ABSTRACT

NIU, LUYUAN. An Empirical Investigation of Consumer Demand for Fruits and Vegetables in the U.S. (Under the direction of Dr. Michael K. Wohlgenant).

There are many policy questions concerning how to encourage consumers to buy healthier food. Among these concerns are the effectiveness of subsidizing low-calorie, high-nutrition foods such as fruits and vegetables to improve public health. Many food policies and food assistance programs, such as the Supplemental Nutrition Assistance Program (SNAP), are targeted at low-income households where per capita fruit and vegetable consumption is the lowest. For policies and programs like 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. In addition to providing information on demographic and other socio-economic variables, this dissertation focuses on quantifying consumer responses to prices and incomes in a complete demand system framework.

Chapter One estimates demand for three fruit and vegetable products (fresh fruits, fresh vegetables, and processed fruits and vegetables) by two income groups (high and low) as a weakly separable group using data from the Consumer Expenditure Survey (CEX) from 2002 to 2006. A censored demand system, using a correlated random effect specification, is applied and the model is estimated by quasi-Maximum Likelihood Estimation. Results show that the demand for fresh fruits, fresh vegetables, and processed fruits and vegetables (conditional on expenditure for all fruits and vegetables) for low-income households are more price elastic than those for high-income households. All three fruit and vegetable products are found to be gross complements and net substitutes. Season, region, and a number of demographic variables such as household heads’ race and gender and household size and composition affect consumption of fruits and vegetables for both categories of households. In contrast, urban status, household heads’ educational level, and age have an impact on high-income households only.

Chapter Two uses a different approach to reexamine conditional demand for fruits and vegetables using the 1996–2010 CEX data. Because the panel data used in the first chapter only have two time periods, this study overcomes the limitation by constructing a "pseudo-panel" from the mean of the grouped observations. The groups, which are selected according to the time and income variables, can be tracked over time, although individuals within the group cannot. This study extends the theoretical framework of single-equation pseudo demand to a demand system where cross-equation restrictions can be imposed. A common problem in the empirical work is that time-invariant group effects are neglected during the estimation, which leads to an inconsistent Ordinary Least Squares (OLS) estimator and an invalid hypothesis testing of the fixed effects estimator. However, this study accounts for the time-invariant group effects and proposes an efficient Generalized Method of Moments (GMM) estimator. Results are consistent with Chapter One except that processed fruits and vegetables and fresh vegetables are found to be net substitutes, but the corresponding estimated cross-price elasticities are insignificant.

Chapter Three investigates how a price subsidy affects demand for three fruit and vegetable products for two income groups of households. This study combines the results from Chapter Two and develops a two-stage budgeting approach to estimate a complete demand for fruits and vegetables using 1986–2010 quarterly CEX data. Results show that low-income households have larger total expenditure elasticities but smaller unconditional price elasticities than high-income households. Fruits and vegetables and all other goods are found to be net substitutes. Assuming that supplies for fruits and vegetables are perfectly elastic, a 10% price subsidy increases consumption of processed fruits and vegetables, fresh vegetables, and fresh fruits by 3.27% (10.68%), 3.29% (10.73%), and 3.50% (11.42%) for low-income (high-income) households and only causes a small change in consumption of all other goods.

An Empirical Investigation of Consumer Demand for Fruits and Vegetables in the U.S.

by
Luyuan Niu

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina

2013

APPROVED BY:

Michael K. Wohlgenant Chair of Advisory Committee	Atsushi Inoue
Xiaoyong Zheng	Ivan T. Kandilov

DEDICATION

To my parents.

BIOGRAPHY

Luyuan Niu is a PhD candidate in the Department of Agricultural and Resource Economics at North Carolina State University. Her research has focused on the areas of consumer demand behavior, food studies, applied econometrics, and policy analysis.

ACKNOWLEDGMENTS

I would like to express my deep gratitude to my graduate advisor, Dr. Michael Wohlgenant, for his excellent guidance, constant encouragements, and rooted knowledge. Without his immeasurable support, this dissertation would not have been possible. I would also like to thank the other three committee members: Dr. Atsushi Inoue for his expertise on econometrics, patience in answering my questions, and constant support; Dr. Xiaoyong Zheng for his ingenuity, insights, and great comments; and Dr. Ivan Kandilov for sharing his experience and knowledge not only in research but in job hunting. I would also like to thank Dr. Tamah Morant for departmental support, and all other faculty or staff members who have provided me help during my Ph.D. study. Special thanks to my parents and my husband for their love and always being there for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 A Censored Demand System Analysis of Fruits and Vegetables by Different Income Groups Using Micro Data

1.1 Introduction

On average, Americans' consumption of fruits and vegetables falls short of the recommended amount outlined in the Dietary Guidelines for Americans, whose 2010 guidelines recommend average daily intakes of 2.0 cups of fruits and 2.5 cups of vegetables at the 2,000 calorie level of the U.S. Department of Agriculture (USDA) Food Patterns.[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 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, respectively, compared to 1.14 and 1.72 cups consumed by high-income households using 1999–2002 National Health and Nutrition Examination Survey (NHANES) data.[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

[1] See what constitutes a cup at USDA's ChooseMyPlate.gov.

