is in line with Aschemann-Witzel and Zielke (2017)'a meta-study conclusion that "psychographics determining favorable beliefs about and attitudes toward organics and, thus, respective motives and preferences seem to be far better explanatory variables” for organic purchase behavior than prices and income (p. 227, Aschemann-Witzel and Zielke 2017).

5.3. Policy analysis

In Tables 10 and 11 we present the results of policy simulations. In the first set of simulations we jointly reduce the observed prices of organic fruits by 10%.[30] In a second set of simulations we impose a 10% tax on all conventional fruits. To measure the relative impacts of these policies we calculate the expected changes in monthly unconditional and conditional household purchases (in ounces) and expenditure (in dollars) for each fruit type × variety *i* and income class combination given the policy-induced price changes. We change these measures to elasticities by dividing by +10% (subsidy) or -10% (tax).

As expected, both policies would have generally increased the expected monthly consumption of organic fruits across all household types. However, in some cases individual organics were consumed less dues to substitution effects across organics. Further, in some cases the amount of money spent on organics fell despite the subsidy due to some inelastic demands.

5.3.1 Conditional demand response to an organic fruit subsidy or conventional fruit tax

According to measured purchase elasticities, the joint reduction in the prices of organics would have increased the conditional monthly consumption of organic apples the most. In contrast, the conditional consumption of organic oranges would have fallen the most among the highlighted fruits in the representative middle and rich income households.

A *positive* expenditure subsidy means a representative household would have been expected to spend less per month on an organic fruit before the subsidy. Conditional expenditure elasticities for organic blueberries, oranges, and strawberries are positive across all household types. Therefore, organic apples are the only of the four featured fruits to have elastic conditional demand: a subsidy would have been expected to entice households to spend more on organic apples, and only organic apples, than they did before the subsidy policy.

The joint increase in the prices of conventionals broadly increases the conditional consumption of all organics (measured by purchase or expenditure elasticity).[31] In this case a positive elasticity means greater monthly consumption and expenditure. Across the three estimation methods, there is broad agreement that the representative poor household would have been expected to increase its conditional consumption of organic blueberries and oranges the most and that the representative rich household would have been expected to increase their conditional consumption of organic blueberries and strawberries the most given the tax on conventionals.

For the middle class and rich households the conditional purchase elasticities with the organic subsidy tend to be greater in an absolute sense than the conditional purchase elasticities with the conventional tax. In other words, for these household types, the subsidy would have had a greater impact on organic fruit consumption behavior than the tax. For the low-income household the comparative impact would have been more mixed; for half of the fruit-estimation method purchase elasticity calculations the tax would have been more consequential than the subsidy when consequence is measured by the absolute value of the elasticity.

Encouraging the less well-to-do to eat more fruit, especially organic fruit, which many believe is healthier that conventional analogs, is one argument for the subsidization of organic fruit. Our results indicate that the subsidy would have increased relative conditional demand for organic fruits among middle class households just as much as it would have among low-income households. Therefore, this subsidization policy would not have been particularly effective at targeting the poor; it would have represented a boon to the broad middle as well (we find that the representative rich household is least affected by the organic fruit subsidy as they tend to have the smallest absolute purchase and expenditure elasticity values). For those that argue for a subsidy / tax to encourage more sustainable farming it is not clear how much these policies would have affected agriculture practices during the 2001 to 2013 period.

5.3.2. Unconditional demand response to an organic fruit subsidy or conventional fruit tax

In most cases, the absolute magnitude of the unconditional purchase and expenditure elasticities given the organic fruit subsidy or conventional tax are greater than their conditional counterparts. In other words, these two policies would have had a greater impact on relative demand for these select organic fruits among the general US public than it would have on the relatively small percentage of US households that already tended to buy organics. Another way to put this conclusion: the subsidy and tax policies would have done relatively more to entice households to make organics part of their grocery basket than they would have increased the volume of organics in the basket. This finding seems to be in line with Zhang et al. (2008)'s conclusion that organic produce is often treated as more of a necessity by the occasional to frequent buyer of organic food and often treated as a luxury good by the representative US household. In particular, unconditional monthly purchases of organic blueberries and strawberries would have increased dramatically with the subsidy on organics or tax on conventionals. In addition, the middle class household would have reacted more strongly to these policies than the low-income and rich households when strength is measured by the absolute value of the unconditional purchase and expenditure elasticities. Finally, if we compare the sizes of the purchase elasticities under the subsidy policy versus the sizes of the purchase elasticities under the tax policy, we find that the tax generally would have driven households to buy more organics than the subsidy would have. Therefore, if the goal is to increase the consumption of organics as much as possible across the US public then the tax would have been the preferred policy.

Finally, our results again indicate that the subsidy would have increased relative unconditional demand for organic fruits among middle class households as much as it would have among low-income households. Therefore, we can again say that this subsidization policy would not have been particularly effective at targeting the poor; it would have represented a boon to the broad middle as well.

5.3.3. Differences in policy simulation results across estimation methods

Estimation method 1 almost always generates the highest unconditional purchase and expenditure elasticities (they are all positive) under the conventional tax policy. We were not surprised to find that estimation method 3 was most likely to produce policy elasticities closer to 0 than the other two methods. According to Table 7 latent demand is not consistently explained by fruit prices (the quantity columns). Therefore, a change in price is more likely to be a more muted effect on household behavior under estimation method 3 than the other two demand estimation approaches.

The differences in estimation methods 1 and 2's results are largely explained by the added economic theory structure in estimation method 2. The unconditional responses to the subsidy and tax for a given fruit – household type are fairly similar across estimation methods 1 and 2. The same cannot be said for the conditional responses to the subsidy and tax for a given fruit – household type; not only do conditional elasticity signs for a given fruit – household type often differ between the two methods but the magnitudes can be quite different. All of this suggests that the organic fruit purchasing behavior of occasional to frequent organic fruit buyers does not align with consumer theory fundamentals as much as the organic fruit purchasing behavior of the larger US population.

6. Conclusions and Discussion

In this paper we have used three techniques to estimate the monthly household demand for some of the most popular organic fruits during the 2011 to 2013 period. We estimated demands for three household types, low-income, middle class, and rich. Ours is the first paper to produce detailed demand elasticities for organic fruits in the US across the household income spectrum, and, in addition, to compute demand elasticities for organic fruits with both traditional econometric and ML techniques.

Generally, we find that own-price elasticities of demand for organic fruits are negative and therefore consistent with economic theory. This is the case despite inexplicable positive purchase probability elasticities with respect to prices for organic apples and, in some cases, blueberries. Whether or not a particular fruit's own-price elasticities of demand are elastic or inelastic can vary across estimation methods and household types. However, the data suggest that conditional own-price elasticities of demand for organic apples and unconditional own-price elasticities of demand for organic strawberries tend to be elastic. Of the four organic fruits we study, elasticities of demand for organic oranges often are statistically equivalent to zero.

Income elasticities of demand measurements are inconsistent and often statistically insignificant. This is the case despite expected positive purchase probability elasticities with respect to income for organic apples, blueberries, and strawberries for the representative middle class and rich households. Inconsistent and statistically insignificant income elasticities of demand suggest to us that organic fruit consumption is driven more by lifestyle choices than it is changes in income. Our finding is consistent with the survey literature that has found that for many consumers the decision to buy organic food is a moral or ethical choice (Thøgersen 2011, Juhl et al. 2017).

Our two policy experiments, a 10% subsidy of all organic fruits and then a 10% tax on all conventional fruits, provide nice summaries of the cumulative effect of the many own- and cross-price elasticities of demand. When all organic fruit prices drop by 10% frequent organic consumers particularly gravitate to organic apples and when all conventional fruit prices increase by 10% frequent organic consumers tend to spread their purchases out among the four studied organic fruits (according to conditional purchase and expenditure elasticities). Conditional elasticity magnitudes do not differ that much across income strata, lending further credence to the observation that habitual buyers are less concerned about expenditure minimization no matter their income, and are more interested in sustaining a lifestyle.

Our hypothetical policies engender a stronger reaction among the general public than habitual buyers; unconditional purchase and expenditure elasticities are generally larger (in an absolute sense) than conditional purchase and expenditure elasticities. When all organic fruit prices drop by 10% the general public particularly gravitates to organic strawberries and when all conventional fruit prices increase by 10% the general public particularly gravitates to organic blueberries and strawberries (according to unconditional purchase and expenditure elasticities). Unconditional elasticity magnitudes are largest for the middle-income household.

The policy results generally confirm Zhang et al.'s (2011) findings that a decrease in organic price premiums would lead to a strong increase in the purchase of organic produce (Rödiger and Hamm 2015). We extend this finding by noting that, among habitual buyers of organic fruits, reductions in the organic price premiums generate a modest increase in organic fruit purchases. The response to the reduction in price premiums among the general public is much stronger. However, the tax policy does more to encourage organic fruit consumption among the US general public than the organic fruit subsidy.

Finally, the elasticities measured with the LASSO technique are not radically different than those measured with traditional econometric methods. For example, in most cases the signs on the elasticities calculated with the econometric methods and LASSO are the same. The most noticeable difference between the two analytical techniques is that the LASSO technique is more likely to find price and income elasticities of demand that are equal to zero, either statistically or in magnitude. This is likely in part because LASSO uses a coefficient shrinkage penalty when cross-validating predictive fit, so the technique may tend to yield smaller elasticities as a tradeoff against avoiding overfitting the data. An exploration of and comparison against other ML models is therefore an important area of future research, in this context. However, we do find that the LASSO results are not as amenable to policy simulations as the econometrically derived results, as the solved LASSO models do not consistently find prices and household income to be predictors of organic fruit latent demand.

7. References

Aschemann-Witzel, J., and S. Zielke. 2017. "Can't Buy Me Green? A Review of Consumer Perceptions of and Behavior Toward the Price of Organic Food." *Journal of Consumer Affairs* 51(1): 211-251.

Athey, S., and G. Imbens. 2015. Lectures on Machine Learning. NBER, Saturday, July 18th, 2015. Accessed at: http://www.nber.org/econometrics_minicourse_2015/syllabus.pdf

Athey, S., and G. Imbens. 2016. "The State of Applied Econometrics-Causality and Policy Evaluation." arXiv preprint arXiv:1607.00699.

Bajari, P., D. Nekipelov, S.P. Ryan, and M. Yang, 2015. "Machine learning methods for demand estimation." *The American Economic Review* 105(5):481-485.

Batte, M., N. Hooker, T. Haab, and J. Beaverson. 2007. "Putting their money where their mouths are: Consumer willingness to pay for multi- ingredient, processed organic food products." *Food Policy* 32(2):145-159.

Beghin, J. C., J. C. Bureau, and S. Drogue. 2004. "Calibration of Incomplete Demand Systems in Quantitative Analysis." *Applied Economics* 36:839–847.

Bezawada, R., and K. Pauwels. 2013. "What Is Special About Marketing Organic Products? How Organic Assortment, Price, and Promotions Drive Retailer Performance." *Journal of Marketing* 77(1):31-51.

Buchner, F., J. Wasem, and S. Schillo. 2017. "Regression Trees Identify Relevant Interactions: Can This Improve the Predictive Performance of Risk Adjustment?" Health Econ., 26:74-85. doi: 10.1002/hec.3277.

Bunte, F. H., M. A. van Galen, W. E. Kuiper, and G. Tacken. 2010. "Limits to growth in organic sales." *De Economist* 158(4):387-410.

Byrne, P. J., O. Capps, Jr., and A. Saha. 1996, "Analysis of Food Away from Home Expenditure Patterns for U.S. Households, 1982-89." *American Journal of Agricultural Economics* 78(3):614-627.

Canavari, M., R. Ghelfi, K. Olson, and S. Rivaroli. 2007. "A comparative profitability analysis of organic and conventional farms in Emilia-Romagna and in Minnesota." In M. Canavari and

K.D. Olson, eds., *Organic Food: Consumers' Choices and Farmers' Opportunities*. New York: Springer pp. 31-34.

Carroll, C., D. Charlton, and E. Tjernström. 2012. "Price Premiums for Organic Strawberries." *ARE Update* 15(6):12-14. University of California Giannini Foundation of Agricultural Economics.

Chalfin, A., O. Danieli, A. Hillis, Z. Jelveh, M. Luca, J. Ludwig, and S. Mullainathan. 2016. "Productivity and Selection of Human Capital with Machine Learning." *The American Economic Review* 106(5):124-127.

Cotti, C., Erik Nesson, and Nathan Tefft. 2016. "The Effects of Tobacco Control Policies on Tobacco Products, Tar, and Nicotine Purchases: Evidence from Household Panel Data." *American Economic Journal: Economic Policy* 8(4):103-23. doi: 10.1257/pol.20150268

Crinnion, W. 2010. "Organic foods contain higher levels of certain nutrients, lower levels of pesticides, and may provide health benefits for the consumer." *Alternative Medicine Review* 15(1):4-12.

Crowder, D.W., and J.P. Reganold. 2015. "Financial competitiveness of organic agriculture on a global scale." *Proceedings of the National Academy of Sciences* 112(24):7611-7616.

Dangour, A., K. Lock, A. Hayter, A. Aikenhead, E. Allen, and R. Uauy. 2010. "Nutrition-related health effects of organic foods: a systematic review." *The American Journal of Clinical Nutrition* 92(1):203-210.

Dimitri, C., and C. Greene. 2002. "Recent growth patterns in the US organic foods market." *Agriculture Information Bulletin* 777:1-21.

Dimitri, C., and L. Lohr. 2007. "The US Consumer Perspective on Organic Foods." In M. Canavari and K.D. Olson, eds., *Organic Food: Consumers' Choices and Farmers' Opportunities*. New York: Springer pp. 157-167

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani. 2004. "Least angle regression." *The Annals of Statistics* 32(2):407-499.

Einav, L., and J. Levin. 2014. "Economics in the age of big data." *Science* 346(6210):1243089. doi: 10.1126/science.1243089

El-Hage Scialabba, N. 2013. “Organic agriculture’s contribution to sustainability.” Online. Crop Management. doi:10.1094/CM-2013-0429-09-PS.

Fabiosa, J.F., and H.H. Jenson. 2003. “Usefulness of Incomplete Demand Model in Censored Demand System Estimation.” Paper presented at American Agricultural Economics Association Annual Meeting, Montreal, Canada, 27-30 July.

Fourmouzi, V., M. Genius, and P. Midmore. 2012. "The demand for organic and conventional produce in London, UK: A system approach." *Journal of Agricultural Economics* 63.3:677-693.

Friedman, J., T. Hastie, and R. Tibshirani. 2010. “Regularization Paths for Generalized Linear Models via Coordinate Descent.” *Journal of Statistical Software*, 33(1). doi: 10.18637/jss.v033.i01.

Gattinger, A., A. Muller, M. Haeni, C. Skinner, A. Fliessbach, N. Buchmann, P. Mäder, et al. 2012. "Enhanced top soil carbon stocks under organic farming." *Proceedings of the National Academy of Sciences* 109(44):18226-18231.

Hallam, D. 2003. "The organic market in OECD countries: past growth, current status and future potential." In *Organic Agriculture: Sustainability, Markets and Policies*. Paris: OECD Publishing, pp. 179-186.

Heien, D., and C.R. Wessells. 1990. “Demand System Estimation with Microdata: A Censored Regression Approach.” *Journal of Business and Economic Statistics* 8 (July) 365-71.

Hjelmar, U. 2011. "Consumers’ purchase of organic food products. A matter of convenience and reflexive practices." *Appetite* 56(2):336-344.

Holzman, D. C. 2012. “Organic food conclusions don’t tell the whole story.” *Environmental Health Perspectives* 120(12):A458. Accessed at: https://ehp.niehs.nih.gov/120-a458/

Huang, C. 1996. "Consumer preferences and attitudes towards organically grown produce." *European Review of Agricultural Economics* 23(3):331-342.

Hughner, R. S., P. McDonagh, A. Prothero, C. Shultz, and J. Stanton. 2007. "Who are organic food consumers? A compilation and review of why people purchase organic food." *Journal of Consumer Behaviour* 6(2-3):94-110.

Janssen, M., and U. Hamm. 2012. "Product labelling in the market for organic food: Consumer preferences and willingness-to-pay for different organic certification logos." *Food Quality and Preference* 25(1):9-22.

Juhl, H.J., H.J. Morten, and J.T. Fenger. 2017. "Will the Consistent Organic Food Consumer Step Forward? An Empirical Analysis." *Journal of Consumer Research*, ucx052. doi: 10.1093/jcr/ucx052.

