The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

A Censored Demand System Analysis of Fruits and Vegetables by Different Income Groups Using Micro Data

Luyuan Niu and Michael Wohlgenant
Department of Agricultural and Resource Economics
North Carolina State University
Box 8109 NCSU
Raleigh, NC 27695

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2012 AAEA Annual Meeting, Seattle, Washington, August 12-14, 2012

Copyright 2012 by Luyuan Niu and Michael Wohlgenant. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

A Censored Demand System Analysis of Fruits and Vegetables by Different Income Groups Using Micro Data*

Luyuan Niu and Michael Wohlgenant†

May 8, 2012

Abstract

This study examines demand for fruits and vegetables segmented by income levels in a complete demand system framework using the Consumer Expenditure Survey (CEX) from 2002 to 2006. Results show that disparities are found between high-income households and low-income households. Seasonal effects and demographic variables, such as household heads' race and gender, region, household size and household composition, play an important role in fruit and vegetable consumption for both categories of household. In contrast, urban status, household heads' educational level and age are not found to have a statistically significant impact for low-income households. Conditional price elasticities indicate that processed fruits and vegetables, fresh fruits and fresh vegetables are "gross complements" and "net substitutes". Moreover, compared to high-income households, the low-income households are more responsive to own price changes for all three categories; these households are also more responsive to expenditure changes for processed fruits and vegetables.

* We thank Xiaoyong Zheng, Atsushi Inoue, and Ivan Kandilov for helpful comments. All errors are our own.
† Niu and Wohlgenant, Department of Agricultural and Resource Economics, North Carolina State University
Email address: lniu@ncsu.edu and michael_wohlgenant@ncsu.edu

I. Introduction

On average, Americans' consumption of fruits and vegetables falls short of the recommended amount outlined in the *Dietary Guidelines for Americans*. The *2010 Dietary Guidelines for Americans* recommends average daily intakes of 2.0 and 2.5 cups of fruits and vegetables for the 2000 calorie level.[1] From 1999 to 2000, 90% of Americans failed to consume sufficient fruits and vegetables compared to the recommended amounts (Basiotis et al. 2002).

Furthermore, low-income households are found to consume even fewer fruits and vegetables than high-income households. For example, Dong and Lin (2009) show that low-income households consume only 0.96 and 1.43 cups of fruits and vegetables compared to 1.14 and 1.72 cups consumed by high-income households according to 1999-2002 National Health and Nutrition Examination Survey (NHANES).[2,3] Many food policies and food assistance programs, such as the Supplemental Nutrition Assistance Program (SNAP), are targeted at low-income households. For policies and programs such as these, quantitative information on demand for fruits and vegetables at the household level and for different segments of the population is required to inform public policy. Although there are several studies researching consumption of low-income households, some do not use an economic model in their analysis (e.g., Lin 2005 and Basiotis et al. 2002), and some use only a single-equation model rather than a demand system (e.g., Dong and Lin 2009). This study uses a complete demand system analysis and examines the question whether any disparities exist in fruit and vegetable consumption patterns between low-income and high-income households. In addition to provide the information on demographic and other socio-economic variables, this study focuses on quantifying household responses to prices and expenditures of fruits and vegetables segmented by income levels.

Compared to the existing literature on demand for fruits and vegetables, this study is different from other studies in the following aspects. First, the demand system is estimated using micro-level data rather than aggregate data that most of the previous studies use (see Tables 1 and 2). Second, due to the fact that price data for individuals are unavailable in the CEX, the Stone-

[1] See what constitutes a cup at USDA's ChooseMyPlate.gov.
[2] In the NHANES sample, average calorie intake is 2164 calories per day.
[3] The high-income household has income greater than 300 percent of the poverty line in the study of Dong and Lin (2009). In this study, the low-income (high-income) household is defined as the one which income is below or equal to (above) 185 % of the federal poverty guidelines after adjusting for the household size. This cutoff is consistent with the definitions in the National Household Food Acquisition and Purchase Survey.

Lewbel price indices are derived, allowing the prices to have sufficient variation to identify the demand function. Third, this study accounts for the reported zero observations that are often neglected in previous studies that use micro-level data (e.g., Dong and Lin 2009 and Huang and Lin 2000). Fourth, elasticities of fresh fruits, fresh vegetables and processed fruits and vegetables are derived for individual consuming units, which are different from other studies where fruits or vegetables are aggregated in all forms (e.g., Dong and Lin 2009, Huang and Lin 2000 and Park et al. 1996). Finally, a correlated random effects model is utilized and estimated by two-step Quasi-Maximum Likelihood Estimation (QMLE): the first stage addresses the problem of censoring and derives a reduced-form estimator, and the second stage derives a minimum distance estimator while imposing economic restrictions.

In this study, all "fruits and vegetables" are divided into three categories: fresh fruits, fresh vegetables, and processed fruits and vegetables. Figure 1 in chapter one shows the trends in expenditures for these three categories from 1999 to 2009. In 2009, expenditures for fresh fruits were $162 per capita, a 19% increase from 1999. Expenditures for fresh vegetables are $143 in 2009 and $146 in 1999, and have remained stable during the entire period. Processed fruits and vegetables show a decline in expenditures during the period of 1999-2009. Per capita expenditures were $153 in 2009, 24% less than those in 1999. Per capita real dollar expenditures are calculated as average annual expenditures of all households divided by the Laspeyers Consumer Price Index (CPI) (Dec 1997=100).

From figure 1, we also see that Americans have changed their consumption mix of fruits and vegetables. People eat relatively fewer processed fruits and vegetables and more fresh fruits. Changes in consumption may be due to the changes in relative prices. Figure 2 illustrates the relative price changes for fruits and vegetables during the year 1999-2009. Although relative prices of all three categories increased steadily from 1999 to 2008, relative prices of fresh fruits and vegetables were always higher than those of processed fruits and vegetables. This explains why people spent more on processed products than fresh products during the same period. Relative prices of fresh fruits and vegetables began to decline in 2008. It is worth mentioning that, at some time between 2008 and 2009, relative prices of processed fruits and vegetables exceed the relative prices of fresh fruits. The changes may be explained from the supply side. Improved packing and shipping technology allows fruits and vegetables to maintain higher quality when shipped over long distances, and storage facilities increase the availability of fresh produce year round with high quality. Increasing fresh selections and quality in grocery stores improves households' choices, which may decrease their need for processed fruits and vegetables.

Figure 3 gives a general picture of fruit and vegetable consumption by different income groups. Average annual expenditures for fresh fruits in the "$70000 and more" income group are 139 % above the "less than $5000" group, 158 % more for fresh vegetables and 70% more for processed fruits and vegetables, respectively. Based on the CEX data during 2002-2006 periods, we also found that the average weekly expenditures for processed fruits and vegetables, fresh fruits and fresh vegetables are $4.61, $4.17, and $4.02, respectively, for low-income households, while they are $5.45, $4.99 and $5.04 for high-income households, respectively.

Results show that statistically significant differences are found between high-income households and low-income households. Seasonality is very important for both income groups, especially for consumption of processed fruits and vegetables and fresh fruits. Demographic variables, such as region, race, and household composition, play an important role in the fruit and vegetable consumption for both income groups. Education and household heads' age have no statistically significant impact on the consumption of fruits and vegetables for low-income households. Conditional price elasticities indicate that three goods categories are "gross complements" and "net substitutes". Moreover, compared to high-income households, low-income households are more responsive to own price changes of all three categories of fruits and vegetables; these households are also more responsive to income changes when consuming processed fruits and vegetables.

The remainder of the paper is organized as follows. The next section discusses CEX data and gives variable construction and descriptive statistics. In the third section, demand model and econometric methodology are given. The fourth section presents the results and makes comparisons between both income groups of households. The final section summarizes and concludes the first part of the paper.

Figure 1 Per Capita Real Dollar Expenditures for Fruits and Vegetables, U.S., 1999-2009