[2] In the NHANES sample, the average calorie intake is 2,164 calories per day.

[3] Dong and Lin (2009) defined a high-income (low-income) household as the one in which the annual income is greater than 300% (below or equal to 130%) of the poverty guidelines. Differently, my study defines a high-income (low-income) household as the one in which the annual income is above (below or equal to) 185 % of the federal poverty guidelines after adjusting for the household size.

model in their analysis (e.g., Lin 2005; 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 demand system and examines the differences between fruit and vegetable consumption patterns in low-income and high-income households. In addition to providing 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.

This study differs from other studies in the following aspects: First, the demand system is estimated using micro-level data rather than aggregate data (see Tables 1.1 and 1.2). Second, due to the fact that price data for individuals are unavailable, the Stone-Lewbel price indexes 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; Huang and Lin 2000).[4] Fourth, elasticities are estimated by income group. Also, elasticities of fresh fruits, fresh vegetables, and processed fruits and vegetables are derived for individual consuming units, which is different from other studies where elasticities are estimated at the market level or fruits or vegetables are aggregated in all forms (e.g., Dong and Lin 2009; Huang and Lin 2000; Park et al. 1996). Finally, following Meyerhoefer, Ranney and Sahn (MRS) (2005), a correlated random effects model is used 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. Different from MRS (2005), this study estimates the impacts of demographic and other socio-economic variables on demand for fruits and vegetables by disentangling the coefficients in the second-stage estimation. In addition, this study estimates both Tobit and two-limit Tobit models.

In this study, all “fruits and vegetables” are divided into three categories: fresh fruits, fresh vegetables, and processed fruits and vegetables. Figure 1.1 shows the trends in per

[4] Zero observations are those with zero consumption of one or more goods.

capita real dollar 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 were $143 in 2009 and $146 in 1999. Processed fruits and vegetables show a decline in expenditures during the period between 1999 and 2009, down to $153 in 2009, 24% less than in 1999. 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 higher quality. Increasing fresh selections and quality in grocery stores improves households' choices, which may decrease their need for processed fruits and vegetables. Per capita real dollar expenditures are calculated as average annual expenditures of all households divided by the Laspeyres Consumer Price Index (CPI) (Dec 1997=100). Figure 1.2 illustrates the trends in the CPI for fruits and vegetables during the period from 1999 to 2009. The CPI for fresh fruits and fresh vegetables increased steadily from 1999 to 2008 and began to decline in 2008. The CPI for processed fruits and vegetables remained relatively steady before increasing noticeably in 2007. Figure 1.3 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 Consumer Expenditure Survey (CEX) data from 2002 to 2006, the average weekly expenditures for processed fruits and vegetables, fresh fruits, and fresh vegetables are $4.65, $4.20, and $4.08, respectively, for low-income households, while they are $5.56, $5.07 and $5.19 for high-income households, respectively.

The remainder of the chapter is organized as follows. The next section reviews the literature on fruit and vegetable demand. The third section presents an overview of demand theory. The fourth section discusses CEX data and gives variable construction and descriptive statistics. In the fifth section, demand model and econometric methodology are given. The sixth section presents the results and makes comparisons between both income groups of households. The final section summarizes and concludes the first part of the chapter.

1.2 Literature Review

This review only focuses on demand for fruits and vegetables in the United States because different markets and commodity and consumer characteristics across countries would introduce different interpretations when making comparisons with the U.S. data. Although only combined categories of fruits or vegetables (for example, fresh fruits and fresh vegetables) are studied, several studies on individual types (for example, apples) are also reviewed.[5] A summary of estimated own-price elasticities and expenditure elasticities from previous studies are shown in Table 1.1 and Table 1.2.

Several approaches are applied to estimating demand equations. They start with different specified functional forms of direct utility, indirect utility or cost, or directly with a demand system. They also use different datasets.

Time series (aggregate) data are used in a number of studies. Huang (1993) develops an estimation methodology to estimate large-scale disaggregate demand systems with limited sample observations. He uses constrained maximum likelihood with a substitution approach and estimates the first-order differential-form demand system. In the first stage, an aggregate demand system is estimated. Food is divided into seven food categories and a nonfood sector. He solves two problems that plague the estimation: insufficient degree of freedom due to the fact that number of sample observations is less than the parameters in each demand equation and possible multicollinearity between price and expenditure variables. The own-price and total expenditure elasticities of fresh fruits, fresh vegetables, and processed fruits and vegetables are estimated. Cross price elasticities show that fresh fruits and fresh vegetables are substitutes, while processed fruits and vegetables are complements of both fresh fruits and fresh vegetables.