Johansson E., A. Hussain, R. Kuktaite, S.C. Andersson, and M.E. Olsson. 2014. "Contribution of organically grown crops to human health." *International Journal of Environmental Research and Public Health* 11(4):3870-3893.

The JPMorgan Chase Institute 2016. "Consumption Inequality: What's in Your Shopping Basket?" 8 June. Accessed at https://www.jpmorganchase.com/corporate/institute/insight-consumption-inequality.htm

Kasteridis, P.,and S.T. Yen. 2012. "US demand for organic and conventional vegetables: a Bayesian censored system approach." *Australian Journal of Agricultural and Resource Economics* 56(3):405-425.

Kilts Nielsen Center. 2014. Kilts Nielsen Consumer Panel Dataset Manual and FAQs. The Kilts Center Archive of The Nielsen Company, 2 April.

LaFrance, J. T. 1990. "Incomplete Demand Systems and Semilogarithmic Demand Models." *Australian Journal of Agricultural Economics* 34(2):118-131.

LaFrance, J. T., and W.M. Hanemann. 1989. "The dual structure of incomplete demand systems." *American Journal of Agricultural Economics*. 71:262–274.

Lairon, D. 2010. "Nutritional quality and safety of organic food. A review." *Agronomy for sustainable development* 30(1):33-41.

Lin, B-H., S. Yen, C. Huang, and T. Smith. 2009. "U.S. Demand for Organic and Conventional Fresh Fruits: The Roles of Income and Price." *Sustainability* 1:464-478. doi:10.3390/su1030464

Lockeretz, W. 2003. "What are the key issues for conusmers?" In *Organic Agriculture: Sustainability, Markets and Policies*. Paris: OECD Publishing, pp. 239-243.

Loureiro, M., and S. Hine. 2002. "Discovering niche markets: A comparison of consumer willingness to pay for local (Colorado grown), organic, and GMO-free products." *Journal of Agricultural and Applied Economics* 34(3):477-488.

Meyerhoefer, C.D., C.K. Ranney, and D.E. Sahn. 2005. "Consistent estimation of censored demand systems using panel data." *American Journal of Agricultural Economics* 87.3:660-672.

Monier, S., D. Hassan, V. Nichèle, and M. Simioni. 2009. "Organic food consumption patterns." *Journal of Agricultural & Food Industrial Organization* 7(2):1-23.

Offermann, F., and H. Nieberg. 2000. "Economic Performance of Organic Farms in Europe." *Organic Farming in Europe: Economics and Policy* 5. Stuttgart-Hohenheim. https://www.uni-hohenheim.de/i410a/ofeurope/organicfarmingineurope-vol5.pdf

Organisation for Economic Co-Operation and Development (OECD). 2003. Organic Agriculture: Sustainabiblity, Markets and Policies. 406 pages.

Pearson, D., J. Henryks, and H. Jones. 2011. "Organic food: What we know (and do not know) about consumers." *Renewable Agriculture and Food Systems* 26(2):171-177.

Reganold, J.P., and J.M. Wachter. 2016. "Organic agriculture in the twenty-first century." *Nature Plants* 2:15221.

Rödiger, M., and U. Hamm. 2015. "How are organic food prices affecting consumer behaviour? A review." *Food Quality and Preference*, 43:10-20. doi:10.1016/j.foodqual.2015.02.002.

Rousseau, S., and L. Vranken. 2013. "Green market expansion by reducing information asymmetries: Evidence for labeled organic food products." *Food Policy* 40:31-43.

Saha, A., O. Capps, and P. Byrne. 1997a. "Calculating marginal effects in models for zero expenditures in household budgets using a Heckman-type correction." *Applied Economics* 29(10):1311-1316

—. 1997b. "Calculating marginal effects in calculating marginal effects in dichotomous-continuous models", *Applied Economic Letters* 4(3):181-185.

Seidu, A., and J. Seale. 2015. "Estimating Danish Consumers' Preference for Organic Foods: Application of a Generalized Differential Demand System." Paper presented at Southern Agricultural Economics Association annual meeting, Atlanta, GA, 31 January—3 February.

Shonkwiler, J.S., and Yen, S. T. 1999. “Two-step Estimation of a Censored System of Equations.” *American Journal of Agricultural Economics*, 81:972-82.

Smith, T., C. Huang, and B-H. Lin. 2009. "Does price or income affect organic choice? Analysis of US fresh produce users." *Journal of Agricultural and Applied Economics* 41(3):731-744.

Smed, S. 2012. "Information and consumer perception of the ‘organic’ attribute in fresh fruits and vegetables." *Agricultural Economics* 43(1):33-48.

Smith-Spangler C., et al. 2012. “Are organic foods safer or healthier than conventional alternatives? A systematic review.” *Annals of Internal Medicine* 157(5):348–366. http://www.ncbi.nlm.nih.gov/pubmed/22944875.

Stevens-Garmon, J., C. Huang, and B-H. Lin. 2007. "Organic demand: a profile of consumers in the fresh produce market." *Choices* 22(2):109-116.

Thompson, G., and J. Kidwell. 1998. "Explaining the choice of organic produce: cosmetic defects, prices, and consumer preferences." *American Journal of Agricultural Economics* 80(2):277-287.

Tuck, S. L., C. Winqvist, F. Mota, J. Ahnström, L.A. Turnbull, and J. Bengtsson. 2014. “Land-use intensity and the effects of organic farming on biodiversity: a hierarchical meta-analysis.” *Journal of Applied Ecology* 51(3):746-755.

U.S. Bureau of the Census. 2017. *Real Median Household Income in the United States* [MEHOINUSA672N]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MEHOINUSA672N, June 18, 2017.

U.S. Department of Agriculture. 2015. NASS. Quick Stats.

U.S. Department of Agriculture 2016. NASS. *Organic Survey* (2014). Volume 3. Special Studies. Part 4. AC-12-SS-4. 2012 Census of Agriculture. Updated April 2016.

U.S. Department of Agriculture 2016A. ERS*. Organic Market Overview*. Accessed at https://www.ers.usda.gov/topics/natural-resources-environment/organic-agriculture/organic-market-overview.aspx, December 1, 2016.

U.S. Department of Agriculture – 2016B. ERS. *Organic Prices*. Accessed at https://www.ers.usda.gov/data-products/organic-prices/organic-prices/, December 1, 2016.

Wier, M., K. Jensen, L. Andersen, and K. Millock. 2008. "The character of demand in mature organic food markets: Great Britain and Denmark compared." *Food Policy* 33(5):406-421.

Willer, H., and D. Schaack. 2015. “Organic Farming and Market Development in Europe.” In H. Willer, and J. Lernoud, eds. *The World of Organic Agriculture. Statistics and Emerging Trends*. Frick, Switzerland: Research Institute of Organic Agriculture (FiBL) and IFOAM–Organics International, pp. 181–214.

Winter, C., and S. Davis. 2006. "Organic foods." *Journal of Food Science* 71(9):R117-R124.

Varian, H.R. 2014. “Big data: New tricks for econometrics.” *The Journal of Economic Perspectives* 28(2):3-27.

Yen, S. T., and B.H. Lin. 2002. “Beverage consumption among US Children and Adolescents: full-information and quasi maximum likelihood estimation of a censored system.” *European Review of Agricultural Economics* 29(1):85-103.

Yen, S. T., B.H. Lin, and D.M. Smallwood. 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* 85:458–478.

Yiridoe, E., S. Bonti-Ankomah, and R. Martin. 2005. "Comparison of consumer perceptions and preference toward organic versus conventionally produced foods: a review and update of the literature." *Renewable Agriculture and Food Systems* 20(4):193-205.

Zepeda, L., and C. Nie. 2012. "What are the odds of being an organic or local food shopper? Multivariate analysis of US food shopper lifestyle segments." *Agriculture and Human Values* 29(4):467-480.

Zhang, F., C. Huang, B-H. Lin, and J. Epperson. 2008. "Modeling fresh organic produce consumption with scanner data: A generalized double hurdle model approach." *Agribusiness* 24(4):510-522.

Zhang, F., C.L. Huang, B-H. Lin, J.E. Epperson, and J.E. Houston. 2011. “National Demand for Fresh Organic and Conventional Vegetables: Scanner Data Evidence.” *Journal of Food Products Marketing* 17(4):441-458.

Figures

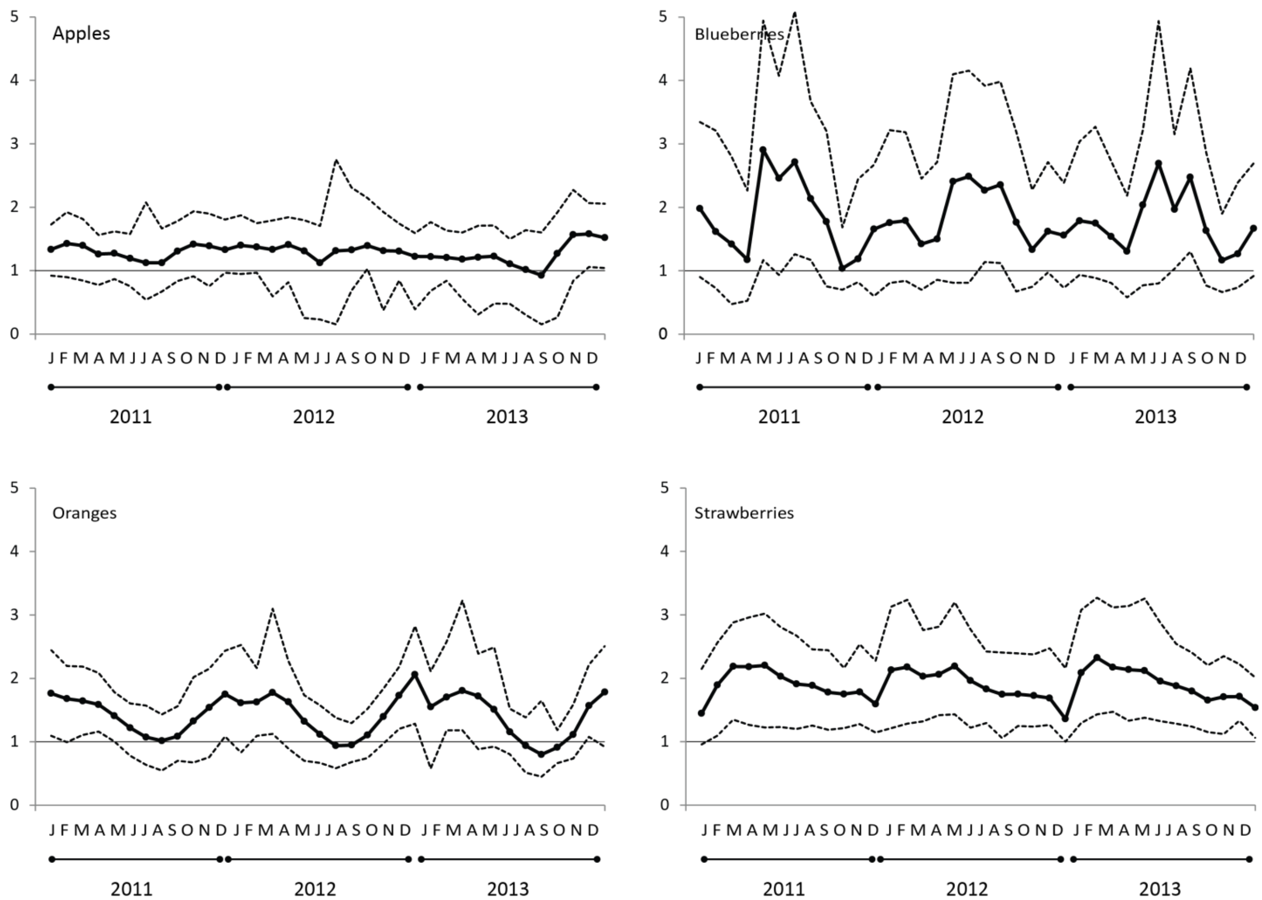

Fig. 1. 5th percentile, mean, and 95th percentile organic to conventional price ratios during the 2011-2013 period for apples, blueberries, oranges, and strawberries. The price premiums for popular organic fruits cycle above and below that threshold across the seasons. Hallam (2003) found organic price premiums of 20 to 30% in OECD countries in the early 2000s. USDA-ERS (2016A) also analyzed prices for 18 fruits with 2005 data and found that the organic premium was less than 30% for most items. Blueberries were the anomaly; its price premium exceeded 100%. Prices depicted in this graph are monthly national averages not weighted by sampled household projection factors.

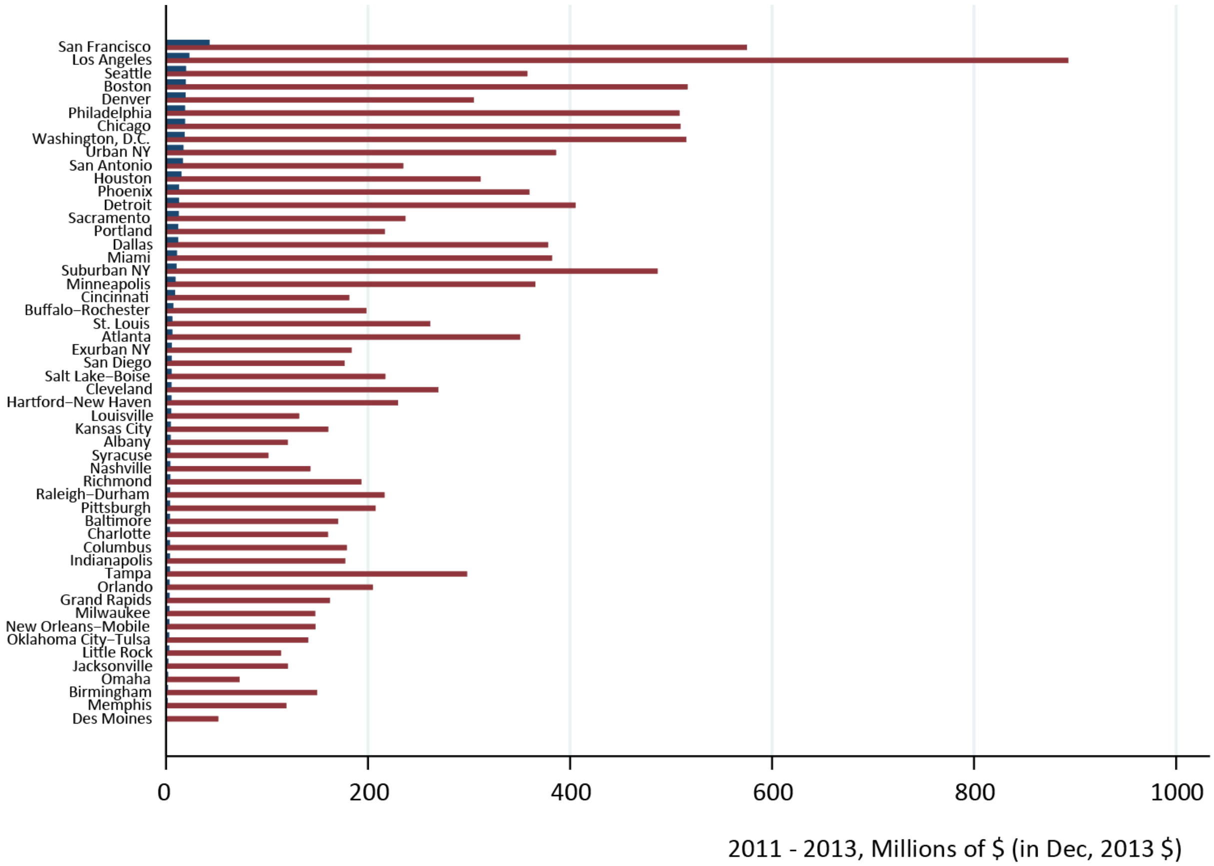

Figure 2. Conventional (red bars) and organic (blue bars) fruit expenditures by Nielsen Scantrack market during the years 2011 through 2013 (December, 2013 dollars). All household expenditures in year *y* are inflated with households' year *y* projection factor to arrive at market totals. A household's year *y* projection factor indicates how many other households in that market the household in question represents in year *y*. The y-axis is ordered by level of gross expenditures on organic fruit.