Source: Consumer Expenditure Survey, Bureau Labor Statistics

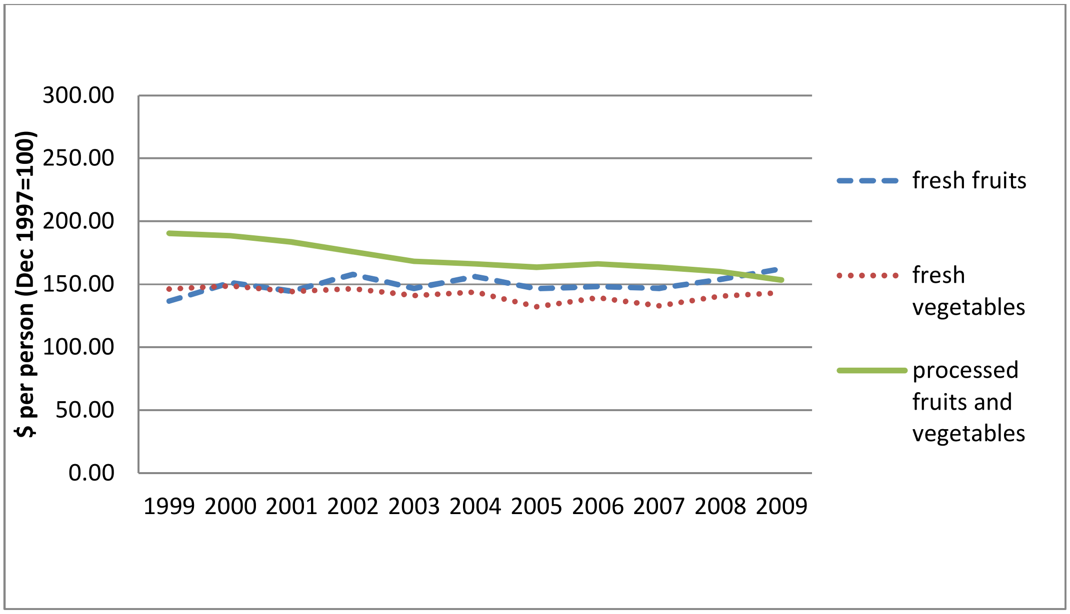


Figure 2 Relative Prices for Fruits and Vegetables, U.S., 1999-2009

Source: Consumer Expenditure Survey, Bureau Labor Statistics

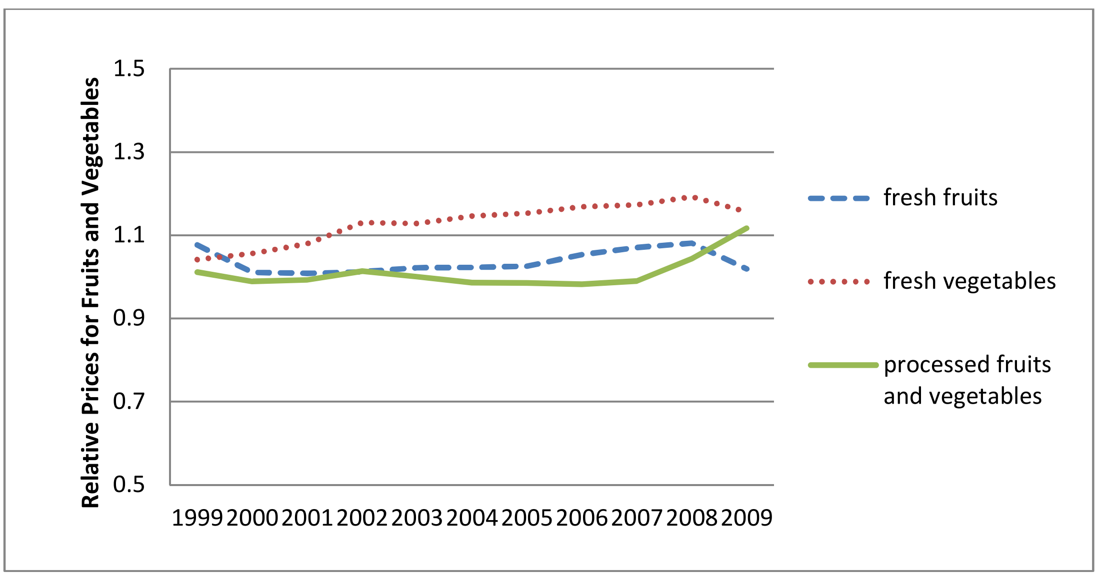
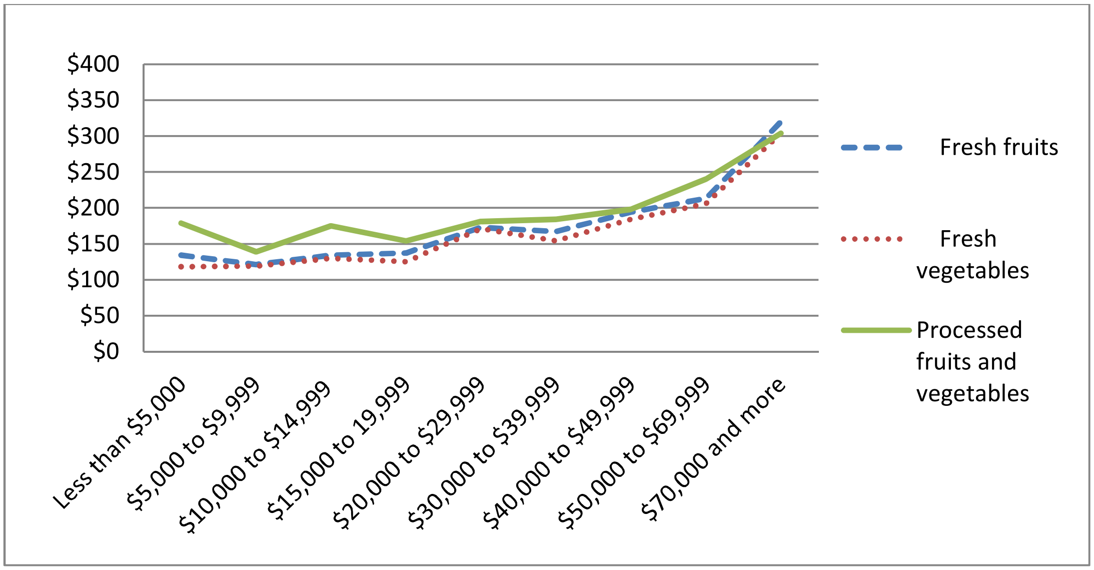


Figure 3 Average Annual Expenditures on Fruits and Vegetables, U.S., 2009

Source: Consumer Expenditure Survey, Bureau Labor Statistics



II. Data[4]

This study uses 2002-2006 CEX Diary Survey. The CEX is conducted by the Census Bureau for the Bureau of Labor Statistics (BLS) in the U.S. Department of Labor and is considered as one of the most comprehensive datasets in the United States. The CEX data is widely used in economic research and analysis, and is also used as the database to support and maintain the CPI.

The CEX is designed as a national probability sample of households representing total civilian noninstitutional population of the United States and portion of the institutional population. From 1980, CEX is issued annually to contain more sufficient and correct data. It contains two surveys: the Interview Survey (IS) and the Diary Survey (DS). IS conducts the interview once every three months over five consecutive quarters to obtain an entire year's data, while DS keeps records of small and frequently purchased items, such as food, collected over two consecutive one-week periods. In this study, DS data is used from 2002 to 2006.

[4] Data descriptions refer to "U.S. Department of Labor, Bureau of Labor Statistics, Consumer Expenditure Survey, Diary Survey, 2006".

The Dairy Survey consists of four data files, where FMLY and MEMB files include demographic and socioeconomic information of all the Consumer units (CUs).[5] Considering the current literature, the following variables are included in the model: region, urban/ rural, household size, number of children under 18 and number of persons over 64 in each household, and the reference person's education level, race, sex and age.[6]

Beginning from 2004, imputed income data have been implemented in the CEX. Five imputed income values and their mean values are reported. The imputation of income data provides estimates for unreported or invalid income values and improves the dataset's utilization. Although both not imputed and imputed data are available in 2006, the imputed income values are chosen in order to be consistent with data in 2004 and 2005.

Incomplete and topcoded data are deleted, and the observations of zero or negative income values and of households not purchasing any fruits and vegetables during the survey period are deleted. [7,8] Since the sample selection rule is exogenous to the model, no sample selection problem is involved. Finally, 33,660 observations remain.

In light of the fact that the goal of this study is to examine the differences in demand for fruits and vegetables between different income groups, data are divided into two groups: the low-income and high-income groups. The low (high) income households refer to the households whose annual income is below (above) 185% of the federal poverty guidelines.[9] Federal poverty guidelines are issued each year by the Department of Health and Human Services (denoted as HHS) and vary by household size. Federal poverty guidelines are a simplified version of the poverty thresholds (updated by the Census Bureau) and are mainly used for administrative purposes in order to determine financial eligibility for certain programs, such as the SNAP. According to the data, there are 4,722 households from the low-income group and 12,108

[5] In the remainder of the study, "CUs" and "households" are used interchangeably.

[6] Reference person is the first member mentioned by the respondent when asked to "Start with the name of the person or one of the persons who owns or rents the home." In most cases, it refers to the household head, so in the remainder of the study, "household head" is used. Other CU members are determined with respect to this person.

[7] Topcoding refers to the data replacement when the value of the original data exceeds prescribed critical values. For income variable, about 1/8 data are topcoded.

[8] Negative income value can occur for people who are self-employed or own a farm. Zeros can occur when respondents don't provide any income data.

[9] This study uses before-tax income. The use of income with federal poverty guidelines depends on the research purpose.

households from the high-income group. Table 3 reports the descriptive statistics for both groups of households.

As written above, CEX is a micro-level dataset, so excessive zero observations are present in expenditure shares. Table 4 shows the proportion of zeros in each group in each time period. There are over 20% zero budget shares for each commodity, suggesting they are censored. This issue is again discussed in the model section.

Because no price data are provided by the CEX, they are constructed based on the CPI. The CPI is defined by the BLS as "a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services".[10] A quarterly CPI series is calculated in order to be consistent with the quarterly based CEX data[11]. All the indices in use are changed to December 1997=100 base.[12] Due to the insufficient variation for the CPI, Stone-Lewbel (S-L) price indices are created for individual households following the approach proposed by Lewbel (1989). The construction method of the S-L price indices can be found in Appendix B.[13]

III. Model and Methodology

Since our CEX data is a short panel, there are two models that can be considered to solve the individual heterogeneity problem. One is the fixed effects model and the other is the random effects model. In both models, unobserved individual effect c_j (j denotes individual household in our case) is treated as a random variable.[14] The difference is that the fixed effect model allows c_j to have an arbitrary correlation with other observed regressors and no distribution is assumed for

[10] "Consumer Price Index - Frequently Asked Questions (FAQs)". *Bureau of Labor Statistics*. Accessed September 10, 2010. http://www.bls.gov/cpi/cpifaq.htm.

[11] The CPI of processed fruits and vegetables does not include fruit juice, while fruit juice is included in the category of processed fruits and vegetables in this study. So the CPI is only an approximate aggregate price of the processed fruit and vegetable expenditure series studied here. More details on how to construct the categories of fruits and vegetables can be found in Table A1.

[12] Due to data availability, the CPI in use is that for all urban consumers.