[5] Demand for individual types would be expected to be more elastic due to the availability of substitutes within the group.

You et al. (1996), following the method proposed by Huang (1993), estimate a composite demand system for U.S. fruits and vegetables. They use annual data covering eleven food categories and a nonfood sector. The results indicate that income does not have an important impact on aggregate consumption of fresh fruits and vegetables because the expenditure elasticities are found to be insignificant. The estimated cross price elasticities show that both fresh and processed vegetables are substitutes to fresh fruits. Processed fruits and fresh vegetables are complements. Processed vegetables and fresh fruits are also found to be complements.

Málaga and Williams (1999) study the U.S. seasonal demand for fresh vegetables using Barten's (1993) synthetic model. Both the winter and the summer season models are evaluated. The three-stage least squares estimator is used to solve the endogeneity problem in the estimation. Although all the price and expenditure elasticities are within the expected ranges, strong seasonal differences are indicated for some commodities such as cucumbers and bell peppers. The authors suggest that this may be due to different salad consumption habits.

Brown and Lee (2002) develop a Rotterdam specification allowing for changes in preferences on demand through marginal utilities. Their approach accounts for the effects of the female labor force participation (FLP) rate on the demand for fresh fruits. The estimated elasticities show that FLP only has a negative effect on the marginal utility for oranges, reflecting the preferences for convenience in fresh fruit consumption.

Okrent and Alston (2011) estimate the elasticity of demand for fruits and vegetables by estimating both Barten's (1993) model and the generalized ordinary differential demand system applied to two datasets: the Bureau of Economic Analysis (BEA) and the Bureau of Labor Statistics (BLS) datasets. They use a two-stage budgeting method, where estimated elasticities of demand based on Barten's (1993) model with annual BEA data and monthly CEX data, respectively, are used for the first- and second-stage demand estimation. They find that BEA-data–based own-price elasticities for food (excluding food away from home, abbreviated FAFH) are smaller than the average ones in the literature. Also BEA-data–based estimated elasticities have a lower average mean absolute error (MAE) from conditional-on-price forecasts than the CEX-data–based estimates and the averages in the studies that are based on data distinguishing FAFH from food at home (FAH).

Henneberry et al. (1999) apply the Linear Approximate Almost Ideal Demand System (LA/AIDS) to see if the risk information has a significant impact on consumption of fresh fruits and vegetables. They find that the impact is very small and statistically insignificant on almost all the produce studied. They also find that different consumers have different responses for each commodity.

Feng and Chern (2000) estimate a modified LA/AIDS model to study the demand for healthy food in United Sates. They select eight healthy food groups and conclude that fresh fruits and fresh vegetables have larger own-price elasticities than processed fruits and vegetables. Moreover, fresh fruits and fresh vegetables are complements, as are processed fruits and processed vegetables. In contrast, fresh fruits and processed fruits are substitutes, as are fresh vegetables and processed vegetables. They find that the larger the value of food stamps received by the households, the smaller the demand for fresh and processed fruits.

Durham and Eales (2010) compare four demand models: the Double-Log model, the LA/AIDS, the AIDS, and the Quadratic Almost Ideal Demand System (QUAIDS). They use the store-level Pacific Northwest datasets of fresh fruits in estimation and conclude that the QUAIDS had the lowest overall root mean error. The QUAIDS is then used to re-estimate the elasticities. The own-price elasticities turn out to be larger than the ones in previous studies.

The study by Nayga (1995b) is one of the early works that introduces socio-demographic variables in the demand model of fruits and vegetables. Engel curves are estimated with 1992 CES data. Fruits and vegetables are divided into four categories: fresh fruits, processed fruits, fresh vegetables, and processed vegetables. The author uses a generalized Heckman procedure developed by Heien and Wessells (1990) and Nayga (1995a) to deal with the zero expenditure problem in the observations. Results show that household size is positively related to household expenditures for all four groups. All the expenditure elasticities with respect to household size are inelastic (different from McCracken 1992). Households with senior heads spend more on all groups and households headed by a black person spend less on processed fruits and fresh vegetables than those headed by a white person (different from Lutz et al. 1992). Households with higher education levels consume more fresh vegetables (consistent with Lutz et al. 1992).