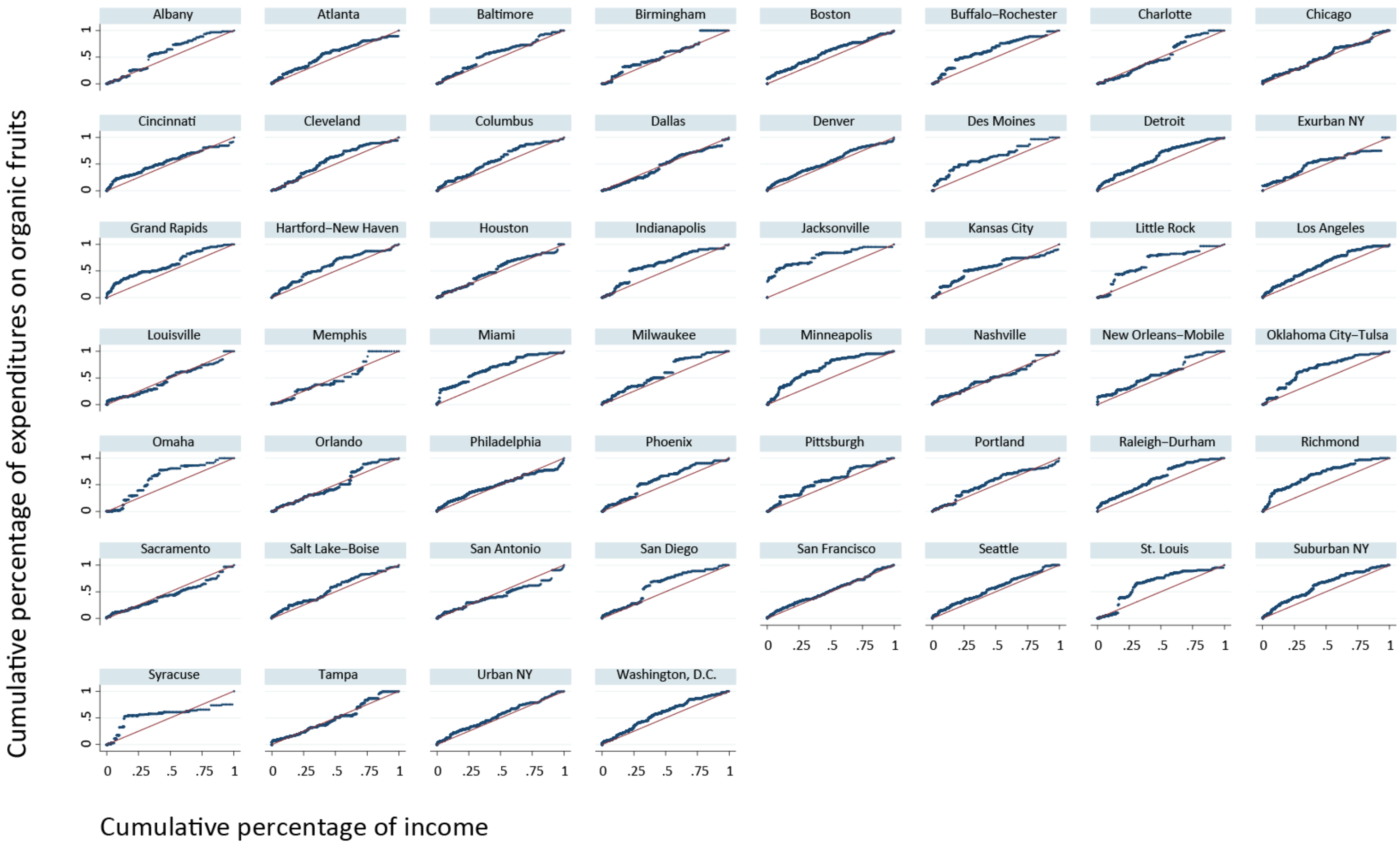

Figure 3: **Lorenz curves of organic fruit expenditures by Nielsen Scantrack markets during the years 2011 through 2013 (December, 2013 dollars).** The dark line in each plot is the actual cumulative expenditure curve and the lighter line is the 45 degree line. In the top five organic fruit markets by gross expenditure, San Francisco, LA, Seattle, Boston, and Denver, spending on organic fruit was proportional across the income spectrum. The markets with the most uneven distribution tend to be the smaller markets for overall fruit consumption. In these markets the top 50^{th} percentile households by income purchased less organic fruit than an even distribution of expenditures across the income spectrum would predict.

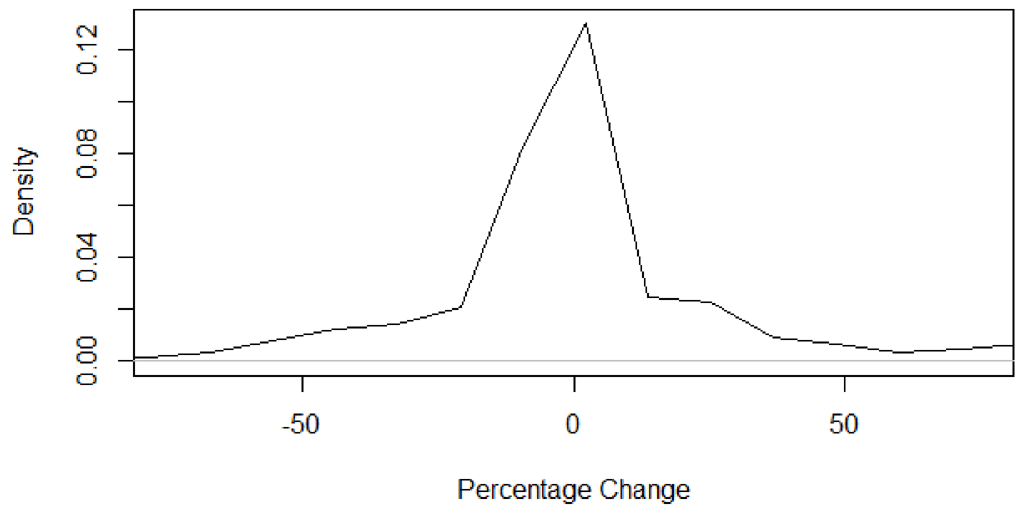

Figure 4: The density of percentage change in household income from 2011 to 2013 across the households in our dataset. Median: -1.86% change. Mean: 12.43% change. Standard deviation: 161.6% change. To be included in this density function the household had to be in the Consumer Panel dataset in 2011, 2012, and 2013.

Tables

Table 1: Real US expenditures on fruit by household income group. Only includes households that bought fruit, organic or conventional, at some point in a given year. Household projection factors are used to extrapolate panel totals to national totals. All dollar values are measured in December, 2013 dollars.

		2011		2012		2013	
	Household Status	**Organic**	**Conventional**	**Organic**	**Conventional**	**Organic**	**Conventional**
Total Expenditures (M $)	**Low income**	9.97	571.76	12.38	636.23	16.82	692.64
	Middle	57.16	2392.18	65.68	2621.65	76.88	2633.01
	Rich	77.70	2504.07	89.93	2375.93	117.83	2519.55
	All	144.82	5468.00	167.99	5633.82	211.53	5845.20
Expenditures / Household	**Low income**	0.65	37.06	0.77	39.67	0.98	40.32
	Middle class	1.16	48.59	1.27	50.70	1.53	52.40
	Rich	2.08	66.99	2.66	70.16	3.38	72.26
	All	1.42	53.59	1.65	55.45	2.07	57.14

Notes: Low income households have a household income that is 130% or less of the poverty line conditional on year and household size. Middle class households have a household income that is between 130% and 500% of the poverty line conditional on year and household size. Rich households have a household income greater than 500% of the poverty line conditional on year and household size. See https://aspe.hhs.gov/2011-hhs-poverty-guidelines; https://aspe.hhs.gov/2012-hhs-poverty-guidelines; and https://aspe.hhs.gov/2013-poverty-guidelines.

Table 2: Number of US households that bought organic and conventional fruit from 2011 to 2013 (Millions of HHs). Projection factors are used to extrapolate panel level results to national estimates.

	2011			2012			2013		
Household Status	**Only Organic**	**Only Conventional**	**Both**	**Only Organic**	**Only Conventional**	**Both**	**Only Organic**	**Only Conventional**	**Both**
Low income	0.02	14.34	1.06	0.00	14.84	1.19	0.04	15.48	1.65
Middle class	0.06	44.39	4.78	0.10	45.84	5.76	0.15	43.35	6.75
Rich	0.06	31.52	5.79	0.06	27.51	6.30	0.08	27.39	7.40
All	0.15	90.25	11.64	0.16	88.20	13.25	0.27	86.22	15.80

Notes: Low income households have a household income that is 130% or less of the poverty line conditional on year and household size. Middle class households have a household income that is between 130% and 500% of the poverty line conditional on year and household size. Rich households have a household income greater than 500% of the poverty line conditional on year and household size. See https://aspe.hhs.gov/2011-hhs-poverty-guidelines; https://aspe.hhs.gov/2012-hhs-poverty-guidelines; and https://aspe.hhs.gov/2013-poverty-guidelines.

Table 3: Domestic Organic Fruit Production and Imports in 2011 and 2014.

	Domestic Acres Harvested			Domestic Production (M of Pounds)			Imports (M of Dollars)		
	2011	2014	% Change	2011	2014	% Change	2011	2014	% Change
Apples	26,721	64,985	143.2%	595.85	1,969.76	230.6%	5.74	29.77	418.8%
Blueberries	3,073	5,307	72.7%	13.92	24.16	73.6%	2.92	6.24	113.7%
Oranges	6,610	7,822	18.3%	123.03	121.08	-1.6%	NA	122.64	NA
Strawberries	1,638	2,961	80.8%	37.79	56.40	49.2%	3.67	11.45	212.1%

Source: USDA Census of Agriculture, USDA-NASS, QuickStats, https://quickstats.nass.usda.gov/; https://www.ers.usda.gov/topics/natural-resources-environment/organic-agriculture/organic-trade/

Table 4: The Impact of Time on Organic Fruit Prices. We measured the impact of time on organic fruit prices by regressing household-month prices for organic fruit *i* in season *s* on all other fruit type × variety *i* prices in season *s*, year dummy variables, and several spatial location dummy variables. The estimated OLS coefficients on the year dummy variables are presented in panel A of the table (2011 is the omitted year). The seasonal shares of organic fruit *i*'s national consumption in 2012 and 2013 are given in panel B. Household projection factors are used to extrapolate Consumer Panel seasonal shares of organic fruit *i*'s consumption to national seasonal shares (panel B). See Appendix Section 8 for more information on obtaining the Stata code and data files used to compile this table.

	Winter		**Spring**		**Summer**		**Fall**	
A	**2012**	**2013**	**2012**	**2013**	**2012**	**2013**	**2012**	**2013**
Apples	0.005***	0.007***	0.005***	0.010***	0.012***	**-0.002*** **	0.007***	0.011***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	**(0.0002)**	(0.0001)	(0.0001)
Blueberries	**-0.034*** **	**-0.013*** **	0.065***	**-0.030*** **	0.0002	**-0.054*** **	0.125***	0.118***
	(0.0008)	**(0.0012)**	(0.001)	**(0.0014)**	(0.0005)	**(0.0006)**	(0.0008)	(0.0008)
Oranges	**-0.003*** **	**-0.004*** **	0.003***	0.003***	**-0.003*** **	**-0.004*** **	0.006***	0.001***
	(0.0001)	**(0.0001)**	(0.0001)	(0.0001)	**(0.0001)**	**(0.0001)**	(0)	(0.0001)
Strawberries	0.013***	0.014***	**-0.013*** **	**-0.008*** **	**-0.015*** **	**-0.004*** **	**-0.007*** **	0.013***
	(0.0004)	(0.0005)	**(0.0002)**	**(0.0002)**	**(0.0001)**	**(0.0002)**	**(0.0002)**	(0.0002)
B								
Apples	0.18	0.25	0.30	0.31	0.13	0.12	0.38	0.31
Blueberries	0.44	0.30	0.19	0.28	0.32	0.36	0.06	0.06
Oranges	0.22	0.23	0.44	0.39	0.23	0.22	0.12	0.17
Strawberries	0.05	0.05	0.31	0.35	0.45	0.41	0.18	0.19

Table 5: Real organic expenditures by variety across US households. Projection factors are used to extrapolate panel level results to national estimates. All dollar values are measured in December, 2013 dollars.

	2011	2012	2013
Apples	28.80	33.20	43.80
Blueberries	19.70	27.10	38.90
Oranges	7.16	7.48	8.42
Strawberries	47.90	53.70	58.00
Other	41.20	46.50	62.40
Blackberries	8.86	6.86	9.32
Grapes	2.48	5.10	8.70
Grapefruit	1.32	1.07	1.23
Lemons	2.51	2.30	3.48
Raspberries	19.90	22.20	21.30
Misc.	6.14	8.96	18.40
Total	144.76	167.98	211.52

Notes: 'Other' is the sum of expenditures on blackberries, grapes, grapefruit, lemons, raspberries, and miscellaneous.

Table 6: Estimated Purchase Probability Elasticities for Four Organic Fruits, 2011 – 2013

	Est. Method	Organic Purchase Probability Elasticity with Respect to Price ($PPEP_{jj}$)				Organic Purchase Probability Elasticity with Respect to Income ($PPEI_i$)			
		Apples	Blueberries	Oranges	Strawberries	Apples	Blueberries	Oranges	Strawberries
Low income HH	1	0.45* (0.27)	0.38* (0.22)	-0.74 (0.53)	-1.46*** (0.42)	0.06 (0.21)	-0.45** (0.20)	0.16 (0.23)	-0.15 (0.15)
	2	0.45* (0.27)	0.38* (0.22)	-0.74 (0.53)	-1.46*** (0.42)	0.06 (0.20)	-0.45** (0.20)	0.16 (0.23)	-0.15 (0.15)
	3	0.00 (0.37)	0.43 (0.27)	0.00 (0.96)	-1.70*** (0.38)	0.00 (0.15)	-0.35** (0.15)	0.00 (0.36)	0.00 (0.1)
Middle HH	1	0.63*** (0.14)	-0.16* (0.08)	-0.58** (0.24)	-1.64*** (0.16)	0.46*** (0.12)	0.50*** (0.11)	-0.03 (0.16)	0.57*** (0.09)
	2	0.63*** (0.14)	-0.16 (0.11)	-0.56** (0.24)	-1.64*** (0.16)	0.46*** (0.12)	0.50*** (0.11)	-0.03 (0.16)	0.57*** (0.09)
	3	0.35** (0.15)	-0.30*** (0.09)	-0.33 (0.34)	-1.85*** (0.16)	0.50*** (0.13)	0.46*** (0.16)	0.00 (0.13)	0.49*** (0.11)
Rich HH	1	0.84*** (0.15)	0.21*** (0.07)	-0.26 (0.24)	-0.75*** (0.15)	0.30** (0.14)	0.69*** (0.10)	0.15 (0.19)	0.59*** (0.09)
	2	0.84*** (0.15)	0.21*** (0.07)	-0.26 (0.24)	-0.75*** (0.15)	0.30** (0.14)	0.69 *** (0.10)	0.15 (0.19)	0.59*** (0.09)
	3	0.59*** (0.15)	0.14* (0.09)	-0.10 (0.41)	-1.00*** (0.14)	0.00 (0.09)	0.31** (0.12)	-0.21 (0.21)	0.39*** (0.09)

Notes: Standard errors of estimates are in parentheses. Est. method 1 refers to individual fruit Heckman models of consumption (see section 4.2). Est. method 2 refers to the incomplete demand system of consumption (see section 4.3). The elasticities from est. methods 1 and 2 assume the Boston market. Est. method 3 refers to the LASSO model (see section 4.4). '***' indicates significant at $p < 0.01$; '**' indicates significant at $p < 0.05$; and '*' indicates significant at $p < 0.10$

Table 7: The frequency with which a variable's estimated coefficient is non-zero across 101 iterations of the LASSO model. This table indicates the fraction of 101 LASSO iterations where a variable's estimated coefficient was non-zero in the selection stage (S) (eq. 11) and the quantity stage (Q) (eq. 12) of the organic fruit consumption model. 'P,' M,' or 'R' indicates the LASSO iterations run over the dataset of low income, middle class, or rich households. Dark green indicates that the variable was selected under most or all iterations. Dark red indicates that the variable was selected under few or no iterations. Yellow is the median color on the 0 to 1 scale.