[13] Hoderlein and Mihaleva (2008) compare the results of using the usual aggregate price indices and the Stone-Lewbel price indices in the food demand estimation, and conclude that the S-L price indices greatly increase the precision of the estimates in both parametric and nonparametric modeling.

[14] Traditionally, c_j is treated as a parameter to be estimated, however, Wooldridge (2002) argues that it makes more sense to treat it as random draws from the population along with the other variables.

it. In contrast, random effects model assumes c_j a conditional distribution and orthogonal to other regressors.

Consider our data structure. Each household has two-period observations: One is recorded in the first week and the other in the second week. For most households, the observed demographic variables included in the model are constant during this short two-week survey period. If we use the fixed effect approach, time-constant demographic variables would not play a role in the estimation because they would be dropped out of the estimable equation after the fixed effects transformation, or, in other words, the time-demeaned explanatory variables would contain zero columns, which fails rank condition requirements. By intuition, if the time-constant explanatory variables not included in the model are correlated with the unobserved c_j, it would be difficult to distinguish these two effects on the dependent variable, which would lead to inconsistent estimates of the coefficients of the observed time-constant explanatory variables (Wooldridge 2002).

We use the random effects approach. However, the zero correlation assumption between c_j and observed regressors is not appropriate in our context because the unobservable household effect is likely to be correlated with observed demographic variables, total expenditure and constructed prices.[15] Hence, a correlated random-effect method (Jakubson 1988) is adopted in the study.

Following Meyerhoefer, Ranney and Sahn (2005), assume a cost function takes the PIGLOG form:

$$\log \breve{c}(p,u,\varepsilon;d,c) = \alpha_0 + \sum_k \alpha_k \log p_{kjt} + \sum_k \mu_k \log p_{kjt} T_q + \sum_k \sum_l \lambda_{kl} \log p_{kjt} d_{lj} + \frac{1}{2}\sum_k \sum_i \tilde{\gamma}_{ki} \log p_{kjt} \log p_{ijt} + u^*_{jt}\beta_0 \prod_k p_{kjt}^{\beta_k} + \sum_k \psi_k \log p_{kjt} c_j + \sum_k \log p_{kjt} \varepsilon_{kjt},$$

where $\log \breve{c}(\cdot)$ represents the cost function, $\log p_{kjt}$ is the price of good k (=1,…,N) in the survey week t for household j (=1,…,J), T_q is a vector of dummy variables for quarter q, d_{lj} denotes the lth (l=1,…,L) demographic variables for household j, u^*_{jt} is household j's utility level, c_j is

[15] Recall that prices faced by individual households are a function of the budget shares of goods in the subgroup consumption. Details can be found in Appendix B.

unobserved household specific effects, ε_{kjt} represents some components deterministic for the households j but unobservable to the researchers and treated by the researchers as a random variable. Assume a vector of ε_{kjt} has a multivariate normal distribution with mean zero and covariance matrix $\sum_{\varepsilon}$, and $E(\varepsilon_{kjt} | \text{all independent variables}) = 0$.

The PIGLOG cost function is general enough to act as a second-order approximation to any arbitrary cost or indirect utility function. Time dummies, individual specific effects and stochastic error terms are incorporated into the demand model in the same way as demographic variables. The procedure is called "demographic translating" (Pollack and Wales 1981), which is very general in the sense that it does not require the functional form of the original demand system but can be used in combining with any complete demand system while maintaining its plausibility.

By applying logarithm version of Shephard's lemma, the Almost Ideal Demand System (AIDS) (Deaton and Muellbauer 1980) is obtained as

$$w^*_{njt} = \alpha_n + \mu_n T_q + \sum_l \lambda_{nl} d_{jl} + \sum_k \gamma_{nk} \log p_{kjt} + \beta_n (\log y_{jt} - \log P_{jt}) + \psi_n c_j + u_{njt},$$

where

$$\log P_{jt} = \alpha_0 + \sum_k \alpha_k \log p_{kjt} + \sum_k \mu_k \log p_{kjt} T_q + \sum_k \sum_l \lambda_{kl} \log p_{kjt} d_{lj} + \frac{1}{2} \sum_k \sum_i \gamma_{ki} \log p_{kjt} \log p_{ijt} + \sum_k \psi_k \log p_{kjt} c_j,$$

$u_{njt} = \varepsilon_{njt} - \beta_n \sum_k \log p_{kjt} \varepsilon_{kjt}$, $\gamma_{ki} = 1/2\,(\tilde{\gamma}_{ki} + \tilde{\gamma}_{ik})$, w^*_{njt} is the expenditure share of good n at time t for household j, and $\log y_{jt}$ represents total expenditure for household j at time t. For simplicity, set $\psi_k = 1$.[16] We know that, when incorporating demand shifters in the intercepts, the AIDS model is not invariant to units of measurement (Alston, Chalfant, and Piggott 2001). One way to solve the problem is to use a "corrected" Stone price index, $\log P^S_{jt} = \sum_k w^o_k \log p_{kjt}$ where w^o_k is the mean share for good k across all the households and all the times, to replace $\log P_{jt}$ in the AIDS model (Moschini 1995). In addition, the new price index can also avoid the potential multicollinearity problem while reducing the burden of estimating the original model.

[16] In intuition, since c is unobservable and almost has no measurement unit, it would not make sense to estimate its partial effect (Wooldridge, 2002).

Homogeneity and symmetry restrictions are imposed on the demand equation, which implies

$$\sum_k \gamma_{ik} = 0 \quad \text{and} \quad \gamma_{ki} = \gamma_{ik}\,.^{17}$$

The unobservable household specific effect c_j is expected to be correlated with individual observable demographics, prices and total expenditure, so a correlated random effect approach is applied by modeling c_j as a linear projection on all other independent variables across all time periods. That is,

$$c_j = \sum_l \eta_l d_{lj} + \sum_i \sum_t \theta_{it} \log p_{ijt} + \sum_t \delta_t (\log y_{jt} - \log P_{jt}) + v_j$$

where v_j is an error term with normal distribution of mean zero and variance σ_v^2 and it is assumed to be uncorrelated with u_{njt}. By the definition of linear projection, v_j is also uncorrelated with other regressors in the expression of c_j.

By substituting c_j into the demand function, one can get

$$\begin{aligned} w^*_{njt} &= \alpha_n + \mu_n T_q + \sum_l (\lambda_{kl} + \eta_l) d_{lj} + \sum_i (\gamma_{ni} \log p_{ijt} + \sum_t \theta_{it} \log p_{ijt}) \\ &+ \beta_n (\log y_{jt} - \log P_{jt}) + \sum_t \delta_t (\log y_{jt} - \log P_{jt}) + \tilde{u}_{njt}, \end{aligned}$$

where $\tilde{u}_{njt} = v_j + u_{njt} = v_j + \varepsilon_{njt} - \beta_n \sum_k \log p_{kjt} \varepsilon_{kjt}$. From this expression of the new error term, $\tilde{u}_{njt}$ has a normal distribution with a zero mean and a heteroscedastic variance.

As written in the data section, there are over 20% of zero observations for expenditures for each good, which is a relatively large amount that could not be simply neglected. Moreover, expenditures are censored from below at point zero due to the fact that zeros or positive amounts are always observed. Thus, the Tobit model is chosen to account for the zeros.[18] The Tobit model is specified as

$$w_{njt} = \max(0,\ w^*_{njt}),$$

[17] The adding-up restrictions are not imposed for simplicity.
[18] A two-limit Tobit model is also estimated in this study. See the setup and results in the Appendix C.

where w^*_{njt} is a latent variable. This means, when w^*_{njt} is larger than zero, the observed w_{njt} equals to w^*_{njt}; when w^*_{njt} is less than or equal to zero, the observed w_{njt} equals to zero.

In the case where censoring is not a problem, the demand system can be estimated with jointly Maximum Likelihood Estimation (MLE). Under the case of censoring, this likelihood function needs to be modified to account for the censored expenditures. In practice, it is difficult to manipulate because the likelihood function of the censored demand requires evaluation of multiple integrals. Although the data used in this study is a short panel, there are in total six (N by T) demand equations that may involve larger than three dimensional integrals, which make the estimation infeasible.

A methodology is adopted by using a QMLE to avoid evaluating high dimensional integrals (Jakubson 1998; Meyerhoefer, Ranney and Sahn 2005). By manipulating the marginal distributions of each univariate Tobit model, joint ML is approximated using the method of moment techniques. Specifically, the QMLE is derived in two stages. In the first stage, ML is applied to the Tobit model equation by equation in order to derive the reduced-form parameter estimates for each equation in each time period. In the second stage, through setting up both sample and population moment conditions, a minimum distance estimator is used to derive consistent structural parameter estimates while imposing the cross-equation economic restrictions.

According to the structure of the error term in the demand equation, the variance $\tilde{u}_{njt}$ is heteroscedastic, which would lead to inconsistent Tobit estimator (Pudney 1989). So the model is modified by specifying the form of the variance in the following way.

$$\rho_{njt} \equiv E(\tilde{u}^2_{njt}) = \sigma^2_n \exp(s_{njt}\xi_n)\,.$$

where s_{njt} is the vector of variables that are expected to be the source of heteroscedasticity and is assumed to vary by good, time and household, while the coefficients σ_n and ξ_n are assumed to vary by good and are estimated by MLE along with the parameters in the share function.

By comparing the magnitude of Akaike Information Criterion (AIC) values of each demand equation, the following nine variables are selected to be included in the s_{njt}: the second and third quarter, high school degree or below, the Midwest and South Region, number of persons who are over 64 in a household, own-price $\log p_{njt}$ and total expenditure $\log y_{jt}$.

The parameter set stacked over time for good n has the form of

$$\kappa_n = [\ell\alpha_n \,|\, \ell\mu' \,|\, \ell(\lambda' + \eta') \,|\, \ell\theta_1' + \gamma_{n1}I \;\cdots\; \ell\theta_3' + \gamma_{n3}I \,|\, \ell\delta' + \beta_n I \,|\, \ell\sigma_n^2 \;\; |\, \ell\xi_n']$$

Where $\mu' = (\mu_{n1} \cdots \mu_{n3})$, $\lambda' = (\lambda_{n1} \cdots \lambda_{n15})$, $\eta' = (\eta_{n1} \cdots \eta_{n15})$, $\theta_i' = (\mu_{i1} \cdots \mu_{i3})$ for i=1,2,and 3, $\delta' = (\delta_1 \;\delta_2)$, $\xi_n' = (\xi_{n1} \cdots \xi_{n9})$, ℓ is the vector of ones, and I is the identity matrix. The parameter set above includes the parameters in the share function and the ones composing the variance of the heteroscedastic error term. Thus, the reduced-form parameters for all goods are denoted as $\kappa = [\kappa_1 \;\kappa_2 \;\kappa_3]$. For the following calculation, κ needs to be transformed into a vector, denoted as $K = \text{vec}(\kappa')$.

Structural parameters, called ϕ, are derived by minimizing the following objective function.

$$\min_{\phi}(\hat{K} - a(\phi))'\hat{W}(\hat{K} - a(\phi)),$$

Where W is the weighting matrix to measure the distance between the sample moments and the corresponding population moments, where the former one involves the consistent estimated reduced-form parameters K, and the latter involves the structural parameters ϕ. The relationship between K and ϕ is described by a function $a(\cdot)$, where $K_0 = a(\phi_0)$, where $a(\cdot)$ is used to disentangle the coefficients and impose the restrictions required by the economic theory, and the subscript "o" means the true values of the parameters. Since the restrictions are linear, $a(\cdot)$ is also linear with the form of $A\phi_0$. The minimum estimator ϕ is efficient if $W = \Xi_0^{-1}$, where Ξ_0^{-1} is the inverse of the asymptotic covariance matrix of $\hat{K}$, that is,

$$\sqrt{J}(\hat{K} - K_0) \xrightarrow{d} N(0,\; \Xi_0)$$

which can be obtained from the univariate Tobit estimation.

Since Ξ_0 is unknown and needs to be estimated, Meyerhoefer (2002) derives $\Xi_0 = H_0^{-1} S_0 H_0^{-1}$ according to the asymptotic property of MLE, in which H_0 is a block diagonal matrix with the asymptotic variance matrix of the univariate Tobit model in the diagonal; S_0 consists of the cross products of the quasi-scores within and across each demand equation. Both H_0 and S_0 enter Ξ_0 since in the first stage Tobit model is estimated equation by equation rather than being jointly estimated.

Nevertheless, difficulties in empirical applications are reported based on efficient minimum distance estimator (e.g., Abowd and Card 1989). Most studies recommend the use of an identity matrix as the weighting matrix instead of the optimal weighting matrix (Kennedy 2003, p. 151). This study follows this recommendation. It is noteworthy that the overidentification test is valid only when one employs the optimal weighting matrix.

There are more than two hundred parameters to be estimated. To reduce the complexity of the computation and to avoid the asymptotic normality assumptions for minimum distance estimator, the estimates of $se(\hat{\phi})$, the standard error of structural parameter estimates, are calculated by bootstrapping. Specifically, 1000 new samples are randomly drawn (allowing repeated sampling) from the original data, in which each new sample has the same number of observations as the original one. The same estimation procedure is conducted on these new samples, and 1000 new sets of structural parameter estimates are derived. The standard deviation of these estimates is the standard error.

Elasticities are derived as follows. Start from the definition of the share function,

$$\log w(p,y) = \log p + \log q(p,y) - \log y,$$

where p is price and y is expenditure or income. By taking the derivatives with respect to $\log p$ and $\log y$ respectively and rearranging the terms, price and expenditure elasticity formulas can be derived for any demand equation. In the Tobit model, the income and price elasticities have the expression of

$$e_{ijt} = \frac{\partial E(w_{ijt})}{\partial \log y_{jt}} \cdot \frac{1}{E(w_{ijt})} + 1$$

and

$$e_{ik\_jt} = \frac{\partial E(w_{ijt})}{\partial \log p_{kjt}} \cdot \frac{1}{E(w_{ijt})} - \Delta_{ik\_jt},$$

where Δ_{ik_jt} is the Kronecker delta (Δ_{ik_jt} =1 for i=k; Δ_{ik_jt} =0, for i≠k)[19]. When i=k, e_{ik_jt} represents the own-price elasticities; when i≠k, it represents the cross-price elasticities.

The expected shares $E(w_{ijt})$ in the above expressions are computed as

$$E(w_{ijt}) = \Phi_{ijt}(x_{jt}K_{ijt} + \rho_{ijt}\frac{\phi_{ijt}}{\Phi_{ijt}}) \ ,$$

where $x_{jt}K_{ijt}$ is short for the demand equation for which x represents the variables and K represents the parameters; $\Phi(\cdot)$ is the cumulative density function of standard normal distribution with $\Phi_{ijt} \equiv \Phi(z_{ijt}) = \Phi(x_{jt}K_{ijt} / \rho_{ijt})$; $\phi(\cdot)$ is the probability density function of standard normal distribution with the same argument defined as above. Thus, expenditure elasticities have the form of

$$\frac{\partial E(w_{ijt})}{\partial \log y_{jt}} = \Phi_{ijt}\frac{\partial z_{ijt}}{\partial y_{jt}}(x_{jt}K_{ijt} + \rho_{ijt}\frac{\phi_{ijt}}{\Phi_{ijt}}) + \Phi_{ijt}(\beta_i + \frac{\partial \rho_{ijt}}{\partial \log y_{jt}}\frac{\phi_{ijt}}{\Phi_{ijt}} - \rho_{ijt}\frac{\phi_{ijt}}{\Phi_{ijt}}(z_{ijt} + \frac{\phi_{ijt}}{\Phi_{ijt}})\frac{\partial z_{ijt}}{\partial \log y_{jt}}),$$

where $\frac{\partial z_{ijt}}{\partial \log y_{jt}} = \frac{\beta_i}{\rho_{ijt}} - (\frac{x_{jt}K_{ijt}}{\rho_{ijt}^2})\frac{\partial \rho_{ijt}}{\partial \log y_{jt}}$, $\frac{\partial \rho_{ijt}}{\partial \log y_{jt}} = \frac{1}{2}\rho_{ijt}\xi_{i\_y}$,and ξ_{i_y} is the coefficient of $\log y_{jt}$ in the heteroscedastic variance ρ_{njt} . The own-price elasticities have very similar expressions to the expenditure elasticities, except that the coefficients β_i and ξ_{i_y} in the above formula must be changed to correspond to the prices. In contrast, the expressions for the cross-price elasticities can be simplified a great deal since the heteroscedastic variance ρ_{ijt} is constructed in the way that only own price $\log p_{ijt}$ is included. Specifically,

$$\frac{\partial E(w_{ijt})}{\partial \log p_{kjt}} = \Phi_{ijt}(\gamma_{ik} - \beta_i w_k^o) \quad \forall \ i \neq k.$$

It is noteworthy that, when computing elasticities, structural form parameter estimates are used instead of reduced-form estimates. This is because elasticities should not be affected by any correlation between the economic variables and unobserved household characteristics.

IV. Results

[19] The formula is also suggested by Meyerhoefer (2002).

Results are presented in Table 5 for both high-income households and low-income households[20]. The percentage change of parameter estimates is shown in Table 6.

First, seasonal effects are found to be very important in explaining the demand for processed fruits and vegetables and fresh fruits for both groups of households. Results suggest that both income groups of households consume fewer processed fruits and vegetables in the third quarter than in the first two quarters. They consume most in the fourth quarter. For example, high-income households and low-income households increase budget shares by 15.58% and 11.44%, respectively, in the fourth quarter compared to the third quarter.[21] In contrast, they purchase more fresh fruits in the third quarter than in the previous two quarters, and consume the fewest fresh fruits in the fourth quarter. For example, high-income households and low-income households purchase 21.25% and 13.58% more fresh fruits in the third quarter than in the fourth quarter, respectively. The reason is straightforward. As it is known, fresh food is more available during the summer time, so people tend to buy more fresh products in season and fewer processed ones. The reason that households buy more processed fruits and vegetables in the winter time is due to Thanksgiving and Christmas Days, both of which are in the fourth quarter, and during this time people may prefer fruit juice to fresh fruits, which may decrease the consumption of fresh fruits.

Second, household heads' education level and age are found to have an effect on demand of fruits and vegetables for high-income households only. The results show that, compared to the households whose heads have college degrees or above, households with heads having high school degrees demand 4.23% more processed fruits and vegetables, and those with heads without degrees demand 4.6% less fresh fruits. However, education levels have no significant effects on the demand for fruits and vegetables for low-income households. Age is also a factor influencing only high-income households. For each additional year of age of the head of household, households decrease the expenditure share by 0.12% for processed fruits and vegetables and increase 0.23% for fresh fruits.

[20] Definition of variables used in the model can be found in Table A2.
[21] The percentage change stated in the results only refers to the change in budget shares if there is no further information.

Third, household heads' gender and race also influences the consumption of fruits and vegetables. High-income households headed by males purchase 3.31% more processed fruits and vegetables and 3.43% less of fresh fruits than those headed by females. Similarly, low-income households headed by males purchase 3.42% more processed fruits and vegetable than those headed by females. However, the effects of gender are not significant for fresh fruits and fresh vegetables for low-income households. Compared to white households, black households buy more processed fruits and vegetables and fewer fresh vegetables and fresh fruits, and Asian-headed households buy fewer processed fruits and vegetables and more fresh fruit for both income groups of households.