	Apple						**Blue Berries**						**Oranges**						**Strawberries**					
	Selection			**Quantity**			**Selection**			**Quantity**			**Selection**			**Quantity**			**Selection**			**Quantity**		
Price variables	**P**	**M**	**R**	**P**	**M**	**R**	**P**	**M**	**R**	**P**	**M**	**R**	**P**	**M**	**R**	**P**	**M**	**R**	**P**	**M**	**R**	**P**	**M**	**R**
Organic apple	**0.72**	**1.00**	**1.00**	**0.79**	**1.00**	**1.00**	0.75	0.74	0.92	0.20	0.91	0.96	0.86	0.64	0.65	0.86	0.20	0.18	0.79	0.94	0.82	0.37	0.86	0.96
Conventional apple	**1.00**	**1.00**	**1.00**	**0.23**	**1.00**	**0.97**	0.92	1.00	1.00	0.54	0.63	0.69	0.80	1.00	1.00	0.80	0.19	0.40	0.98	1.00	1.00	0.45	0.67	0.74
Organic blueberries	0.80	0.76	0.73	0.29	0.87	0.48	**0.96**	**1.00**	**0.96**	**0.99**	**1.00**	**1.00**	0.76	0.92	0.72	0.76	0.28	0.17	0.68	0.80	0.88	0.67	0.81	0.55
Conventional blueberries	0.91	1.00	1.00	0.45	0.78	0.51	**0.82**	**0.97**	**0.99**	**0.27**	**0.95**	**0.52**	0.74	0.52	0.55	0.74	0.22	0.50	1.00	1.00	1.00	0.31	0.74	0.93
Organic oranges	0.69	0.74	0.61	0.17	0.67	0.59	0.98	1.00	1.00	0.44	0.71	0.52	**0.73**	**0.97**	**0.81**	**0.73**	**0.43**	**0.39**	0.92	0.97	0.93	0.63	0.55	0.95
Conventional oranges	0.40	0.69	0.63	0.25	0.76	0.83	0.91	1.00	0.99	0.29	0.54	0.92	**0.83**	**0.54**	**0.68**	**0.83**	**0.26**	**0.33**	0.65	0.71	1.00	0.33	0.50	0.70
Organic strawberries	0.89	0.82	0.68	0.67	0.97	0.62	0.93	0.95	0.82	0.69	0.52	0.44	0.96	0.56	0.52	0.96	0.13	0.30	**1.00**	**1.00**	**1.00**	**0.34**	**0.88**	**0.88**
Conventional strawberries	0.74	0.99	1.00	0.53	0.72	0.54	0.79	1.00	1.00	0.51	0.70	0.36	0.67	0.61	0.86	0.67	0.33	0.17	**1.00**	**1.00**	**1.00**	**0.49**	**0.97**	**0.66**
Other organics	0.54	0.97	0.99	0.16	0.99	0.95	0.95	1.00	1.00	0.84	0.56	0.89	0.82	0.84	0.75	0.82	0.32	0.50	0.78	0.99	1.00	0.56	0.68	0.70
Other conventional	0.81	1.00	1.00	0.43	0.65	0.55	1.00	1.00	1.00	0.95	0.44	0.70	1.00	0.99	1.00	1.00	0.34	0.26	1.00	1.00	1.00	0.26	0.74	0.90
Household – Economic variables																								
HH income	0.50	1.00	0.62	0.10	0.95	0.95	0.97	1.00	1.00	0.44	0.67	0.71	0.71	0.70	0.96	0.71	0.13	0.53	0.60	1.00	1.00	0.26	0.54	0.56
H of H female hours worked	0.58	0.72	0.74	0.26	0.41	0.52	0.55	0.61	0.44	0.34	0.50	0.45	0.58	0.50	0.63	0.58	0.31	0.21	0.63	0.71	0.54	0.36	0.66	0.53
H of H male hours worked	0.51	0.72	0.57	0.22	0.56	0.45	0.69	0.58	0.52	0.24	0.35	0.59	0.61	0.68	0.50	0.61	0.26	0.35	0.58	0.55	0.61	0.41	0.51	0.51
H of H female occupation	0.62	0.80	0.72	0.22	0.52	0.44	0.64	0.72	0.66	0.17	0.45	0.39	0.59	0.60	0.69	0.59	0.20	0.17	0.73	0.70	0.82	0.19	0.58	0.60
H of H male occupation	0.70	0.84	0.77	0.24	0.65	0.52	0.74	0.76	0.77	0.15	0.51	0.40	0.71	0.71	0.68	0.71	0.22	0.21	0.81	0.80	0.83	0.26	0.61	0.64
HH composition	0.54	0.66	0.66	0.16	0.66	0.42	0.71	0.65	0.65	0.27	0.39	0.62	0.64	0.46	0.58	0.64	0.20	0.29	0.58	0.64	0.83	0.22	0.57	0.61
HH Size	0.61	0.78	0.80	0.24	0.61	0.45	0.62	0.83	0.84	0.24	0.51	0.40	0.65	0.73	0.75	0.65	0.31	0.09	0.70	0.81	0.89	0.31	0.58	0.58
Residential type	0.67	0.85	0.85	0.28	0.64	0.63	0.82	0.81	0.94	0.24	0.61	0.55	0.82	0.73	0.80	0.82	0.27	0.20	0.79	0.82	0.93	0.41	0.76	0.67
Marital status	0.54	0.80	0.76	0.35	0.43	0.50	0.66	0.69	0.69	0.18	0.45	0.56	0.58	0.69	0.73	0.58	0.21	0.25	0.84	0.70	0.79	0.51	0.71	0.62
Children	0.69	0.91	0.87	0.26	0.65	0.48	0.85	0.86	0.69	0.19	0.65	0.52	0.75	0.69	0.76	0.75	0.24	0.13	0.82	0.94	0.95	0.25	0.66	0.67
H of H female age	0.65	0.74	0.72	0.23	0.56	0.59	0.72	0.79	0.71	0.30	0.55	0.50	0.73	0.64	0.61	0.73	0.30	0.20	0.75	0.73	0.82	0.39	0.61	0.65
H of H male age	0.66	0.71	0.78	0.21	0.53	0.48	0.73	0.76	0.71	0.36	0.52	0.51	0.81	0.66	0.68	0.81	0.20	0.16	0.78	0.71	0.75	0.33	0.66	0.63
H of H female education	0.73	0.72	0.83	0.13	0.50	0.56	0.73	0.82	0.84	0.37	0.55	0.48	0.75	0.61	0.67	0.75	0.28	0.19	0.82	0.77	0.83	0.25	0.61	0.70
H of H male education	0.79	0.78	0.71	0.29	0.58	0.48	0.84	0.80	0.76	0.31	0.52	0.47	0.77	0.67	0.64	0.77	0.27	0.20	0.76	0.77	0.80	0.34	0.65	0.64
Race Variables																								
Hispanic	0.61	0.78	0.91	0.26	0.96	0.89	0.95	0.93	0.86	0.19	0.52	0.61	0.80	0.67	0.69	0.80	0.31	0.34	0.99	0.85	0.90	0.41	0.79	0.59
Race	0.68	0.93	0.79	0.33	0.72	0.91	0.81	0.90	0.84	0.24	0.55	0.96	0.88	0.65	0.78	0.88	0.35	0.17	0.91	0.96	0.88	0.26	0.81	0.92
Season variables	0.56	0.80	0.79	0.27	0.57	0.52	0.65	0.82	0.59	0.19	0.34	0.36	0.58	0.81	0.82	0.58	0.18	0.21	0.70	0.87	0.85	0.36	0.50	0.46
Region Variables																								
Rural-Urban Continuum	0.74	0.85	0.86				0.83	0.95	0.75				0.79	0.77	0.82				0.79	0.87	0.90			
Scantrack Market	0.71	0.87	0.83				0.86	0.84	0.83				0.78	0.81	0.83				0.87	0.90	0.93			

Notes: If a variable is a categorical variable then each category less one is included in the count. For example, there are 10 female head of the household age categories. Therefore, this variable could be selected 101 x 9 = 909 times in the selection or quantity models. 'HH' indicates household, 'H of H' indicates head of household, 'RUCC' indicates the Rural-Urban Continuum Code (see https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/), and 'Market' refers to the 52 Scantrack markets.

Table 8: Estimated Conditional Elasticities (*$CPED_{ij}$* and *$CIED_i$*). Own-price elasticities are in bold.

	Organic fruit *i*	Est. Method	Price Elasticity w.r.t Organic Apples	w.r.t Organic Blueberries	w.r.t Organic Oranges	w.r.t Organic Strawberries	w.r.t. to Conventional *i*	Income Elasticity
Low income HH	**Apples**	**1**	**-1.74*** (0.55)**	-0.06 (0.09)	-0.05 (0.24)	0.44* (0.26)	-0.05 (0.09)	-0.02 (0.10)
		2	**-2.25* (1.34)**	0.01 (0.041)	-0.01 (0.44)	0.05 (0.09)	-0.08 (0.19)	-0.14 (0.64)
		3	**0.00 (0.68)**	0.00 (0.11)	0.00 (0.11)	0.00 (0.61)	0.00 (0.12)	0.00 (0.04)
	Blueberries	**1**	0.09 (0.20)	**-0.69*** (0.10)**	-0.22 (0.28)	-0.24 (0.37)	-0.08 (0.1)	-0.15 (0.18)
		2	0.38 (0.98)	**-0.69* (0.38)**	1.69 (1.57)	-0.35 (0.35)	0.43 (0.35)	0.05 (0.31)
		3	0.00 (0.14)	**-0.51*** (0.18)**	0.00 (0.25)	0.00 (0.49)	0.00 (0.08)	0.00 (0.17)
	Oranges	**1**	0.10 (0.18)	0.08 (0.08)	**-0.28 (0.22)**	-0.21 (0.18)	0.11 (0.14)	-0.20*** (0.08)
		2	-0.03 (0.39)	0.05 (0.06)	**0.08 (1.33)**	-0.20 (0.16)	0.40** (0.19)	-0.39 (0.50)
		3	-0.33 (0.24)	0.00 (0.06)	**0.00 (0.09)**	0.00 (0.09)	-0.03 (0.1)	0.00 (0.02)
	Strawberries	**1**	0.18 (0.11)	0.01 (0.06)	-0.17* (0.1)	**-0.34*** (0.09)**	-0.02 (0.11)	-0.04 (0.08)
		2	0.55 (0.57)	-0.10 (0.11)	-1.61 (1.32)	**-0.59 (0.88)**	0.22 (0.14)	-0.16 (0.30)
		3	0.00 (0.07)	0.00 (0.12)	-0.02 (0.18)	**-0.09 (0.1)**	0.04 (0.1)	0.00 (0.03)
Middle HH	**Apples**	**1**	**-2.32*** (0.55)**	0.05 (0.04)	0.15 (0.13)	-0.01 (0.13)	0.32*** (0.12)	0.27** (0.12)
		2	**-3.00** (1.44)**	0.011 (0.015)	0.17 (0.166)	0.027 (0.033)	-0.40*** (0.13)	0.25 (0.27)
		3	**-2.51*** (0.29)**	0.05 (0.07)	0.00 (0.13)	-0.22 (0.14)	0.36** (0.17)	0.08 (0.13)
	Blueberries	**1**	0.17** (0.08)	**-0.79*** (0.04)**	0.03 (0.14)	0.06 (0.08)	-0.11** (0.05)	-0.05 (0.11)
		2	0.10 (0.33)	**-1.63*** (0.19)**	0.22 (0.543)	-0.19** (0.091)	0.13*** (0.05)	-0.09 (0.24)
		3	0.11 (0.12)	**-0.58*** (0.07)**	0.00 (0.29)	0.00 (0.11)	-0.08 (0.06)	-0.02 (0.13)
	Oranges	**1**	0.16* (0.08)	0.04 (0.04)	**-0.13** (0.06)**	0.16* (0.09)	-0.06 (0.07)	-0.08 (0.09)
		2	0.23 (0.17)	0.001 (0.02)	**-0.32 (0.41)**	0.10** (0.04)	0.28*** (0.06)	-0.55 (0.51)
		3	0.00 (0.05)	0.00 (0.04)	**0.00 (0.10)**	0.00 (0.03)	0.00 (0.05)	0.00 (0.03)
	Strawberries	**1**	0.02 (0.04)	-0.04 (0.03)	0.09* (0.05)	**-0.17** (0.07)**	-0.09 (0.05)	-0.04 (0.09)
		2	0.21 (0.26)	-0.08** (0.03)	0.88** (0.36)	**-1.90*** (0.31)**	0.02 (0.07)	0.16 (0.20)
		3	0.00 (0.05)	0.00 (0.03)	0.00 (0.05)	**-0.06 (0.07)**	-0.09 (0.07)	0.00 (0.04)
Rich HH	**Apples**	**1**	**-1.35*** (0.27)**	0.01 (0.03)	0.04 (0.07)	0.17** (0.08)	0.18 (0.12)	0.01 (0.08)
		2	**-0.87* (0.45)**	0.01 (0.01)	0.23 (0.15)	-0.04* (0.03)	-0.17*** (0.04)	0.18 (0.18)
		3	**-1.65*** (0.23)**	0.00 (0.03)	0.00 (0.1)	0.00 (0.11)	0.23* (0.13)	-0.21 (0.14)
	Blueberries	**1**	0.135*** (0.05)	**-0.75*** (0.03)**	0.10* (0.06)	-0.055 (0.06)	0.03 (0.03)	0.11** (0.05)
		2	0.20 (0.24)	**-1.02*** (0.10)**	-0.02 (0.40)	-0.09 (0.09)	0.49*** (0.11)	0.10 (0.15)
		3	0.16** (0.08)	**-0.67*** (0.04)**	0.01 (0.05)	0.00 (0.04)	0.00 (0.03)	0.00 (0.08)
	Oranges	**1**	0.09 (0.07)	0.03 (0.05)	**-0.04 (0.11)**	0.02 (0.09)	0.11 (0.07)	-0.05 (0.11)
		2	0.28** (0.14)	0.003 (0.02)	**0.09 (0.31)**	0.04 (0.05)	0.05 (0.06)	-0.17 (0.36)
		3	0.00 (0.05)	0.00 (0.03)	**0.00 (0.1)**	0.00 (0.07)	0.00 (0.08)	0.00 (0.11)
	Strawberries	**1**	-0.05 (0.03)	0.00 (0.02)	-0.07 (0.05)	**-0.16*** (0.06)**	0.00 (0.07)	0.06 (0.06)
		2	-0.28 (0.18)	-0.02 (0.03)	0.43 (0.35)	**-0.81*** (0.30)**	-0.09 (0.06)	0.05 (0.14)
		3	-0.07 (0.04)	0.00 (0.02)	-0.11** (0.05)	**-0.06 (0.06)**	0.00 (0.05)	0.00 (0.04)

Notes: Standard errors of estimates are in parentheses. Est. method 1 refers to individual fruit Heckman models of consumption (see section 4.2). Est. method 2 refers to the incomplete demand system of consumption (see section 4.3). The elasticities from est. methods 1 and 2 assume the Boston market. Est. method 3 refers to the LASSO model (see section 4.4). '***" indicates significant at $p < 0.01$; '**" indicates significant at $p < 0.05$; and '*" indicates significant at $p < 0.10$

Table 9: Estimated Unconditional Elasticities (*$UPED_{ij}$* and *$UIED_i$*). Own-price elasticities are in bold bolded