Moreover, region is also a significant indicator of demand. Households in both income groups living in the Northeast purchase fewer fresh vegetables than those living in the West, while households in the Midwest purchase fewer fresh vegetables but more fresh fruits than ones in the West. The low-income households in the South purchase 4.95% more fresh vegetables than those in the West. This corresponds with the fact that people in the West and South may have access to more fresh fruits and vegetables. As a result, fresh products may take up the majority of their expenditure shares.

Other contributors also attract the attention. High-income households in the urban areas tend to purchase 4.46% more fresh vegetables than those in rural areas. This may be due to the fact that households in urban areas have more access to fresh vegetables than those in rural areas. However, urban status is not a significant factor for low-income households. Furthermore, it is interesting to note that, as household size increases, both groups of households purchase fewer fresh fruits. This may be because people consume more juice, such as orange and apple juice, for convenience. However, one more child (lower than 18) in a household induces greater demand for fresh fruits but less for fresh vegetables for both income groups and induces greater demand for processed fruits and vegetables (2.14%) only for high-income households. Moreover, persons over 64 in a household are not found to have a significant effect on demand for fruits and vegetables for both household groups.

Economic variables are all significant. The coefficients γ's can provide some preliminary evidence to relationships among goods. The negative signs indicate that goods are "gross complements”. Further inference needs to be made based on the estimates of the price elasticities.

Elasticities are calculated for representative households in each group. A representative household is defined as one who has the median income in each group. The results are presented in Table 7, Table 8 and Table 9.[22] Expenditure elasticities of three categories are all positive for both groups of households, as expected, meaning all three categories of goods are “normal”. The values are less than “1”, meaning they are “necessities”, and are close to “1” because only conditional elasticities are studied, meaning that households only take expenditures on fruits and vegetables rather than total income constant. It is also worth noting that the expenditure elasticities of fresh fruits and fresh vegetables are larger than those of processed fruits and vegetables. This indicates that, when the expenditure on fruits and vegetables increases, households demand more fresh fruits and vegetables than processed fruits and vegetables.

In addition, low-income households have a higher expenditure elasticity of processed fruits and vegetables than high-income households. Subsidizing fresh fruits and fresh vegetables for low-income households may be the reason for the lower response to the expenditure change.

Uncompensated own-price elasticities for all the three goods are all negative, as expected. All the cross-price elasticities are negative, meaning that all of the goods studied are “gross complements”.[23] After accounting for the income effects, compensated own-price elasticities of all three goods are negative, as expected, and all cross-price elasticities are positive, meaning all the three goods are “net substitutes”.[24] The own-price elasticities for fresh vegetables are a bit larger (2.413 for high-income household and 2.81 for low-income households in absolute values) than the values derived from the literature (see Table 1), which may attribute to the micro-level data used in this study. Moreover, all the own-price elasticities and most cross-price elasticities for low-income households are relatively large compared to high-income households, meaning

[22] As written in the model section, a two-limit Tobit model is also used and the results of elasticities can be found in the Appendix C for comparison.

[23] Huang (1993) estimates an unconditional food demand system (including both Food At Home (FAH) and Food Away From Home (FAFH)) using aggregate data. He also finds that processed fruits and vegetables are “gross complements” with fresh fruits and fresh vegetables.

[24] Conditional on the expenditure for food (including FAH and FAFH), Feng and Chern (2000) show that processed fruits and processed vegetables are both “net substitutes” for fresh vegetables. They also show that fresh fruits are “net substitutes” for processed fruits and “net complements” with processed vegetables.

that low-income households are more sensitive to price changes[25]. Based on the compensated elasticities, Slutsky matrix is derived and all the eigenvalues are negative. So the negativity condition holds at the points where elasticities are evaluated.

V. Conclusion

This study addresses the question whether there are any disparities in fruit and vegetable consumption patterns between low-income households and high-income households using household-level CEX Diary data from 2002 to 2006. In order to account for the zero observations, a censored demand system is estimated. A correlated random effect approach is utilized to solve for the individual heterogeneity and heteroscedasticity problems. Due to the infeasibility of dealing with multiple integrals in estimating demand system, a two-stage QMLE is used with the following two steps. In the first step, consistent reduced form parameter estimates are derived from a univariate Tobit model. In the second step, structural parameter estimates are derived using a minimum distance estimator after imposing economic restrictions.

Results show that there is obvious seasonality in fruit and vegetable consumption. Moreover, demographic characteristics, such as household heads' race and sex, region, household size, and number of children under 18 in a household, play an important role in the demand for fruits and vegetables. In contrast, urban status, household heads' educational level and age are suggested to affect only high-income households' demand decisions. In addition, region has no impact on demand for processed fruits and vegetables for both income groups of households, while the number of persons over 64 in a household does not influence demand for fruits and vegetables.

Conditional elasticities show that processed fruits and vegetables, fresh fruits and fresh vegetables are "necessities" and demand for them is inelastic. They appear to be "gross complements" and "net substitutes". In general, own-price elasticities for low-income households are larger than those for high-income households, meaning low-income households are more responsive to price changes for all three goods categories. Moreover, low-income households have larger expenditure elasticities for processed fruits and vegetables and smaller expenditure elasticities for fresh fruits and fresh vegetables than high-income households. This

[25] Dong and Lin (2009) report that low-income households have larger price elasticities for vegetables but smaller ones for fruits than high-income households, while Huang and Lin (2000) found both of the own-price elasticities of fruits and vegetables are lower for poverty households.

may be due to the fact that the low-income households are subsidized for fresh fruits and fresh vegetables, so they are not very responsive to changes of total expenditure for fruits and vegetables.

There are some issues worthy of further research. First, in this study, processed fruits and vegetables are considered as one category, so the demand for subcategories of fruits and vegetables in each category cannot be differentiated. To know more about the disparities of demand for disaggregate fruits and vegetables between two income groups of households, more detailed classification is desired. Second, different ways of grouping fruits and vegetables may lead to different results. For example, Okrent and Alston (2011) put fruit juices in the nonalcoholic beverages and Huang (1993) put potatoes in the group of “Staple foods”, while this study puts fruit juices in the category of processed fruits and vegetables, and potatoes in the category of fresh vegetables. Third, this study focuses only on the consumption of fruits and vegetables prepared at home. Food away from home may also influence the results.

Table 1. Own-price and Expenditure Elasticities of Demand in the Previous Studies (Not Segmented by Income)

Paper (Year)	Main data Source	Micro or Aggregregate	Data Frequency	Data Years	Table # in paper	Conditional on the expenditure of
Okrent and Alston (2011)	CEX[1]	aggregate	monthly	1998-2010	A.7.	fruits and vegetables
					7	all goods
Okrent and Alston (2010)	CEX	aggregate	monthly	1998-2006	24	fruits and vegetables
	PCE[2]		annually	1960-2006	26	all goods
Durham and Eales (2010)	Two grocery stores in the Pacific Northwest	aggregate	weekly		5 6	fresh fruits
Brown and Lee (2002)	Fruit and Tree Nuts[3]	aggregate	annually	1980-1998	3	fresh fruits
Malaga and Williams (2002)	U.S. and Mexico production data and U.S. shipment data	aggregate	seasonally	1971-1993	7 8	fresh vegetables
Feng and Chern (2000)	CEX	aggregate	monthly	1981-1995	3 4	food
Henneberry et al. (1999)	Fruit and Tree Nuts, Food For Less (retail food supermarket)	aggregate	annually	1970-1992	2, 5	fresh vegetables /fresh fruits
You et al. (1996)	FCPE[4], Fruit and Tree Nuts	aggregate	annually	1960-1993	1	all goods
Huang (1993)	FCPE	aggregate	annually	1953-1990	1	all goods
Cox and Wohlgenant (1986)	NFCS[5]	micro	cross section	1977-1978	3	all goods

Note:
1. CEX is short for Consumer Expenditure Survey.
2. PCE is short for Personal Consumption Expenditure collected by the U.S. Bureau of Economic Analysis.
3. It is short for Fruit and Tree Nuts, Situation and Outlook Yearbook.
4. FCPE is short for Food Consumption, Prices, and Expenditures, which is issued by the USDA/Economic Research Service.
5. NFCS are short for National Food Consumption Survey.

Table 1. Own-price and Expenditure Elasticities of Demand in the Previous Studies (Not Segmented by Income) (continued)

Paper (Year)	Uncp. or cp. price elast.[1]	Own price elasticities				Expenditure/Income elasticities			
		PFV[2]		FV	FF	PFV		FV	FF
		PF	PV			PF	PV		
Okrent and Alston (2011)	uncp.	-0.84		-0.45~-0.98	-0.60~-1.01	0.77		0.75~1.43	0.77~1.41
	uncp.	-0.77		-0.42~-0.94	-0.58~-1.1	(0.03)[3]		(0.03)~(0.06)	(0.03)~(0.06)
Okrent and Alston (2010)	uncp.	(-0.17)		(-0.24)~-0.85	(-0.28)~-1.25	0.81		0.69~1.41	0.83~1.66
		-0.07		-0.2~-0.77	-0.28 ~-1.18	0.13		0.11~0.23	0.13~0.27
Durham and Eales (2010)	uncp.				-0.98 ~1.62 (store 1) -0.90~-1.68 (store 2)				
Brown and Lee (2002)	uncp.			-0.52~-1.11					0.40~1.75
Malaga and Williams (2002)	uncp.			(-0.21)~-0.53 (winter)				0.85~1.35 (winter)	
	cp.			(-0.01)~-0.33 (winter)					
	uncp.			(-0.17)~-0.66 (summer)				0.74~1.71 (summer)	
	cp.			(-0.02)~(-0.35) (summer)					
Feng and Chern (2000)	uncp.	-0.27	-0.56	-0.61	-0.82	0.83	0.62	0.87	0.74
	cp.	-0.25	-0.55	-0.59	-0.80				
Henneberry et al. (1999)	uncp.			0.84~-1.65	(-0.04)~-2.10			(0.46) ~2.24	
	cp.			(0.15)~-1.50	(0.06) ~-1.47				0.50~5.22
You et al. (1996)	uncp.	-0.35	(-0.14)	(-0.03)	-0.40	(0.34)	(0.27)	(0.29)	(0.11)
Huang (1993)	uncp	(-0.30)		(-0.13)	-0.20	0.43		0.41	(-0.38)
Cox and Wohlgenant (1986)	uncp	-0.2 (canned), -0.67 (frozen)		(-0.20)		-0.08(canned), 0.20 (frozen)		0.07	

Note:
1. "Uncp. or cp. price elast." represents uncompensated or compensated elasticities.
2. PF denotes processed fruits; PV denotes processed vegetables; PFV denotes processed fruits and vegetables
3. The numbers in the parenthesis means they are not significant.

Table 2. Own-price and Expenditure Elasticities of Demand in the Previous Studies (Segmented by Income)

Paper (Year)	Data Source	Micro or Aggregate	Data Frequency	Data Years	Table # in paper	Cond'l on the exp. of[2]	Uncp. or cp. price elast.[3]	High: Own price elasticities		High: Expenditure/Income Elasticities[1]		Low: Own price elasticities		Low: Expenditure/Income elasticities	
								veg.[4]	fruit	veg.	fruit	veg.	fruit	veg.	fruit
Dong and Lin (2009)	Nielsen homescan Data	micro	weekly	2004	2	all goods	uncp.	-0.57	-0.58			-0.69	-0.52		
Huang and Lin (2000)	NFCS	micro	cross section	1987-1988	8	food at home	uncp.	-0.71	-0.75	0.98	1.19	-0.70	-0.65	1.03	1.26
Park et al. (1996)	NFCS	micro	cross section	1987-1988	7, 8	food	uncp.	-0.45	-0.52	0.61/0.26	0.69/0.30	-0.32	-0.34	0.60/ 0.38	0.56/ 0.36

Note:

1. Without further distinction, the values shown are expenditure elasticities; the values before and after "/" are expenditure and income elasticities respectively.
2. "Cond'l on the exp. of" represents "conditional on the expenditure of".
3. "Uncp. or cp. price elast." represents uncompensated or compensated elasticities.
4. "veg." represents vegetable.

Table 3. Variables in the Model and Sample Statistics

	High income group (N=12108)				Low income group (N=4722)			
Variable	**Mean**	**Std. Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**
Urban***	0.925	0.264	0.000	1.000	0.902	0.297	0.000	1.000
Seasonality								
Quar1**	0.251	0.434	0.000	1.000	0.261	0.439	0.000	1.000
Quar2	0.250	0.433	0.000	1.000	0.250	0.433	0.000	1.000
Quar3	0.253	0.435	0.000	1.000	0.253	0.435	0.000	1.000
Reference person's education***								
No degree	0.083	0.276	0.000	1.000	0.315	0.465	0.000	1.000
High	0.261	0.439	0.000	1.000	0.321	0.467	0.000	1.000
College	0.301	0.458	0.000	1.000	0.248	0.432	0.000	1.000
Reference person's race								
Orace***	0.024	0.154	0.000	1.000	0.018	0.133	0.000	1.000
Black***	0.077	0.267	0.000	1.000	0.156	0.363	0.000	1.000
Asian	0.044	0.205	0.000	1.000	0.039	0.194	0.000	1.000
Reference person's sex***								
Male	0.516	0.500	0.000	1.000	0.381	0.486	0.000	1.000
Reference person's age***	51.94	20.226	14.000	86.000	49.030	15.401	16.000	86.000
Region***								
Northeast	0.192	0.394	0.000	1.000	0.173	0.378	0.000	1.000
Midwest	0.259	0.438	0.000	1.000	0.218	0.413	0.000	1.000
South	0.309	0.462	0.000	1.000	0.360	0.480	0.000	1.000
Household size	2.718	1.395	1.000	14.000	2.734	1.773	1.000	14.000
Under 18 (u18)***	0.680	1.030	0.000	10.000	0.957	1.379	0.000	9.000
Over 64 (o64)***	0.300	0.629	0.000	3.000	0.471	0.683	0.000	4.000
log income***	10.947	0.523	9.705	12.543	9.525	0.793	0.000	11.371
log price***								
Processed fruits and vegetables	4.061	0.368	3.746	4.941	4.023	0.350	3.746	4.940
Fresh vegetables	4.035	0.452	3.539	5.115	4.005	0.450	3.539	5.112
Fresh fruits	3.928	0.448	3.432	4.997	3.889	0.443	3.432	4.978
Weekly expenditure ($)***								
Processed fruits and vegetables	5.561	6.580	0.000	130.109	4.652	5.593	0.000	57.010
Fresh vegetables	5.068	6.459	0.000	118.400	4.195	5.471	0.000	79.690
Fresh fruits	5.188	6.894	0.000	127.410	4.077	5.155	0.000	48.880
Budget share (%)								
Processed fruits and vegetables***	0.364	0.311	0.000	1.000	0.371	0.325	0.000	1.000
Fresh vegetables	0.317	0.281	0.000	1.000	0.319	0.294	0.000	1.000
Fresh fruits***	0.319	0.285	0.000	1.000	0.310	0.291	0.000	1.000

Note: ** and *** represent the mean difference between high-income group and low-income group are significant at 5% level and 1% level, respectively.

Table 4. Proportions of Zero Budget Shares

	First week	Second week
Processed fruits and vegetables	21.84%	24.34%
Fresh vegetables	22.82%	24.56%
Fresh fruits	23.25%	24.64%

Table 5. Structural Parameter Estimates

Category n	High						Low					
	PFV 1		Fresh vegetables 2		Fresh fruits 3		PFV1		Fresh vegetables 2		Fresh fruits 3	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.044	0.059	0.285***	0.055	0.184***	0.055	0.099	0.072	0.283***	0.071	0.167***	0.066
Urban	-0.008	0.008	0.014**	0.007	-0.004	0.007	1.85E-04	0.011	0.013	0.009	-0.007	0.010
QUAR1	-0.016***	0.005	-0.006	0.005	0.026***	0.005	-0.024***	0.009	-0.002	0.008	0.027***	0.007
QUAR2	-0.046***	0.005	-0.014***	0.005	0.059***	0.005	-0.041***	0.010	0.009	0.008	0.032***	0.008
QUAR3	-0.058***	0.005	-0.003	0.005	0.066***	0.005	-0.044***	0.009	0.005	0.009	0.041***	0.008
No degree	0.012	0.008	0.006	0.007	-0.014**	0.006	0.004	0.011	0.009	0.010	-0.004	0.010
High	0.016***	0.005	-0.004	0.005	-0.003	0.005	0.002	0.010	0.004	0.010	-0.002	0.010
College	0.004	0.005	0.002	0.004	4.37E-04	0.004	-0.002	0.011	0.001	0.011	-0.001	0.011
Other race	0.003	0.016	0.001	0.014	-0.010	0.015	-0.056*	0.032	0.042*	0.022	-0.002	0.022
Black	0.049***	0.007	-0.029***	0.007	-0.023***	0.007	0.038***	0.009	-0.029***	0.008	-0.017**	0.008
Asian	-0.032***	0.010	0.054***	0.009	-0.023***	0.008	-0.101***	0.020	0.082***	0.017	-0.01	0.014
Sex	0.012***	0.004	-0.004	0.003	-0.011***	0.004	0.013*	0.007	-0.005	0.006	-0.006	0.006
Age	-4.51E-04***	1.75E-04	-6.12E-05	1.65E-04	0.001***	1.66E-04	-3.41E-04	2.73E-04	3.78E-04	2.33E-04	-2.51E-05	2.35E-04
Northeast	0.003	0.006	-0.011**	0.005	-0.003	0.005	0.008	0.011	-0.017**	0.009	-0.011	0.009
Midwest	0.007	0.005	-0.009*	0.005	0.013***	0.005	0.002	0.010	-0.02**	0.010	0.025***	0.009
South	0.005	0.005	0.003	0.005	-0.001	0.005	-0.006	0.009	-0.011	0.008	0.015*	0.008
Household size	0.003	0.002	4.48E-04	0.003	-0.011***	0.002	-0.001	0.004	0.008**	0.004	-0.016***	0.004
Under 18	0.008***	0.003	-0.013***	0.003	0.01***	0.003	0.006	0.005	-0.018***	0.005	0.013***	0.005
Over 64	0.001	0.004	9.50E-05	0.004	0.004	0.004	-0.004	0.007	-0.006	0.007	0.01	0.006
γ_n1	0.359***	0.004	-0.18***	0.003	-0.179***	0.003	0.36***	0.007	-0.181***	0.005	-0.18***	0.005
γ_n2	-0.18***	0.003	0.33***	0.004	-0.15***	0.003	-0.181***	0.005	0.339***	0.006	-0.158***	0.004
γ_n3	-0.179***	0.003	-0.15***	0.003	0.329***	0.004	-0.18***	0.005	-0.158***	0.004	0.338***	0.006
Θ_nt												
Θ_n1	0.004***	0.001	0.007***	0.001	0.004***	0.001	0.003	0.002	0.008***	0.002	0.005**	0.002
Θ_n2	0.005***	0.001	0.007***	0.001	0.004***	0.001	0.006***	0.002	0.006***	0.002	0.006***	0.002
β_n	0.029***	0.003	0.015***	0.003	0.013***	0.003	0.053***	0.006	0.013***	0.005	0.003	0.005
σ_n	5.767***	0.453	2.593***	0.217	2.335***	0.211	5.502***	0.747	3.065***	0.468	1.974***	0.263
δ_1	-0.005***	0.001		δ_2	-0.003***	0.001	δ_1	-0.005**	0.002	δ_2	-1.98E-04	0.002
Heteroscedastic part in variance												
QUAR2	0.017	0.030	-0.012	0.031	0.112***	0.033	0.039	0.049	0.035	0.052	0.1*	0.052
QUAR3	0.104***	0.032	-0.017	0.031	0.128***	0.031	0.059	0.048	-0.024	0.054	0.193***	0.051
No degree	-0.038	0.046	-0.066	0.048	-0.031	0.048	-0.087*	0.050	-0.015	0.058	-0.104**	0.049
High	-0.048*	0.028	-0.047	0.029	-0.102***	0.031	-0.097**	0.049	-0.106*	0.058	-0.109**	0.053
Midwest	-0.112***	0.033	-0.162***	0.034	-0.105***	0.033	-0.035	0.055	-0.084	0.077	-0.076	0.053
South	-0.057*	0.029	-0.096***	0.030	-0.020	0.029	-0.011	0.047	-0.102**	0.047	-0.081	0.050
Over 64	-0.077***	0.019	-0.09***	0.019	-0.054***	0.020	-0.029	0.031	-0.065**	0.030	-0.031	0.033
logp (own)	-1.69***	0.038	-1.312***	0.035	-1.31***	0.038	-1.733***	0.066	-1.424***	0.063	-1.284***	0.059
logy-logP (own)	-0.605***	0.017	-0.566***	0.020	-0.544***	0.018	-0.716***	0.028	-0.614***	0.033	-0.665***	0.030