	Organic fruit *i*	Est. Method	Price elasticity w.r.t Organic Apples	w.r.t Organic Blueberries	w.r.t Organic Oranges	w.r.t Organic Strawberries	w.r.t. to Conventional *i*	Income Elasticity
Low income HH	**Apples**	**1**	**-1.05 (1.00)**	-0.15 (0.25)	-0.23 (0.52)	0.71 (0.53)	1.06*** (0.29)	0.02 (0.52)
		2	**-0.90 (0.66)**	1.11*** (0.11)	0.18 (0.19)	0.66*** (0.12)	0.14* (0.08)	-0.04 (0.28)
		3	**0.00 (0.76)**	0.00 (0.11)	0.00 (0.11)	0.00 (0.63)	0.00 (0.14)	0.00 (0.04)
	Blueberries	**1**	-0.36 (0.73)	**-0.69 (0.59)**	-0.85 (0.69)	-0.72 (0.87)	1.14* (0.69)	-0.65 (0.58)
		2	0.071 (0.08)	**-0.10 (0.17)**	0.162 (0.12)	0.03 (0.09)	0.25** (0.10)	-0.22 (0.15)
		3	0.00 (0.25)	**-0.55* (0.32)**	0.00 (0.42)	0.00 (0.64)	0.00 (0.18)	0.00 (0.29)
	Oranges	**1**	-0.06 (0.49)	-0.36 (0.49)	**-0.81 (1.25)**	-0.28 (0.79)	1.12 (1.07)	-0.03 (0.49)
		2	-0.52* (0.28)	-2.39*** (0.23)	**-0.39 (0.82)**	-1.81*** (0.32)	-0.15* (0.09)	-0.13 (0.29)
		3	-0.32 (0.24)	0.00 (0.06)	**0.00 (0.11)**	0.00 (0.13)	-0.03 (0.13)	0.00 (0.00)
	Strawberries	**1**	0.57 (0.66)	-0.44 (0.49)	-1.05 (1.06)	**-2.84 (2.26)**	2.85* (1.48)	-0.53 (0.57)
		2	-0.09 (0.10)	-1.00** (0.09)	-0.39** (0.19)	**-0.77* (0.42)**	-0.22*** (0.04)	-0.12 (0.13)
		3	0.00 (0.08)	0.00 (0.14)	-0.01 (0.21)	**-0.02 (0.13)**	0.01 (0.11)	0.00 (0.04)
Middle HH	**Apples**	**1**	**-1.03* (0.54)**	-0.03 (0.07)	0.45** (0.19)	-0.07 (0.16)	1.10*** (0.10)	0.56*** (0.20)
		2	**-1.62 (1.01)**	2.62*** (0.06)	0.54*** (0.10)	1.44*** (0.07)	0.17** (0.08)	0.52*** (0.17)
		3	**-2.41*** (0.30)**	0.05 (0.07)	0.00 (0.12)	-0.21 (0.14)	0.34** (0.17)	0.07 (0.11)
	Blueberries	**1**	0.61 (0.37)	**-3.05*** (1.02)**	-0.20 (0.5)	-0.22 (0.31)	1.54*** (0.42)	0.48 (0.46)
		2	0.00 (0.04)	**-1.10*** (0.12)**	0.01 (0.06)	-0.11*** (0.03)	0.00 (0.02)	0.20 (0.13)
		3	0.13 (0.14)	**-0.73*** (0.10)**	0.00 (0.35)	0.00 (0.13)	-0.10 (0.07)	-0.02 (0.14)
	Oranges	**1**	0.37 (0.33)	-0.07 (0.16)	**-0.66 (0.56)**	0.35 (0.34)	-0.21 (0.32)	-0.49 (0.53)
		2	-0.25* (0.13)	-2.26*** (0.10)	**-0.61** (0.29)**	-1.00*** (0.10)	-0.15*** (0.03)	-0.38 (0.34)
		3	0.00 (0.05)	0.00 (0.04)	**0.00 (0.10)**	0.00 (0.03)	0.00 (0.05)	0.00 (0.03)
	Strawberries	**1**	0.11 (0.25)	-0.23 (0.15)	0.44 (0.36)	**-7.33*** (1.8)**	2.95*** (0.73)	1.02** (0.47)
		2	-0.22*** (0.06)	-1.47*** (0.04)	-0.06 (0.07)	**-1.97*** (0.20)**	-0.41*** (0.03)	0.34*** (0.10)
		3	0.00 (0.06)	0.00 (0.03)	0.00 (0.05)	**-0.07 (0.1)**	-0.10 (0.08)	0.00 (0.04)
Rich HH	**Apples**	**1**	**0.08 (0.47)**	0.08 (0.09)	0.11 (0.21)	0.29 (0.22)	0.82*** (0.13)	0.54* (0.29)
		2	**-0.02 (0.31)**	3.12*** (0.04)	0.67*** (0.09)	1.57*** (0.05)	0.37*** (0.02)	0.33*** (0.12)
		3	**-1.57*** (0.21)**	0.00 (0.02)	0.00 (0.09)	0.00 (0.1)	0.21* (0.12)	-0.19 (0.12)
	Blueberries	**1**	0.30* (0.17)	**-1.00*** (0.19)**	0.17 (0.21)	0.06 (0.24)	1.52*** (0.30)	0.75*** (0.25)
		2	0.05* (0.03)	**-0.52*** (0.06)**	0.02 (0.04)	0.04 (0.03)	0.27*** (0.04)	0.45*** (0.08)
		3	0.17** (0.08)	**-0.75*** (0.04)**	0.02 (0.05)	0.00 (0.04)	0.00 (0.03)	0.00 (0.09)
	Oranges	**1**	0.06 (0.3)	-0.17 (0.16)	**-0.12 (0.47)**	0.24 (0.41)	0.37 (0.35)	-0.03 (0.46)
		2	-0.04 (0.11)	-1.36*** (0.07)	**-0.17 (0.22)**	-0.66*** (0.10)	-0.15*** (0.03)	0.02 (0.24)
		3	0.00 (0.05)	0.00 (0.03)	**0 (0.11)**	0.00 (0.07)	0.00 (0.09)	0.00 (0.11)
	Strawberries	**1**	-0.13 (0.13)	-0.09 (0.1)	-0.10 (0.22)	**-2.26*** (0.61)**	1.83*** (0.45)	0.71*** (0.26)
		2	-0.23*** (0.05)	-0.89*** (0.04)	-0.05 (0.08)	**-1.03*** (0.21)**	-0.30*** (0.03)	0.39*** (0.08)
		3	-0.08 (0.05)	0.00 (0.02)	-0.12** (0.06)	**-0.07 (0.07)**	0.00 (0.06)	0.00 (0.04)

Notes: Standard errors of estimates are in parentheses. Est. method 1 refers to individual fruit Heckman models of consumption (see section 4.2). Est. method 2 refers to the incomplete demand system of consumption (see section 4.3). The elasticities from est. methods 1 and 2 assume the Boston market. Est. method 3 refers to the LASSO model (see section 4.4). '***' indicates significant at $p < 0.01$; '**' indicates significant at $p < 0.05$; and '*' indicates significant at $p < 0.10$

Table 10: Expected Conditional Consumer Reaction to a Subsidy of Organic Fruit or Tax on Conventional Fruit. Under the subsidy plan all organic fruit prices are 90% of observed means. Under the tax plan all conventional fruit prices are 110% of observed means.

			Purchased (ounces / month)			**Subsidy Elasticities**		**Tax Elasticity**
	Organic fruit *i*	**Est. Method**	No subsidy or tax	Subsidy on all organics	Tax on all conventionals	Purchase	Expenditure	Purchase and Expenditure
Low income	**Apples**	**1**	73.95	86.95	70.94	-1.76	-0.58	-0.41
		2	86.84	105.27	86.77	-2.12	-0.91	-0.01
		3	82.98	82.98	82.98	0.00	1.00	0.00
	Blueberries	**1**	18.28	18.22	20.75	0.03	1.03	1.35
		2	22.32	22.77	22.59	-0.20	0.82	0.12
		3	19.00	19.27	21.90	-0.14	0.87	1.53
	Oranges	**1**	69.48	66.87	70.70	0.37	1.34	0.18
		2	72.25	75.62	88.79	-0.47	0.58	2.29
		3	71.51	73.82	71.87	-0.32	0.71	0.05
	Strawberries	**1**	25.34	26.09	25.13	-0.30	0.73	-0.08
		2	28.64	29.18	31.83	-0.19	0.83	1.11
		3	24.71	25.03	25.05	-0.13	0.88	0.14
Middle	**Apples**	**1**	83.33	105.65	84.59	-2.68	-1.41	0.15
		2	125.70	152.62	120.98	-2.14	-0.93	-0.38
		3	88.62	117.81	92.05	-3.29	-1.96	0.39
	Blueberries	**1**	17.07	17.37	17.44	-0.17	0.85	0.21
		2	19.67	19.82	19.73	-0.08	0.93	0.03
		3	17.42	18.23	17.34	-0.46	0.58	-0.05
	Oranges	**1**	71.96	70.85	72.61	0.15	1.14	0.09
		2	73.61	70.99	78.63	0.36	1.32	0.68
		3	73.00	73.00	73.00	0.00	1.00	0.00
	Strawberries	**1**	25.75	25.94	25.79	-0.07	0.94	0.01
		2	28.12	26.23	28.82	0.67	1.61	0.25
		3	26.26	26.41	26.02	-0.06	0.95	-0.09
Rich	**Apples**	**1**	77.70	88.25	81.68	-1.36	-0.22	0.51
		2	90.49	96.93	88.84	-0.71	0.36	-0.18
		3	82.95	99.08	85.52	-1.94	-0.75	0.31
	Blueberries	**1**	17.21	17.58	17.47	-0.22	0.80	0.15
		2	20.89	21.69	21.56	-0.38	0.65	0.32
		3	16.65	17.22	16.97	-0.35	0.69	0.19
	Oranges	**1**	70.00	69.62	73.90	0.06	1.05	0.56
		2	73.79	70.19	74.71	0.49	1.44	0.12
		3	68.30	68.30	68.30	0.00	1.00	0.00
	Strawberries	**1**	26.70	27.49	27.78	-0.30	0.73	0.40
		2	29.18	29.15	30.06	0.01	1.01	0.30
		3	26.51	27.16	26.87	-0.25	0.78	0.14

Table 11: Expected Unconditional Consumer Reaction to a Subsidy of Organic Fruit or Tax on Conventional Fruit. Under the subsidy plan all organic fruit prices are 90% of observed means. Under the tax plan all conventional fruit prices are 110% of observed means.

			Purchased (ounces / month)			**Subsidy Elasticities**		**Tax Elasticity**
	Organic fruit *i*	**Est. Method**	No subsidy or tax	Subsidy on all organics	Tax on all conventionals	Purchase	Expenditure	Purchase and Expenditure
Low income	**Apples**	**1**	0.13	0.14	0.15	-0.80	0.28	1.91
		2	0.10	0.11	0.12	-1.02	0.08	1.22
		3	0.21	0.20	0.24	0.34	1.31	1.32
	Blueberries	**1**	0.02	0.03	0.03	-1.30	-0.17	4.60
		2	0.02	0.03	0.03	-0.42	0.62	2.86
		3	0.04	0.05	0.06	-1.84	-0.65	2.65
	Oranges	**1**	0.15	0.14	0.19	0.51	1.46	2.18
		2	0.12	0.14	0.16	-1.09	0.02	2.76
		3	0.06	0.06	0.07	0.12	1.10	0.66
	Strawberries	**1**	0.05	0.07	0.10	-4.18	-2.76	8.76
		2	0.07	0.09	0.11	-3.59	-2.23	6.26
		3	0.12	0.15	0.15	-2.57	-1.31	2.40
Middle	**Apples**	**1**	0.18	0.19	0.23	-0.76	0.32	2.93
		2	0.15	0.17	0.18	-0.97	0.13	2.12
		3	0.38	0.47	0.46	-2.44	-1.19	2.23
	Blueberries	**1**	0.02	0.02	0.03	-1.81	-0.63	5.48
		2	0.02	0.03	0.03	-2.59	-1.33	2.33
		3	0.05	0.06	0.05	-2.40	-1.16	1.40
	Oranges	**1**	0.10	0.10	0.12	-0.27	0.76	1.67
		2	0.09	0.09	0.10	-0.02	0.98	2.02
		3	0.08	0.08	0.09	-0.04	0.97	1.14
	Strawberries	**1**	0.06	0.10	0.10	-6.14	-4.53	6.06
		2	0.09	0.13	0.12	-5.11	-3.60	3.05
		3	0.16	0.21	0.18	-3.31	-1.98	1.39
Rich	**Apples**	**1**	0.19	0.18	0.25	0.75	1.68	2.82
		2	0.21	0.20	0.24	0.45	1.40	1.61
		3	0.48	0.51	0.58	-0.84	0.25	2.31
	Blueberries	**1**	0.05	0.05	0.07	-0.27	0.75	4.07
		2	0.06	0.06	0.08	-0.60	0.46	2.69
		3	0.11	0.12	0.13	-1.44	-0.29	1.79
	Oranges	**1**	0.22	0.22	0.27	-0.39	0.64	2.53
		2	0.19	0.19	0.22	0.03	1.03	1.76
		3	0.09	0.09	0.10	-0.10	0.91	1.23
	Strawberries	**1**	0.22	0.29	0.33	-3.22	-1.90	5.36
		2	0.25	0.30	0.33	-2.10	-0.89	3.23
		3	0.31	0.38	0.38	-2.17	-0.95	2.46

Appendix for US Household Demand for Organic Fruit

1. Imputed prices

Using the raw Nielsen data on household expenditures and purchases on all consumable items, we first created household-month expenditure (represented with *e* and measured in dollars) and purchase (represented with *o* and measured in ounces) variables for 20 fruit type × varieties,

$$e_{fkm} = \sum_{d \in m} e_{fkd} \tag{A}$$

$$o_{fkm} = \sum_{d \in m} o_{fkd} \tag{B}$$

where *f* indexes fruit type × variety, *d* indexes each shopping trip taken that ended up in a fruit purchase, and *m* indexes months in the years 2011, 2012, and 2013. In some cases purchases were reported in the number of items purchased, not ounces purchased. In these cases we had to first convert number of items purchased to ounces purchased before we could add the purchase to other purchases. The weight per item of fruit is given in Table B.

Note that e_{fkm} and o_{fkm} for two fruit type × varieties, 'other × organic' and 'other × conventional', are created by summing *km*'s monthly expenditures and purchases of minor organic fruit types and minor conventional fruit types, respectively. See section XX of the paper for more details.

Then we calculated the nominal price per ounce of *f* faced by household *k* in month *m* for each unique *fkm* combination,

$$P_{fkm} = e_{fkm} / o_{fkm} \tag{C}$$

When e_{fkm} and o_{fkm} were 0 (household *k* did not purchase *f* during the month *m*) we had to impute P_{fkm}.

We created two sets of imputed prices. The first set of imputed prices were created with the following method. Let, $\hat{P}_{fmry}$ be the imputed price of fruit type × variety *f* in month *m* in market *r* in year *y*,

$$\hat{P}_{fmry} = \frac{\sum_{km \in y}[w_{ky} I(k \in r, y) e_{fkm}]}{\sum_{km \in y}[w_{ky} I(k \in r, y) I(e_{fkm} > 0)]} \Bigg/ \frac{\sum_{km} w_{km} I(km \in r) o_{fkm}}{\sum_{km} w_{km} I(km \in r) I(e_{fkm} > 0)} \tag{D}$$

where $km \in y$ is the set of all *km* observations in year *y*, w_{ky} is household *k*'s projection factor in year *y*, $I(k \in r, y) = 1$ if household *k* resides in market *r* in year *y* and equals 0 otherwise, and $I(e_{fkm} > 0) = 1$ if $e_{fkm} > 0$ and equals 0 otherwise. Therefore, $\hat{P}_{fmry}$ is the (weighted) average price of *f* across all household-month purchases of *f* in month *m* of year *y* in market *r*. When $\sum_{km \in y}[w_{ky}I(k \in r, y)I(e_{fkm} > 0)] = 0$ for fruit *f* then $\hat{P}_{fmry}$ does not exist. This occurs when no household *k* that resides in market *r* in year *y* purchases *f* in month *m*.

We created another set of imputed prices with the following method. The observed price of fruit type × variety *f* in month *m* in in market *r* in year *y* is explained by the following model,

$$E[P_{fkmy}] = \alpha + \sum_{m=2}^{12} \beta_m I(m) + \sum_{r=2}^{76} \gamma_r I(r) + \sum_{f=2}^{20} \theta_f I(f) \qquad \text{(E)}$$

where *I*(*m*) is a dummy variable that equals 1 if the observed month is equal to *m* and equals 0 otherwise, *I*(*r*) is a dummy variable that equals 1 if the observed household *k* is in region *r* in year *y* and equals 0 otherwise, and *I*(*f*) is a dummy variable that equals 1 if the observed fruit is equal to *f* and equals 0 otherwise. The set of α, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\theta}$ are model coefficients to be estimated. We estimate (Z) for each year in our dataset using weighted OLS where *k*'s projection factor in year *y* as the weight (i.e., we use the [pweight=projection factor] option after reg in Stata). Let the expected price of fruit *f* in month *m* in region *r* in year y be given by,

$$\hat{\hat{P}}_{fmry} = \hat{\alpha} + \hat{\beta}_m + \hat{\gamma}_r + \hat{\theta}_f \qquad \text{(F)}$$

where $\hat{\beta}_m = 0$ if *m* = 1, $\hat{\gamma}_r = 0$ if *r* = 1, and $\hat{\theta}_f = 0$ if *f* = 1.