Note: *, **, and *** denote significance at 10%, 5% and 1% level, respectively; PFV denotes processed fruits and vegetables. The numbers in the parenthesis are standard errors.

Table 6. Structural Parameter Estimates (%Change)

	High			Low		
Category	PFV	Fresh vegetables	Fresh fruits	PFV	Fresh vegetables	Fresh fruits
Intercept	11.79%	90.62%***	58.82%***	25.86%	90.06%***	54.67%***
Urban	-2.21%	4.46%**	-1.23%	0.05%	4.20%	-2.28%
QUAR1	-4.18%***	-1.97%	8.28%***	-6.23%***	-0.60%	8.89%***
QUAR2	-12.34%***	-4.47%***	18.79%***	-10.79%***	2.76%	10.47%***
QUAR3	-15.58%***	-1.10%	21.25%***	-11.44%***	1.53%	13.58%***
No degree	3.26%	1.88%	-4.6%**	1.01%	2.99%	-1.28%
High	4.23%***	-1.24%	-0.98%	0.52%	1.16%	-0.66%
College	0.98%	0.64%	0.14%	-0.53%	0.37%	-0.47%
Other race	0.85%	0.42%	-3.05%	-14.61%*	13.27%*	-0.69%
Black	13.04%***	-9.25%***	-7.37%***	9.95%***	-9.39%***	-5.56%**
Asian	-8.46%***	17.19%***	-7.34%***	-26.5%***	26.07%***	-3.28%
Sex	3.31%***	-1.39%	-3.43%***	3.42%*	-1.55%	-1.82%
Age	-0.12%***	-0.02%	0.23%***	-0.09%	0.12%	-0.01%
Northeast	0.81%	-3.38%**	-0.96%	2.18%	-5.44%**	-3.71%
Midwest	1.78%	-2.97%*	4.28%***	0.65%	-6.44%**	8.08%***
South	1.29%	1.11%	-0.44%	-1.53%	-3.60%	4.95%*
Household size	0.94%	0.14%	-3.53%***	-0.15%	2.6%**	-5.2%***
Under 18	2.14%***	-4.12%***	3.11%***	1.65%	-5.82%***	4.18%***
Over 64	0.28%	0.03%	1.13%	-1.05%	-2.05%	3.41%

Note: *, **, and *** denote significance at 10%, 5% and 1% level, respectively; PFV denotes processed fruits and vegetables.

Table 7. Uncompensated Price Elasticities

	High			Low		
Price of:	PFV	Fresh vegetables	Fresh fruits	PFV	Fresh vegetables	Fresh fruits
PFV	-1.027***	-0.097***	-0.096***	-1.157***	-0.109***	-0.109***
	(0.053)	(0.003)	(0.003)	(0.063)	(0.004)	(0.005)
Fresh vegetables	-0.156***	-2.925***	-0.129***	-0.138***	-3.268***	-0.12***
	(0.004)	(0.116)	(0.002)	(0.005)	(0.19)	(0.004)
Fresh fruits	-0.13***	-0.109***	-0.711***	-0.118***	-0.104***	-0.827***
	(0.004)	(0.002)	(0.022)	(0.005)	(0.004)	(0.038)

Note: *** denote significance at 1% level; PFV denotes processed fruits and vegetables. The numbers in the parenthesis are standard errors.

Table 8. Compensated Price Elasticities

	High			Low		
Price of:	PFV	Fresh vegetables	Fresh fruits	PFV	Fresh vegetables	Fresh fruits
PFV	-0.852***	0.316***	0.203***	-0.896***	0.293***	0.208***
	(0.069)	(0.017)	(0.014)	(0.096)	(0.028)	(0.023)
Fresh vegetables	0.062***	-2.413***	0.242***	0.159***	-2.81***	0.24***
	(0.017)	(0.137)	(0.02)	(0.031)	(0.225)	(0.027)
Fresh fruits	0.073***	0.369***	-0.364***	0.15***	0.31***	-0.501***
	(0.016)	(0.023)	(0.042)	(0.027)	(0.03)	(0.064)

Note: *** denote significance at 1% level; PFV denotes processed fruits and vegetables. The numbers in the parenthesis are standard errors.

Table 9. Expenditure Elasticities

	High			Low		
	PFV	Fresh vegetables	Fresh fruits	PFV	Fresh vegetables	Fresh fruits
Elasticities	0.773***	0.959***	0.896***	0.788***	0.896***	0.81***
	(0.025)	(0.01)	(0.013)	(0.029)	(0.022)	(0.028)

Note: *** denote significance at 1% level; PFV denotes processed fruits and vegetables. The numbers in the parenthesis are standard errors.

Appendix A: Construction of the Fruit and Vegetable category and Definitions of Some Variables Used in the Model

Table A1. Construction of fruit and vegetable categories

Category	Disaggregates of Fruits and vegetables
Processed fruits and vegetables	Frozen fruits, frozen fruit juices, fresh fruit juices, canned and bottled fruit juices, canned fruits, dried fruits, frozen vegetables, canned beans, canned corn, miscellaneous canned vegetables, dried peas, dried beans, other processed dried vegetables, frozen vegetable juices, fresh/canned vegetable juices, other processed fruits and vegetables
Fresh vegetables	Potatoes, lettuce, tomatoes, others fresh vegetables
Fresh fruits	Apples, bananas, oranges, citrus fruits excluding oranges, others fresh fruits

Table A2. Definitions of Some Variables Used in the Model

Variable	Variable Definition
Reference person's urban status	
Base	Rural
Seasonality	
QUAR1	The first quarter
QUAR2	The second quarter
QUAR3	The third quarter
Base	The fourth quarter
Reference person's education	
No degree	Never attended school; Frist through eighth grade; Ninth through twelve grade;
High	High school graduate
College	Some college, less than college graduate; Associate's degree (occupation/vocational or academic)
Base	Bachelor's degree; Master's degree; Professional/Doctorate degree
Reference person's race	
Other race	Multi-race, mainly including American Indian, Native Hawaiian or Other Pacific Islander
Base	White
Reference person's sex	
Base	Female
Region	
Base	West
Household size	Number of members in a CU
Under 18	Number of children under 18 in a CU
Over 64	Number of persons over 64 in a CU

Appendix B. Construct S-L Price Indices

One difficulty in estimating consumer demand is that there is no price data in the CEX. Although CPI is used to overcome the problem, there is no sufficient price variation in aggregate prices compared to the demand variation. Lewbel (1989) proposes an S-L price index to solve this problem.

Assume a "between-group" utility function is weakly separable. Lewbel shows that if the "within-group" utility function has a Cobb-Douglas function form, then the S-L price index for each group can be derived using expenditure shares of goods in that group. That is, $P_i = 1/k_i \prod_{j=1}^{N_i} w_{ij}^{-w_{ij}}$, where w_{ij} is the expenditure share of good j in group i, N_i is number of goods in group i, and $k_i = \prod_{j=1}^{N_i} w_{ij}^{*-w_{ij}^*}$, in which w_{ij}^* is the expenditure share for the reference household and is derived as the sample average across all the households and times.

The S-L price index has sufficient variation since it introduces demographic variation into the prices through budget shares. It is noteworthy that, although the sub-utility function takes the Cobb-Douglas form, there is no restriction on the form of the between-group utility function. The between-group utility function is the LA/AIDS model in this study.

Appendix C: Two-Limit Tobit Model

As it is known, observed budget shares are bounded by 0 and 1. In order to account for this requirement and provide comparable results, a two-limit Tobit model is also estimated. The two-limit Tobit model is specified as $w_{njt} = 0$ if $w_{njt}^* \leq 0$, $w_{njt} = 1$ if $w_{njt}^* \geq 1$, $w_{njt} = w_{njt}^*$ otherwise, in which, as defined in the main body, w_{njt} is observed share of good n for household j at time t and w_{njt}^* is the latent variable. The results are shown in Tables C1, C2 and C3.

Compared to Tables 7, 8 and 9, Tables C1, C2 and C3 show similar results. The signs of the estimated conditional elasticities are same as those derived from the Tobit model, so the relationships between the three fruit and vegetable categories are consistent across the two models. We can also see that, in Table 7 and 8, the own-price elasticities of fresh vegetables are the largest among all the own-price elasticities; in contrast, those elasticities become the smallest after applying the new model (see Table C1 and C2). However, own-price elasticities of fresh fruits become larger compared to the previous results. Moreover, the expenditure elasticities are larger for all the three categories in the two-limit Tobit model (see Table C3). The comparisons of elasticities between low-income households and high-income households show that the results are consistent with the previous results except that own-price elasticities of processed fruits and vegetables are smaller for low-income households than those for high-income households.

Table C1. Uncompensated Price Elasticities

	High			Low		
Price of:	PFV	Fresh vegetables	Fresh fruits	PFV	Fresh vegetables	Fresh fruits
PFV	-1.449***	-0.253***	-0.25***	-1.222***	-0.15***	-0.15***
	(0.059)	(0.008)	(0.008)	(0.054)	(0.009)	(0.009)
Fresh vegetables	-0.212***	-1.003***	-0.176***	-0.16***	-1.039***	-0.14***
	(0.007)	(0.024)	(0.006)	(0.01)	(0.045)	(0.01)
Fresh fruits	-0.227***	-0.191***	-1.128***	-0.165***	-0.144***	-1.133***
	(0.007)	(0.006)	(0.027)	(0.01)	(0.009)	(0.038)

Note: *** denote significance at 1% level; PFV denotes processed fruits and vegetables. The numbers in the parenthesis are standard errors.

Table C2. Compensated Price Elasticities

	High			Low		
Price of:	PFV	Fresh vegetables	Fresh fruits	PFV	Fresh vegetables	Fresh fruits
PFV	-1.21***	0.131***	0.066***	-0.879***	0.264***	0.213***
	(0.08)	(0.017)	(0.016)	(0.087)	(0.022)	(0.022)
Fresh vegetables	0.092***	-0.513***	0.226***	0.219***	-0.58***	0.262***
	(0.015)	(0.051)	(0.017)	(0.02)	(0.083)	(0.023)
Fresh fruits	0.056***	0.265***	-0.753***	0.185***	0.279***	-0.762***
	(0.015)	(0.019)	(0.05)	(0.023)	(0.026)	(0.068)

Note: *** denote significance at 1% level; PFV denotes processed fruits and vegetables. The numbers in the parenthesis are standard errors.

Table C3. Expenditure Elasticities

	High			Low		
	PFV	Fresh vegetables	Fresh fruits	PFV	Fresh vegetables	Fresh fruits
Elasticities	0.787***	1.002***	0.932***	0.884***	0.979***	0.902***
	(0.029)	(0.014)	(0.016)	(0.029)	(0.025)	(0.029)

Note: *** denote significance at 1% level; PFV denotes processed fruits and vegetables. The numbers in the parenthesis are standard errors.

REFERNCES

Abowed, J. M., and Card, D. (1989). On the Covariance Structure of Earnings and Hours Changes. *Econometrica*, 57, 411-445.

Alston, J. M., J. A. Chalfant, and N.E. Piggott. (2001). Incorporating Demand Shifters in the Almost Ideal Demand System. *Economics Letters, 7,* 73-78.

Basiotis, P., A. C. Gerrior, W. Juan, and M. Lino. (2002). The Healthy Eating Index: 1999-2000, CNPP-12, U.S. Department of Agricalture, Center for Nutrition Policy and Promotion.

Blundell, R., P. Pashardes and G. Weber. (1993). What Do We Learn About Consumer Demand Patterns from Micro Data? *American Economic Review, 83,* 570-597.

Brown, M. and J. Lee. (2002). Restrictions on the Effects of Preference Variables in the Rotterdam Model. *Journal of Agricultural Economics, 34*, 17-26

Cox, T.L., A.K. Wohlgenant. (1986). Prices and Quality Effects in Cross-Sectional Demand Analysis. *American Journal of Agricultural Economics, 68,* 908–919.

Deaton, A.S., and J. Muellbauer. (1980). An Almost Ideal Demand System. *American Economics Review 70*, 312-326.

Dietary Guidelines. (2010). United States Department of Agriculture. http://www.cnpp.usda.gov/Publications/DietaryGuidelines/2010/PolicyDoc/PolicyDoc.pdf

Dong, D. and Lin, B. H. (2009). Fruit and Vegetable Consumption by Low-Income Americans: Would a Price Make a Difference? *USDA Economic Research Report, 70*, 7-29.

Durham, C. and J. Eales. (2010). Demand Elasticities for Fresh Fruit at the Retail Level. *Applied Economics, 42*, 1345-1354.

Feng, X., and W. S. Chern. (2000). Demand for Healthy Food in the United States. Selected paper presented at the meetings of the American Agricultural Economics Association, Tampa, FL.

Huang, S. K. (1993). A complete System of U.S. Demand for Food. *Division, Economic Research Service, , U.S. Department of Agriculture. Technical Bulletin No. 1821*.

Henneberry, S.R., K. Piewthongngam and H. Qiang. (1999). Consumer Food Safety Concerns and Fresh Produce Consumption. *Journal of Agricultural and Resource Economics*, *24*, 98-113.

Hoderlein, S. and S. Mihaleva. (2008). Increasing the Price Variation in a Repeated Cross Section. *Journal of Econometrics, 147*, 316-326.

Huang, S. K. (1993). A complete System of U.S. Demand for Food. *Division, Economic Research Service , U.S. Department of Agriculture. Technical Bulletin No. 1821*.

Huang, K., and B. Lin. (2000). Estimation of Food Demand and Nutrient Elasticities from Household Survey Data, Technical Bulletin No. 1887, U.S. Department of Agriculture, Economic Research Service.

Jakubson, G. (1988). The Sensitivity of Labor-Supply Parameter Estimates to Unobserved Indeividaul Effects: Fixed- and Random-Effects Estimates in a Nonlinear Model Using Panel Data. *Journal of Labor Economics, 6,* 302-329.

Kennedy, P. (2003). A Guide to Econometrics. The MIT Press, Cambridge, Massachusetts.

Lewbel, Arthur. (1989). Identification and Estimation of Equivalence Scales Under Weak Separability. *Review of Economic Studies Limited* 56, 311-316.

Lin, B. (2005). Diet Quality Usually Varies by Income Status. *Amber Waves, 3*, 4-5, U.S. Department of Agriculture, Economic Research Service.

Lutz, S.M., D.M. Smallwood, J.R. Blaylock, and M.Y. Hama. (1992). Changes in Food Consumption and Expenditures in American Households during the 1980s. *U.S. Department of Agriculture Statistical Bulletin No. 849,* Human Nutrition Information Service and Economic Research Service, Washington, DC .

Málaga, J. and G. Williams. (1999). Demand Systems and Fresh Vegetables: An Application of the Barten Approach. *Paper presented at SAEA annual meeting, Memphis Tennessee, January.*

McCracken, V. A. (1992). The U.S. Demand for Vegetables. In R. Lopez, & L. Polopolus (Eds.), *Vegetable Markets in the Western Hemisphere: Trends, Policies, and Trade Linkages* , Ames, IA: Iowa State University.

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*, 660-672.

Meyerhoefer, C.D. (2002). The Econometrics of Non-Standard Structural Demand Modeling: Applications to Transition Country Data. PhD dissertation, Cornell University, 2002.

Moschini, G. (1995). Units of Measurement and the Stone Index in Demand Systems Estimation. *American Journal of Agricultural Economics*, *77*, 63-68.

Nayga Jr., R. (1995a). Microdata Expenditure Analysis of Disaggregate Meat Products. *Rev. Agr. Econ.*, *17*, 275-285.

Nayga Jr., R. (1995b). Determinants of U.S. Household Expenditures on Fruit and Vegetables: a Note and Update. *J. Agr. and Applied Econ., 27*, 588-594.

Okrent , A and J.M. Alston. (2011). The Demand for Disaggregated Food-Away-from-Home and Food-at-Home Products in the United States. *Paper presented at 2011 AAEA annual meeting , Pittsburgh, Pennsylvania, July.*

Okrent, A. and J.M. Alston. (2010). Deamnd for Food in the United States. *RMI-CWE working paper* number 1002.

Park J.L., R.B. Holcomb, K.C. Raper and O. Capps, Jr. (1996). A Demand Systems Analysis of Food Commodities by U.S. Households Segmented by Income. *American Journal of Agricultural Economics* 78: 290-300.

Pollad, R.A. and T.J. Wales. (1981) Demographic Variables in Demand Analysis. Econometrica, 49, 1533-1551.

Raper, K.C., M.N. Wanzala, and R.M. Nayga, Jr.(2002). Food Expenditures and Household Demographic Composition in the US: A Demand Systems Approach. *Applied Economics* 34: 981–992.

Stewart H., Blisard N., and Jolliffe D. (2003). Do Income Constraints Inhibit Spending on Fruits and Vegetables Among Low-income Households. *Journal of Agricultural and Resource Economics*, 28, 465-480.

You, Z., J.E. Epperson, and C.L. Huang. (1996). A Composite System Demand Analysis for Fresh Fruits and Vegetables in the United States. *Journal of Food Distribution Research, 27*, 11-22.

Wooldridge, J. (2002). Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge, Massachusetts.