If P_{fkm} does not exist then P_{fkm} is set equal to the appropriate $\hat{P}_{fmry}$ (household-month *km* is assigned the $\hat{P}_{fmry}$ that matches *km*'s market of residence and the month of time). If the appropriate $\hat{P}_{fmry}$ does not exist then P_{fkm} is set equal to the appropriate $\hat{\hat{P}}_{fmry}$.

2. Data manipulation

We reduce the 20 fruit type × varieties to 10 fruit type × varieties before we conduct numerical analysis. We do this by first combining *km*'s expenditures on and ounces purchased of blackberry × conventional; grape × conventional, grapefruit × conventional, lemon × conventional, raspberry × conventional, and other× conventional. Let this new fruit type × variety be known as other× conventional. Next we combine *km*'s expenditures on and ounces purchased of blackberry × organic; grape × organic, grapefruit × organic, lemon × organic, raspberry × organic, and other× organic. Let this new fruit type × variety be known as other×

organic. Now we have ten e_{ikm} and o_{ikm} variables where i indexes the set of 10 fruit type × varieties.

To calculate P_{ikm} for i = other× organic for household k in month m we do the following. First let u index the 6 organic fruit types that we collapsed into one category. Define $\bar{e}_{ukm\in y}$ as,

$$\bar{e}_{uy} = \frac{1}{N}\sum_{km\in y}^{N} e_{ukm} \tag{G}$$

$\bar{e}_{Ukm\in y}$ as,

$$\bar{e}_{Uy} = \frac{1}{N}\left(\sum_{u=1}^{6}\sum_{km\in y}^{N} e_{ukm}\right) \tag{H}$$

and,

$$share_{uy} = \bar{e}_{uy}/\bar{e}_{Uy} \tag{I}$$

Finally,

$$P_{ikm} = \sum_{u=1}^{6} share_{uy}\, P_{ukm\in y} \tag{J}$$

for i = other× organic. To calculate P_{ikm} for i = other× conventional we repeat (G) - (J) using the appropriate expenditures and prices.

In the Nielsen data annual household income is coded in categories. We recode income categories using the following where the number before the equal sign is the category and the number after the equal sign is the nominal dollar amount we assumed,

(3=2500); (4=6500); (6=9000); (8=11000); (10=13500); (11=17500); (13=22500); (15=27500); (16=32500); (17=37500); (18=42500); (19=47500); (21=55000); (23=65000); (26=80000); (27=150000)

We convert annual household income to monthly real household income by dividing by 12 and then inflating according to the CPI.

3. Details on estimation method 1

3.1. Conditional expectations and elasticities

Below are the steps we took to estimate conditional expectations and elasticities using estimation method 1.

1. We drop all households that never bought fruit (organic or conventional) during a calendar year.
2. We drop all households that do not belong to income class *Z*.
3. Let the number of household-month observations that remain be given by given by N_Z.
4. We estimate the probit for fruit type × variety *i* (estimate equation (3) across N_Z household-month observations). Independent variables include monthly household income, at least one child in the house, at least one head of household with a college degree, is the household in a metro county, is the household lead by a married couple, racial dummies, seasonal dummies interacted with year, all market dummies, and all real prices. This gives us the estimates of $\boldsymbol{\alpha}_i, \boldsymbol{\mu}_i, \boldsymbol{\omega}_i$, given by $\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\omega}}_i$.
5. We estimate the inverse mills ratio for each household-month observation across all N_Z household-month observations. The mean mills ratio across all N_Z is given by $\bar{\hat{\lambda}}$.
6. We drop all household-month observations where ounces of fruit type × variety *i* bought are 0. Let the number of household month observations where $o_{ikm} > 0$ be given by given by n_Z.
7. We regress ounces of fruit type × variety *i* bought on household characteristic, seasonal and year dummies, real prices, and the estimated inverse mills ratio across n_Z household-month observations. This gives us the estimates of $\boldsymbol{\beta}_i, \boldsymbol{\sigma}_i, \boldsymbol{\theta}_i$, given by $\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\sigma}}_i, \hat{\boldsymbol{\theta}}_i$.
8. We use the Stata command suest to calculate robust standard errors that are clustered on the individual.
9. Using equations (3) - (7) we can derive monthly conditional price and income elasticity of demand for each fruit type × variety *i* and income class *Z* combination (Saha et al. 1997),

$$\begin{aligned} CPED_{ij} &= \frac{\partial E\left[o_{ikm} \mid Y_{ikm}=1\right]}{\partial p_j} \frac{P_j}{E\left[o_{ikm} \mid Y_{ikm}=1\right]} \\ &= \left[\hat{\theta}_{ij} - \hat{\gamma}_i \hat{\omega}_{ij}\left(\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)\bar{\hat{\lambda}}_i + \left(\bar{\hat{\lambda}}_i\right)^2\right)\right]\frac{\bar{P}_j}{\bar{o}_i} \end{aligned} \tag{K}$$

$$\begin{aligned} CIED_{i} &= \frac{\partial E\left[o_{ikm} \mid Y_{ikm}=1\right]}{\partial s} \frac{s}{E\left[o_{ikm} \mid Y_{ikm}=1\right]} \\ &= \left[\hat{\beta}_{is} - \hat{\gamma}_i \hat{\alpha}_{is}\left(\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)\bar{\hat{\lambda}}_i + \left(\bar{\hat{\lambda}}_i\right)^2\right)\right]\frac{\bar{s}}{\bar{o}_i} \end{aligned} \tag{L}$$

where j also indexes fruit type × variety (where $i = j$ is own-price elasticity), $\hat{\theta}_{ij}$ is the estimated coefficient for price P_j from the ounces purchased equation regression, $\hat{\gamma}_i$ is the estimated coefficient for the inverse mills ratio from the ounces purchased equation regression, $\hat{\omega}_{ij}$ is the estimated coefficient for price P_j from the selection equation regression, $\hat{\beta}_{is}$ is the estimated coefficient for real monthly household income from the ounces purchased equation regression, $\hat{\alpha}_{is}$ is the estimated coefficient for real monthly household income from the selection equation regression, $\bar{s}$ is average real monthly income, $\bar{P}_j$ is average monthly price of fruit type × variety j, $\bar{o}_i$ is average ounces of fruit type × variety i bought by a household in a month, and $\bar{\mathbf{X}}, \bar{\mathbf{C}}$, and $\bar{\mathbf{P}}$ are mean vectors. Means are generated across all N_Z observations (see Table A). We do make one change to $\bar{\mathbf{C}}$: all market dummies are set equal to 0 expect for Boston's, which is set equal to 1. Finally, $\bar{\hat{\lambda}}_i$ is given by,

$$\bar{\hat{\lambda}}_i = \phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right) \Big/ \Phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right) \quad \text{(M)}$$

10. We use Stata's nlcom to find $CPED_{ij}$ and $CIED_i$ and their standard errors. nlcom cannot be used to estimate (A)-(C) if all the market dummies are included. That is why we assume the Boston market in the estimates of (A)-(C). Further, the standard errors for the estimate of $E\left[o_{ikm} \mid Y_{ikm} = 1\right]$ (and thus the standard errors for $CPED_{ij}$ and $CIED_i$ generated with nlcom) are not quite correct given we treat estimated $\hat{\lambda}_{ikm}$ as a constant instead of the random variable it is. The two stage Heckman command in Stata would produce correct standard errors. However, that two stage Heckman command does not lend itself to the derivation of $CPED_{ij}$ and $CIED_i$ as detailed in Saha et al. (1997).

3.2. Unconditional expectations and elasticities and purchase probabilities

Below are the steps we took to estimate unconditional expectations and elasticities using estimation method 1.

1. We drop all households that never bought fruit (organic or conventional) during a calendar year.
2. We drop all households that do not belong to income class Z.
3. Let the number of household-month observations that remain be given by given by N_Z.
4. We estimate the probit for fruit type × variety i (estimate equation (3) across N_Z household-month observations). This gives us the estimates of $\boldsymbol{\alpha}_i, \boldsymbol{\mu}_i, \boldsymbol{\omega}_i$, given by $\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\omega}}_i$.

5. We estimate the inverse mills ratio for each household-month observation across all N_Z household-month observations. The mean mills ratio across all N_Z is given by $\bar{\hat{\lambda}}$.
11. We regress ounces of fruit type × variety *i* bought on CDF of probit estimate times independent variables on household characteristics, seasonal x year dummies, real prices, and the estimated PDFs of the probit estimate across N_Z household-month observations. This gives us the estimates of $\boldsymbol{\beta}_i, \boldsymbol{\sigma}_i, \boldsymbol{\theta}_i$, given by $\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\sigma}}_i, \hat{\boldsymbol{\theta}}_i$.
6. We use the Stata command suest to calculate robust standard errors that are clustered on the individual.
7. Using equations (3) - (7) we can derive monthly unconditional price and income elasticity of demand for each fruit type × variety *i* and income class *Z* combination (Saha et al. 1997),

$$UPED_{ij} = \left[\hat{\omega}_{ij} \phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)\left(\bar{\mathbf{X}}\hat{\boldsymbol{\beta}}_i + \bar{\mathbf{c}}\hat{\boldsymbol{\sigma}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\theta}}_i + \hat{\gamma}_i \bar{\hat{\lambda}}\right) + \right.$$
$$\left. \Phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)\left(\hat{\theta}_{ij} - \hat{\gamma}_i \hat{\omega}_{ij}\left(\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)\bar{\hat{\lambda}}_i + \left(\bar{\hat{\lambda}}_i\right)^2\right)\right)\frac{\bar{P}_j}{\bar{o}_i}\right] \tag{N}$$

$$UIED_{i} = \left[\hat{\omega}_{is} \phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)\left(\bar{\mathbf{X}}\hat{\boldsymbol{\beta}}_i + \bar{\mathbf{c}}\hat{\boldsymbol{\sigma}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\theta}}_i + \hat{\gamma}_i \bar{\hat{\lambda}}\right) + \right.$$
$$\left. \Phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)\left(\hat{\beta}_{is} - \hat{\gamma}_i \hat{\omega}_{is}\left(\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)\bar{\hat{\lambda}}_i + \left(\bar{\hat{\lambda}}_i\right)^2\right)\right)\frac{\bar{s}}{\bar{o}_i}\right] \tag{O}$$

where *j* also indexes fruit type × variety (where *i* = *j* is own-price elasticity), $\hat{\theta}_{ij}$ is the estimated coefficient for price P_j from the ounces purchased equation regression, $\hat{\gamma}_i$ is the estimated coefficient for the inverse mills ratio from the ounces purchased equation regression, $\hat{\omega}_{ij}$ is the estimated coefficient for price P_j from the selection equation regression, $\hat{\beta}_{is}$ is the estimated coefficient for real monthly household income from the ounces purchased equation regression, $\hat{\alpha}_{is}$ is the estimated coefficient for real monthly household income from the selection equation regression, $\bar{s}$ is average real monthly income, $\bar{P}_j$ is average monthly price of fruit type × variety *j*, $\bar{o}_i$ is average ounces of fruit type × variety *i* bought by a household in a month, and $\bar{\mathbf{X}}, \bar{\mathbf{C}}$, and $\bar{\mathbf{P}}$ are mean vectors. Means are generated across all N_Z observations. We do make one change to $\bar{\mathbf{C}}$: all market dummies are set equal to 0 expect for Boston's, which is set equal to 1. Finally, $\bar{\hat{\lambda}}_i$ is given by,

$$\bar{\bar{\lambda}}_i = \phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right) \Big/ \Phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right) \tag{P}$$

where again the market is set equal to Boston.

8. Unfortunately formulas (N)-(O) are too large for Stata's nlcom function to process. Therefore, we estimate use nlcom to estimate the unconditional arc elasticities:

$$UPED_{ij} = \frac{\dfrac{o_i(\bar{\mathbf{X}},\bar{\mathbf{C}},\tilde{\mathbf{P}},\bar{\mathbf{c}},\hat{\boldsymbol{\alpha}}_i,\hat{\boldsymbol{\mu}}_i,\hat{\boldsymbol{\omega}}_i,\hat{\boldsymbol{\beta}}_i,\hat{\boldsymbol{\sigma}}_i,\hat{\boldsymbol{\theta}}_i) - o_i(\bar{\mathbf{X}},\bar{\mathbf{C}},\bar{\mathbf{P}},\bar{\mathbf{c}},\hat{\boldsymbol{\alpha}}_i,\hat{\boldsymbol{\mu}}_i,\hat{\boldsymbol{\omega}}_i,\hat{\boldsymbol{\beta}}_i,\hat{\boldsymbol{\sigma}}_i,\hat{\boldsymbol{\theta}}_i)}{o_i(\bar{\mathbf{X}},\bar{\mathbf{C}},\bar{\mathbf{P}},\bar{\mathbf{c}},\hat{\boldsymbol{\alpha}}_i,\hat{\boldsymbol{\mu}}_i,\hat{\boldsymbol{\omega}}_i,\hat{\boldsymbol{\beta}}_i,\hat{\boldsymbol{\sigma}}_i,\hat{\boldsymbol{\theta}}_i)}}{-0.10} \tag{Q}$$

$$UIED_{i} = \frac{\dfrac{o_i(\tilde{\mathbf{X}},\bar{\mathbf{C}},\bar{\mathbf{P}},\bar{\mathbf{c}},\hat{\boldsymbol{\alpha}}_i,\hat{\boldsymbol{\mu}}_i,\hat{\boldsymbol{\omega}}_i,\hat{\boldsymbol{\beta}}_i,\hat{\boldsymbol{\sigma}}_i,\hat{\boldsymbol{\theta}}_i) - o_i(\bar{\mathbf{X}},\bar{\mathbf{C}},\bar{\mathbf{P}},\bar{\mathbf{c}},\hat{\boldsymbol{\alpha}}_i,\hat{\boldsymbol{\mu}}_i,\hat{\boldsymbol{\omega}}_i,\hat{\boldsymbol{\beta}}_i,\hat{\boldsymbol{\sigma}}_i,\hat{\boldsymbol{\theta}}_i)}{o_i(\bar{\mathbf{X}},\bar{\mathbf{C}},\bar{\mathbf{P}},\bar{\mathbf{c}},\hat{\boldsymbol{\alpha}}_i,\hat{\boldsymbol{\mu}}_i,\hat{\boldsymbol{\omega}}_i,\hat{\boldsymbol{\beta}}_i,\hat{\boldsymbol{\sigma}}_i,\hat{\boldsymbol{\theta}}_i)}}{0.10} \tag{R}$$

where,

$$o_i(\bar{\mathbf{X}},\bar{\mathbf{C}},\bar{\mathbf{P}},\bar{\mathbf{c}},\hat{\boldsymbol{\alpha}}_i,\hat{\boldsymbol{\mu}}_i,\hat{\boldsymbol{\omega}}_i,\hat{\boldsymbol{\beta}}_i,\hat{\boldsymbol{\sigma}}_i,\hat{\boldsymbol{\theta}}_i) = \Phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)\left(\bar{\mathbf{X}}\hat{\boldsymbol{\beta}}_i + \bar{\mathbf{c}}\hat{\boldsymbol{\sigma}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\theta}}_i\right) +$$

$$\hat{\gamma}_i\phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right) \tag{S}$$

$\tilde{\mathbf{P}}$ is the same as $\bar{\mathbf{P}}$ except $\tilde{P}_j = 0.9\bar{P}_j$ and $\tilde{\mathbf{X}}$ is the same as $\bar{\mathbf{X}}$ except $\tilde{s} = 1.1\bar{s}$. Please note that the standard errors for the estimate of $E\left[o_{ikm}\right]$ (and thus the standard errors for *UPED_{ij}* and *UIED_i* generated with nlcom) are not quite correct given we treat estimated $\hat{\lambda}_{ikm}$ as a constant instead of the random variable it is.

9. The only exception to equation (G) is a tax on the conventional version of organic fruit *i*. Let the conventional version be indexed by *g*,

$$UPED_{ig} = \frac{\dfrac{o_i(\bar{\mathbf{X}},\bar{\mathbf{C}},\tilde{\mathbf{P}},\bar{\mathbf{c}},\hat{\boldsymbol{\alpha}}_i,\hat{\boldsymbol{\mu}}_i,\hat{\boldsymbol{\omega}}_i,\hat{\boldsymbol{\beta}}_i,\hat{\boldsymbol{\sigma}}_i,\hat{\boldsymbol{\theta}}_i) - o_i(\bar{\mathbf{X}},\bar{\mathbf{C}},\bar{\mathbf{P}},\bar{\mathbf{c}},\hat{\boldsymbol{\alpha}}_i,\hat{\boldsymbol{\mu}}_i,\hat{\boldsymbol{\omega}}_i,\hat{\boldsymbol{\beta}}_i,\hat{\boldsymbol{\sigma}}_i,\hat{\boldsymbol{\theta}}_i)}{o_i(\bar{\mathbf{X}},\bar{\mathbf{C}},\bar{\mathbf{P}},\bar{\mathbf{c}},\hat{\boldsymbol{\alpha}}_i,\hat{\boldsymbol{\mu}}_i,\hat{\boldsymbol{\omega}}_i,\hat{\boldsymbol{\beta}}_i,\hat{\boldsymbol{\sigma}}_i,\hat{\boldsymbol{\theta}}_i)}}{0.10} \tag{T}$$

where $\tilde{\mathbf{P}}$ is the same as $\bar{\mathbf{P}}$ except $\tilde{P}_g = 1.1\bar{P}_g$.

3.3. Purchase probabilities

Below are the steps we took to estimate conditional and unconditional purchase probabilities using estimation method 1.

1. We drop all households that never bought fruit (organic or conventional) during a calendar year.
2. We drop all households that do not belong to income class *Z*.
3. Let the number of household-month observations that remain be given by given by N_Z.
4. We estimate the probit for fruit type × variety *i* across N_Z household-month observations. This gives us the estimates of $\boldsymbol{\alpha}_i, \boldsymbol{\mu}_i, \boldsymbol{\omega}_i$, given by $\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\omega}}_i$.
5. We estimate the inverse mills ratio for each household-month observation across all N_Z household-month observations. The mean mills ratio across all N_Z is given by $\bar{\hat{\lambda}}$.
6. We calculate the purchase probability elasticity with respect to price of type × variety *j* for type × variety *i* ($PPEP_{ij}$) over all N_Z observations,

$$PPEP_{ij} = \frac{\partial \Phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)}{\partial \bar{P}_j} \frac{\bar{P}_j}{\Phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)} \tag{U}$$

$$= \frac{\phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)\hat{\omega}_{ij}\bar{P}_j}{\Phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)} \tag{V}$$

$$= \bar{\hat{\lambda}}_i \hat{\omega}_{ij} \bar{P}_j \tag{W}$$

   where *i* = *j* is own-price elasticity, $\hat{\omega}_{ij}$ is the estimated coefficient for price P_j from the selection equation regression, $\bar{P}_j$ is average monthly price of fruit type × variety *j*, and $\bar{\mathbf{X}}, \bar{\mathbf{C}}$, and $\bar{\mathbf{P}}$ are mean vectors. Means are generated across all N_Z observations. We do make one change to $\bar{\mathbf{C}}$: all market dummies are set equal to 0 expect for Boston's, which is set equal to 1. As before the Stata command nlcom cannot be used to estimate (K)-(M) if all the market dummies are included. Please note that the standard errors for the estimate of $PPEP_{ij}$ generated with nlcom are not quite correct given we treat estimated $\hat{\lambda}_{ikm}$ as a constant instead of the random variable it is.
7. We calculate the purchase probability elasticity with respect to income for type × variety *i* ($PPEI_i$) over all N_Z observations,

$$PPEI_i = \frac{\partial \Phi\left(\overline{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \overline{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \overline{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)}{\partial s} \frac{\overline{s}}{\Phi\left(\overline{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \overline{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \overline{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)} \tag{X}$$

$$= \frac{\phi\left(\overline{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \overline{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \overline{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)\hat{\omega}_{is}\overline{s}}{\Phi\left(\overline{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \overline{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \overline{\mathbf{P}}\hat{\boldsymbol{\omega}}_i\right)} \tag{Y}$$

$$= \overline{\hat{\lambda}}_i \hat{\omega}_{is}\overline{s} \tag{Z}$$

where $\hat{\omega}_{is}$ is the estimated probit coefficient on *s* and $\overline{s}$ is the average monthly income across all N_Z observations. Again the the market in $\overline{\mathbf{C}}$ it set equal to Boston. Please note that the standard errors for the estimate of $PPEI_i$ generated with nlcom are not quite correct given we treat estimated $\hat{\lambda}_{ikm}$ as a constant instead of the random variable it is.

4. Details on estimation method 2

Below are the steps we took to estimate expectations and elasticities using estimation method 2.

1. We drop all households that never bought fruit (organic or conventional) during a calendar year.
2. We drop all households that do not belong to income class *Z*.
3. Let the number of household-month observations that remain be given by given by N_Z.
4. We estimate the probit for fruit type × variety *i*.
5. From the probit estimates we derive standard normal cdf and pdf values for each observation. These values are used in estimation of the full latent demand system using nlsur in Stata.
6. Then we estimate the latent demand system simultaneously for all eight fruits over all N_Z observations using nlsur (non-linear least squares) as in equation (8). Each equation in the LinQuad system includes terms to correct for the probability of non-zero values per method of Shonkwiler and Yen (1999). We allow correlation in the ε across equations and use a seemingly unrelated regression (SUR) approach.[xxxii] We report standard errors without adjusting for the use of predicted parameters for the cdf and pdf in the system estimation.
7. Fabiosa and Jensen (2003) present formulas for the elasticities from the LinQuad system, including consideration of the selection for non-zero values and those are applied here. In our notation, for example, the equation below is equivalent to their equation 13 where the final selection term is written differently. For price elasticities:

$$UPED_{ij} = \left[\hat{\omega}_{ij} \phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i \right)\left(\bar{\mathbf{X}}\hat{\boldsymbol{\beta}}_i + \bar{\mathbf{c}}\hat{\boldsymbol{\sigma}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\theta}}_i + \hat{\gamma}_i \bar{\hat{\lambda}} \right) + \right.$$
$$\left. \Phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i \right)\left(\hat{\xi}_{ij} - \hat{\gamma}_i \hat{\omega}_{ij} \left(\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i \right) \bar{\hat{\lambda}}_i + \left(\bar{\hat{\lambda}}_i \right)^2 \right) \right) \frac{\bar{P}_j}{\bar{o}_i} \right] \quad \text{(AA)}$$

where $\xi_{ij} = \theta_{ij} - \boldsymbol{\Psi}_i[\boldsymbol{\delta}_i + \bar{\mathbf{X}}_{km}\boldsymbol{\beta}_i + \mathbf{c}_{km}\boldsymbol{\sigma}_i + \bar{\mathbf{P}}_{km}\boldsymbol{\theta}_i]$.

Our income elasticity is equivalent to their equation (14),

$$UIED_{i} = \left[\hat{\omega}_{is} \phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i \right)\left(\bar{\mathbf{X}}\hat{\boldsymbol{\beta}}_i + \bar{\mathbf{c}}\hat{\boldsymbol{\sigma}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\theta}}_i + \hat{\gamma}_i \bar{\hat{\lambda}} \right) + \right.$$
$$\left. \Phi\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i \right)\left(\Psi_i - \hat{\gamma}_i \hat{\omega}_{is} \left(\left(\bar{\mathbf{X}}\hat{\boldsymbol{\alpha}}_i + \bar{\mathbf{C}}\hat{\boldsymbol{\mu}}_i + \bar{\mathbf{P}}\hat{\boldsymbol{\omega}}_i \right) \bar{\hat{\lambda}}_i + \left(\bar{\hat{\lambda}}_i \right)^2 \right) \right) \frac{\bar{S}}{\bar{o}_i} \right] \quad \text{(AB)}$$

The conditional elasticities were calculated using the formula for the elasticity for the selected positive sample and using means for covariates that correspond to the selected sample for each fruit separately. For example, we computed:

$$CPED_{ij} = \frac{\partial E\left[o_{ikm} \mid Y_{ikm} = 1 \right]_i}{\partial p_j} \frac{p_j}{E\left[o_{ikm} \mid Y_{ikm} = 1 \right]} \quad \text{(AC)}$$

with the covariates set to the means for the conditional sample. To be clear, the parameters underlying these conditional elasticities were not measured using only the sample of positive consumers for each fruit—the entire sample was used to estimate the parameters of the system.

8. Our robust standard errors are clustered on the individual and thus allow for a particular heteroskedasticity. As noted in the text, we do not adjust the standard errors for the use of the estimated selection correction parameters when estimating the demand system.

5. Details on estimation method 3

5.1. Bootstrap procedure

We bootstrap a vector of estimated elasticities, the calculation of which is described in the enumerated process below. Each bootstrap is conducted with the *R* function boot, using the full sample to calculate the vector of estimates and an additional 100 bootstrap replicates (random sample of size *N*, with replacement) to calculate standard errors. A separate bootstrap procedure is run for each of the fruit types (apple, blueberry, orange, strawberry) and for each income level (poor, middle, rich).

5.2. Calculating the bootstrapped vector of elasticities

1. Data preparation
    a. We drop all households that never bought fruit (organic or conventional) during a calendar year.
    b. We restrict attention to the primary Nielsen markets (the market variable will enter as a factor variable in estimation of the purchase probabilities).
    c. We generate season-by-year categories to include as a factor variable in estimation.
    d. We convert income to real monthly income by using the midpoint of the annual income categories, dividing by 12, and inflating using the US Urban Consumer Price Index.
    e. We calculate “other” price and expenditure categories by combining all fruit other than apples, blueberries, orange, and strawberries.
    f. We calculate purchase quantities by dividing expenditures by prices.
    g. We drop observations in the top 1% of prices.
    h. We use an 80% random sample for estimating (training) the lasso models.
2. Purchase probabilities
    a. We use the *R* function cv.glmnet function with the binomial distribution family option (logistic link function) and 10-fold cross-validation to predict whether a household purchases any fruit (target variable is an indicator variable of whether fruit was purchased). Predictors (controls) include prices, income, and indicators for market, household size, residential type, household composition, number of children, household head age categories, education categories, occupation categories, marital status, race, ethnicity, Rural-Urban Continuum Code, and seasons-by-year.
    b. We estimate purchase probability by predicting each household’s purchase probability using the *R* function predict and the returned lasso object, and then calculating the mean across households.
3. Unconditional expectations
    a. We use the *R* function cv.glmnet with the gaussian distribution family option (identity link function) and 10-fold cross-validation to predict the quantity of fruit a household purchases (target variable is the quantity of fruit purchased). Predictors (controls) are the same as when estimating purchase probabilities, except market indicator variables are removed and each household’s predicted purchase probability (as described above) is added as a propensity score.
    b. We estimate unconditional expected purchase quantity by predicting each household’s purchase quantity using the *R* function predict and the returned lasso object, and then calculating the mean across households.
4. Conditional expectations

   a. For the fruit type whose purchase probability was estimated as described above, we drop all household-month observations where ounces purchased are 0.
   b. We use the *R* function cv.glmnet function with the gaussian distribution family option (identity function) and 10-fold cross-validation to predict the quantity of fruit a household purchases, conditional on any purchase (target variable is the quantity of fruit purchased). Predictors (controls) are the same as when estimating purchase probabilities, except market indicator variables are removed and each household's predicted purchase probability (as described above) is added as a propensity score.
   c. We estimate conditional expected purchase quantity by predicting each household's purchase quantity using the *R* function predict and the returned lasso object, and then calculating the mean across households.
5. Elasticities
   a. Price and income elasticities of demand, and subsidy and tax elasticities are calculated using a common simulation procedure. In each case we simulate the effects of a 10% change, which is represented as a 10% increase in prices or income, or as a 10% decrease in prices in the subsidy case. Price and income elasticities consist of a 10% increase in a single price (or income), the conventional tax is represented by a 10% price increase on all conventional fruit categories, and the organic subsidy is represented by a 10% price decrease on all organic fruit categories. Simulated predictions are conducted by replacing the observed price with the simulated price for each household and predicting purchase probabilities and quantities, all else held equal.
   b. Each elasticity is calculated by dividing the percent change between the mean simulated prediction and the mean baseline prediction, divided by 0.1 (or -0.1 in the subsidy case), for each fruit type.

6. Dropped Observations

The raw dataset size is 2,119,716 household-months over three years (176,643 households over 3 years). This dataset was used to create Tables 1, 2, and 5 as well as Figures 1, 2, 3, and A. For the econometric and ML analysis we winnowed the dataset down a bit. This smaller dataset is the basis of Tables 4 and 6 – 10. First we dropped any observation that had an extremely high price for organic or conventional apples, blueberries, oranges, or strawberries. Specifically, any observation that had a real price in the 99.9 percentile for organic or conventional apples, blueberries, oranges, or strawberries was dropped. Appendix Table G indicates these cutoff prices and the number of observations with prices above the cutoff prices. The total number of observations dropped due to this winnowing is 20,293.

Next we dropped any remaining household-month for a calendar year if the household did not purchase any fruit, organic or conventional, that calendar year. This resulted in an additional 173,324 household-months observations being dropped. Therefore, overall we dropped 193,617 household-months from our dataset before conducting econometric and ML analysis.

7. Code and Data

The Stata files used to estimate the individual fruit Heckman models of consumption are found in the zip file ‘estmethodone.’ The State files used to estimate the incomplete demand system of consumption are found in the zip file ‘estmethodtwo.’ The R files to run the LASSO models are found in the zip file ‘estmethodthree.’

The State code files needed to recreate Figures 1-3 and Appendix Figure 1 are in the zip file ‘figs.’

The State code files needed to recreate Table 4 are in the zip file ‘tablefour.’

All of the data referred to in the State and R code files are in the data repository. These files include:

AllExp.dta
AllExp2011.txt
AllExp2012.txt
AllExp2013.txt

Appendix Tables

Table A. Ounces per fruit item

Fruit	Ounces	Fruit	Ounces
apples	6.420	melons	40.212
apricots	1.235	mixed	7.018
avocados	7.760	nectarine	5.009
bananas	4.162	oranges	4.938
blackberries	1.227	papayas	16.861
blueberries	27.676	Passion fruit	1.499
cantaloupes	19.471	Peaches	5.291
cherries	56.842	Pears	6.279
citrine	3.104	persimmon	5.926
coconuts	14.004	pineapples	31.923
dragonfruit	13.228	plantains	6.314
figs	1.764	plums	2.328
goldenberry	24.594	pomegranates	9.947
grapefruits	6.226	pummelos	21.482
grapes	24.594	quince	6.279
honeydew	35.274	raspberry	1.227
kiwis	2.434	starfruit	3.210
lemons	2.504	strawberries	10.785
limes	2.363	tangelos	3.104
lychee	0.705	tangerines	3.104
mandarin	3.104	ugli fruit	3.104
mangos	11.852	watermelons	159.368

Table B: Variables in $\mathbf{X}_{km}$ vector

Variable	Description
incomerm	Monthly real household income (Dec., 2013 $)
Child	= 1 if there is one or more children residing in the household and equals 0 otherwise
Hhsize	The number of people in the household
college	= 1 if one or more heads of household have a bachelor's degree or higher and equals 0 otherwise
Married	= 1 if the heads of household are married and equals 0 otherwise
Black	= 1 if the head of household is black and equals 0 otherwise
Asian	= 1 if the head of household is Asian and equals 0 otherwise
otherrace	= 1 if the head of household is some other non-white race and equals 0 otherwise

Table C: Variables in C_{km} vector

Variable	Description
Metro	=1 if the rural-urban continuum category (RUCC) for county of residence is 3 or less and equals 0 otherwise RUCC = 1 if county is in metro areas of 1 million population or more. RUCC = 2 if county is in metro areas of 250,000 to 1 million population. RUCC = 3 if country is in metro areas of fewer than 250,000 population. See https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/ for more details on RUCC.
winter2011; spring2011; summer2011; fall2011; winter2012; spring2012; summer2012; fall2012; winter2013; spring2013; summer2013	Dummy variables for each season x year interaction. We assumed that March, April, and May are the spring months; June, July, and August are the summer months; September, October, November and December are the fall months; and January and February are the winter months
mkt2 - mkt76	Dummy variables for each Nielsen Scantrack market.

Table D: Variables in c_{km} vector

Variable	Description
winter2011; spring2011; summer2011; fall2011; winter2012; spring2012; summer2012; fall2012; winter2013; spring2013; summer2013	Dummy variables for each season x year interaction

Table E: Mean household values for each income class (estimated across all N_Z observations)

Variable	Income class Poor	Middle	Rich
Monthly Income	1,254	3,839	9,206
At least one child in the household (fraction)	0.283	0.237	0.184
Male or female head of household has college degree (fraction)	0.316	0.465	0.741
Live in a metropolitan area (fraction)	0.763	0.820	0.904
Household heads are married (fraction)	0.431	0.649	0.734
Household identifies as black (fraction)	0.105	0.088	0.099
Household identifies as Asian (fraction)	0.017	0.023	0.047
Household identifies as other (fraction)	0.057	0.042	0.037
Household Size	2.525	2.454	2.313
Average price of organic apples ($ per ounce)	0.094	0.094	0.095
Average price of conventional apples ($ per ounce)	0.075	0.075	0.076
Average price of organic blueberries ($ per ounce)	0.491	0.492	0.499
Average price of conventional blueberries ($ per ounce)	0.314	0.315	0.321
Average price of organic oranges ($ per ounce)	0.083	0.084	0.085
Average price of conventional oranges ($ per ounce)	0.064	0.064	0.064
Average price of organic strawberries ($ per ounce)	0.265	0.266	0.267
Average price of conventional strawberries ($ per ounce)	0.149	0.149	0.151
Average price of other organic fruits ($ per ounce)	0.332	0.333	0.335
Average price of other conventional fruits ($ per ounce)	0.153	0.153	0.155

All monetary values are in Dec. 2013 $

Table F: Variables in $\mathbf{X}'_{km}$ vector

Variable	Description	
HH income	Monthly real household income (Dec., 2013 $)	
Children	1	Under 6 only
	2	6-12 only
	3	13-17 only
	4	Under 6 & 6-12
	5	Under 6 & 13-17
	6	6-12 & 13-17
	7	Under 6 & 6-12 & 13-17
	9	No Children Under 18
HH size	1	1
	2	2
	3	3
	4	4
	5	5
	6	6
	7	7
	8	8
	9	9
Residential type	1	One Family House
	2	One Family House (Condo/Coop)
	3	Two Family
	4	Two Family House (Condo/Coop)
	5	Three+ Family House
	6	Three+ Family House (Condo/Coop)
	7	Mobile Home or Trailer
HH composition	1	Married
	2	Female Head Living with Others Related
	3	Male Head Living with Others Related
	5	Female Living Alone
	6	Female Head Living with Non-Related
	7	Male Living Alone
	8	Male Head Living with Non-Related

Variable	Description	
H of H female age	0	No female Head
	1	<25
	2	25-29
	3	30-34
	4	35-39
	5	40-44
	6	45-49
	7	50-54
	8	55-64
	9	65+
H of H male age	0	No Male Head
	1	<25
	2	25-29
	3	30-34
	4	35-39
	5	40-44
	6	45-49
	7	50-54
	8	55-64
	9	65+
H of H male hours worked	0	No Male Head
	1	< 30 hours
	2	30-34 hours
	3	35+ hours
	9	Not employed for pay
H of H female hours worked	0	No Female Head
	1	< 30 hours
	2	30-34 hours
	3	35+ hours
	9	Not employed for pay
H of H male education	0	No Male Head
	1	Grade school
	2	Some high school
	3	Graduated high school
	4	Some college
	5	Graduated college
	6	Post college grad

Variable	Description	
H of H female education	0	No Male Head
	1	Grade school
	2	Some high school
	3	Graduated high school
	4	Some college
	5	Graduated college
	6	Post college grad
H of H male occ.	0	No male head of household
	1	Professional
	2	Office work
	3	Services
	4	Sales
	5	Skilled trade
	6	Factory / Delivery / Driver
	7	Member of armed forces
	8	Personal Services
	9	Agriculture
	10	Student employed less than 30 hours per week
	11	Construction / fishermen
	12	Retired / unable to work / unemployed
H of H female occ.	0	No female head of household
	1	Professional
	2	Office work
	3	Services
	4	Sales
	5	Skilled trade
	6	Factory / Delivery / Driver
	7	Member of armed forces
	8	Personal Services
	9	Agriculture
	10	Student employed less than 30 hours per week
	11	Construction / fishermen
	12	Retired / unable to work / unemployed
Marital status	1	Married
	2	Widowed
	3	Divorced/Separated
	4	Single

Variable	Description	
race	1	White
	2	Black
	3	Asian
	4	Other
hispanic	1	Yes
	2	No

Table G: Variables in $\mathbf{C}'_{km}$vector

Variable	Description	
RUCC	1	Counties in metro areas of 1 million population or more
	2	Counties in metro areas of 250,000 to 1 million population
	3	Counties in metro areas of fewer than 250,000 population
	4	Urban population of 20,000 or more, adjacent to a metro area
	5	Urban population of 20,000 or more, not adjacent to a metro area
	6	Urban population of 2,500 to 19,999, adjacent to a metro area
	7	Urban population of 2,500 to 19,999, not adjacent to a metro area
	8	Completely rural or less than 2,500 urban population, adj. to a metro area
	9	Completely rural or less than 2,500 urban population, not adj. to a metro area
winter2011; spring2011; summer2011; fall2011; winter2012; spring2012; summer2012; fall2012; winter2013; spring2013; summer2013	Dummy variables for each season x year interaction	
mkt2 - mkt76	Dummy variables for each Nielsen Scantrack market.	

Table H: Household-months dropped from dataset due to price outliers

	Cutoff price	Number of observations above cutoff price
Conventional strawberries	0.3124	3576
Conventional apples	0.3067	2123
Conventional oranges	0.1798	2072
Conventional blueberries	0.8935	2313
Organic strawberries	0.5661	2719
Organic apples	0.2325	2185
Organic oranges	0.1757	1548
Organic blueberries	1.3878	3757

Appendix Figures

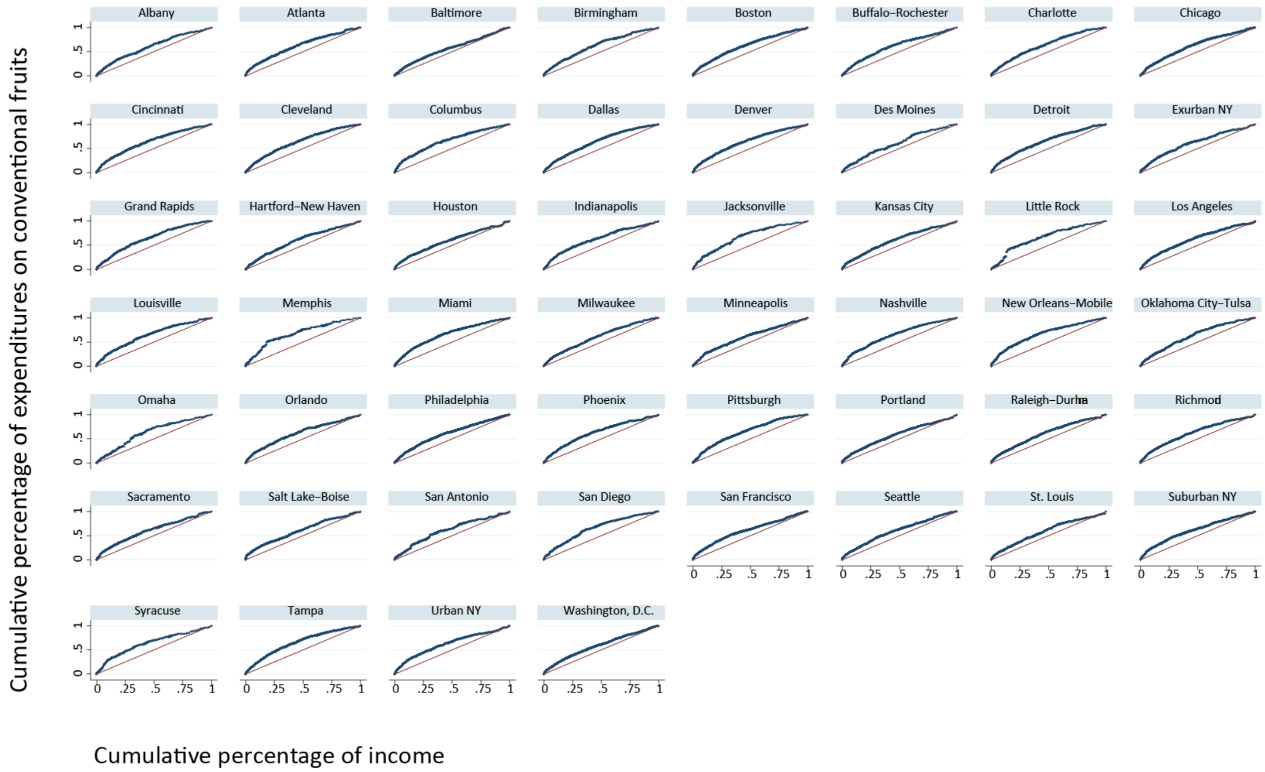

Figure A: Lorenz curves of conventional fruit expenditures by Nielsen Scantrack markets during the years 2011 through 2013 (December, 2013 dollars). The dark line in each plot is the actual cumulative expenditure curve and the lighter line is the 45 degree line.

[1] https://www.ota.com/news/press-releases/19031.
[22] Zhang et al. (2008) estimated US 2003 price and income elasticities for organic produce in general and not for individual fruit and vegetable types.
[3] Ninety-three percent of organic food sales place in conventional and natural food supermarkets and chains according to the Organic Trade Association. See https://www.ers.usda.gov/topics/natural-resources-environment/organic-agriculture/organic-market-overview.aspx
[4] Several reasons have been given for price premiums. In some cases demand for the organic products is relatively high compared to supply (e.g., Carroll et al. 2012, OECD 2003, Stevens-Garmon et al. 2007). In addition organic farming is more costly than conventional farming (Yiridoe et al., 2005; Dimitri and Greene, 2002) and these additional costs are passed on to consumers (OECD 2003). Carroll et al. (2012) also notes the diseconomies of scale in organic food processing and marketing as a reason for an organic price premium. For example, Carroll et al. (2012) found that organic strawberry farmgate and retail prices were not highly correlated. They hypothesize that the higher costs associated with marketing organic strawberries in relation to conventional strawberries is a major source of the organic premium in the strawberry market.
[5] While more educated households are more likely to buy organic produce, there is some evidence that the most highly educated (consumers with graduate or professional degrees) are less likely to buy organic than the typical household (Thompson and Kidwell 1998).
[6] Conversely, Zhang et al. (2008) claim that eastern US households are the most likely of all households to buy organic produce, all else equal. However, their produce totals includes vegetables, a category of food we do not look at. Higher levels of per capita organic produce purchases in the eastern US could be explained by the eastern US population's relative familiarity with organic systems. Among all regions of the US, the eastern US has the highest percentage of certified organic acreage relative to total farmland (USDA-NASS 2015, USDA-NASS 2016).
[7] Purchases of products without bar codes are not typically recorded. For example, restaurant meals and farmer's market purchases are not likely to be included in Panel datasets (Kilts Nielsen Center 2014). We do not use the Nielsen dataset on purchases of products without standard UPC codes. This dataset does include some fruit purchases.
[8] Often a household trip included multiple incidents of fruit purchase. For example, on a shopping trip a household would generates three separate fruit purchase incidents by buying 16 ounces of conventional apples, a bag of organic grapes, and 8 ounces of conventional kiwi.
[9] Technically, Nielsen's panel year variable denotes the Nielsen "data year" which begins on the first Sunday before the start of a new year, or if Sunday is January 1st, that Sunday, and ends on the last Saturday of the calendar year. For example, the 2004 Consumer Panel contains data for December 28, 2003 through December 25, 2004.
[10] A household is an "organic fruit-only" household in year *y* if all fruit purchases recorded by the household in the year y Consumer Panel were of the organic variety.
[11] A household is a "both varieties of fruit" household in year *y* if somefruit purchases recorded by the household in the year y Consumer Panel were of the organic variety and others were of the conventional variety.
[12] All categorization is conditional on year and household size.
[13] Examples of a reason for subsidization include: 1) "Toward a Healthy Sustainable Food System" by the APHA (https://www.apha.org/policies-and-advocacy/public-health-policy-statements/policy-database/2014/07/29/12/34/toward-a-healthy-sustainable-food-system); 2) "Europe Subsidizes Organic Farms While "Market Obsessed" US Does Little (https://www.organicconsumers.org/old_articles/ofgu/Subsidies021206.php); 3) Reisch, L, Eberle, U, and Lorek, S. 2013. Sustainable food consumption: an overview of contemporary issues and policies. Sustainability: Science, Practice, & Policy 9(2): 7-25; 4) Food Insecurity Nutrition Incentive (FINI) Grant Program (https://nifa.usda.gov/program/food-insecurity-nutrition-incentive-fini-grant-program)
[14] An alternative way to subsidize the expansion of organic food is to offer "organic food in schools, cafeterias, and so on, in what can be regarded as green public procurement (which is on the agenda of some European countries)." (p.241, Aschemann-Witzel and Zielke 2017).
[15] 2013 data was not available.

[16] Several studies use Heckman selection models to adjust for zeros in food demand. For an early example, see Heien and Wessells (1990) and we follow further development by Shonkwiler and Yen (1999).
[17] We do not weigh household-month observations with the household's year *y* projection factor when estimating demand for organic fruit with the individual Heckman models of consumption. Nor do we weigh household-month observations with the household's year *y* projection factor when estimating demand for organic fruit with the two other estimation methods discussed in sections 4.3 and 4.4.
[18] In other words, $\partial h_j/\partial p_i = \partial h_i/\partial p_j$ where *i* and *j* index fruit, *h* is a compensated demand curve for shares, and *p* is a price.
[19] A system of Tobit equations could be used but Tobits assume that the same variables and parameters determine the probability of non-zero purchase and the amount purchased for purchasers. The selection method is more general in that the parameters and variables determining non-zero purchase can be different that those that determine quantities for purchasers.
[20] The selection demand equations also could be estimated as a system. However, we will follow Shonkwiler et al. (1999) and use equation by equation treatment of selection for positive quantities. Equation by equation treatment of the positive selection, although consistent, ignores useful cross equation information about selection. Yen and Lin (2002) propose a method with correlated selection terms for a linear demand system and Yen et al. (2003) use correlated selection terms in a nonlinear Translog demand system.
[21] $\theta_{ij} = \theta_{ji}$,for all *i,j* combinations where j also indexes fruit type × variety.
[22] The latter is imposed by deflating all prices and income by a common deflator. Technically, the deflator should be an index of all prices except those for fruit. We deflated by the CPI-U and assumed that the CPI's inclusion of fruit prices would be inconsequential.
[23] Although ML models are not always unbiased estimators, their optimization procedures could lead to lower bias if the true model is complex and there is a sufficiently large number of relevant variables that the econometrician is unable to rely on in a theoretical model.
[24] The variables are transformed to each have a mean of zero and a variance of one.
[25] See https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/.
[26] Theoretically, the product of organic *j*'s purchase probability elasticity with respect to price ($PPEP_j$) and its conditional own-price elasticity of demand ($CPED_{jj}$) should equal its unconditional own-price elasticity of demand ($UPED_{jj}$). However, because estimation method 1-3' models are nonlinear and we evaluate estimated models at the means, this will not hold with our estimates. For example, consider the representative middle income household's expected demand for organic strawberries according to estimation method 2. $UPED_{jj} = PPEP_j \times CPED_{jj} = -(1.64 \times 1.90) = -3.12$. But the estimated $UPED_{jj}$ is –1.97.
[27] "US scanner data indicate that elasticities are highest for produce and lowest for processed categories (Sridhar et al. 2012)." (p. 240, Aschemann-Witzel and Zielke 2017).
[28] In estimation method 2 (see section 4.3) we estimate eight latent demand equations, one each for the four organic and four conventional versions of apples, blueberries, oranges, and strawberries as a system. Therefore, the other organic and other conventional fruit categories are not included in estimation method 2. These two fruit categories are included in estimation methods 1 and 3. Therefore, differences between method 2 estimates and methods 1 and 3 estimates may partially be explained by this structural difference in methods.
[29] The standard errors on estimation method 1 and 2's elasticities may be biased downward given that both methods treat each household-month's inverse mills ratio (or propensity score) as an observation and not the random variable that it actually is. In other words, standard errors on estimation method 1 and 2's elasticities may not be as precise as reported. The LASSO elasticity estimates do not have biased standard errors.
[30] As we noted in the previous footnote, estimation method 2 does not include other organic and other conventional fruit categories in the system of latent demand equations. These two fruit categories *are* included in estimation methods 1 and 3. Therefore, households in the policy simulations with the second method do not experience a price decrease in the "other organic fruit" category.
[31] Conventional tax purchase and expenditure elasticities are the same because the price of organics does not change under this policy, so the percentage change in quantity is the same as the percentage change in expenditure.

[xxxii] With location dummies, the number of parameters for the system as a whole is too large to do as a single maximum likelihood estimation (over 1600 parameters).