Park et al. (1996) use a demand system for 12 food commodities. They employ a Heckman two-step procedure to account for the zero expenditure in the data from 1987-88 National Food Consumption Survey (NFCS) and apply the Linear Expenditure System (LES) in the second step to examine the subsistence expenditures, own-price elasticities, and expenditure and income elasticities for poverty and nonpoverty households. They find that own-price elasticities are lower and income elasticities are higher for poverty households for both fruit and vegetable categories.

Huang and Lin (2000) use a modified AIDS model to estimate the quality-adjusted food demand elasticities for households segmented at three income levels. These elasticities are further employed to estimate nutrient elasticities for low-income households. The results for fruits and vegetables are consistent with Park et al. (1996): Low-income households are less responsive to the price changes of fruits and vegetables.

Stewart et al. (2003) answer the questions of whether the "poor" in general spend significantly less money on consumption of fruits and vegetables than the "non-poor." Fruits and vegetables are combined in an aggregate group. They estimate Engel curves and test for first-order stochastic dominance with the 2000 CES data. Data are censored at zero and the censored least absolute deviations (CLAD) estimator is used. They conclude that the "poor" spend less on fruits and vegetables, but the income coefficients are insignificant. In contrast, the "non-poor" consume more when income increases. However, no elasticities are derived in their paper.

Using 2004 Nielsen Homescan Data, Dong and Lin (2009) estimate a single-equation model instead of a complete demand system for fruit and vegetable consumption, where the purchase quantity of fruits and vegetables is a function of unit values and households' socio-demographic variables. They report that low-income households have larger price elasticities for vegetables but smaller ones for fruits.

Andreyeva et al. (2010) review 160 studies undertaken between 1938 and September 2007 on price elasticities of demand for major food groups. Eleven percent of the studies provide estimates of price elasticities for fruits or vegetables, and only 6 estimates are on aggregate categories. They calculate that mean price elasticity estimates for fruits and vegetables are -0.70 and -0.58, respectively.[6] They also conclude that vegetables are not sensitive to study methodology, but fruits do vary by type of demand system and data. However, their study does not distinguish conditional and unconditional elasticities.

1.3 Introduction to Demand Systems

There are many alternative ways to derive demand equations for econometric estimation.[7] If one starts from an assumed specific form of utility function, first-order conditions of the maximization problems give the form of the demand system. For example, the LES (Klein and Rubin 1948) is derived from an additive utility function. If one manipulates marginal utility of income and the Hotelling-Wold identity, an inverse demand system can be derived.

One can also obtain a demand system through duality theory starting with either the indirect utility function or the cost function. When given enough parameters, it is flexible enough to approximate an arbitrary representation of preferences. For example, the Indirect Translog demand system (Christensen, Jorgenson and Lau 1975) is given by applying the Roy's identity on a quadratic form indirect utility function; the AIDS is given by applying Shephard's lemma on a PIGLOG (short for "price independent generalized log linearity") form cost function.

Another approach is to directly approximate the conceptual demand model and impose and test the general restrictions of consumer behavior. Examples include Theil's Rotterdam Model and Huang's differential-form demand system (Huang 1993).

[6] 95% confidence intervals are -1.07 to -0.56 for fruits and -1.11 to -0.21 for vegetables.
[7] Here, it refers to the quantity-dependent demand equations.

In general, the assumed functional forms of direct utility, indirect utility or cost function are theoretically consistent. However, some of these models produce quite restrictive specifications. For instance, the additive utility function limits the use of the LES in empirical work. First, it does not admit inferior goods and all the goods must be net substitutes. Second, because the LES is obtained from a direct utility function, homogeneity and symmetry will hold at every point and cannot be tested statistically. Third, absolute values of all uncompensated own-price elasticities are constrained to be less than their income elasticities, which contradicts empirical evidence for food demand in the United States.

There are several advantages of Huang's differential-form demand system. First, it is linear in parameters so the computational burden is reduced compared to the LES and AIDS. Second, the parameters can be interpreted as elasticities directly, so it can avoid any differences or theoretically incorrect measurement in elasticities as in the LA/AIDS (Green and Alston 1991). Third, the dependent variables are changes in quantities rather than changes in expenditure shares, which is convenient when using time-series quantity index data. However, the differential-form demand system can only impose theoretical restrictions, such as Engel aggregation and symmetry, locally at the point where fixed expenditure weights are selected. In contrast, models like the Rotterdam Model and AIDS enforce the restrictions globally.

The advantages of using the AIDS, as pointed out by Deaton and Muellbauer (1980), are the following: First, it is a locally flexible function form, providing a first-order approximation for any true demand system, whether derived from utility-maximizing behavior or not. Second, it has the property of exact aggregation over consumers. That is, it permits that market demand is represented by the decision of a rational representative consumer, which is useful especially when conducting welfare analysis. Third, theoretical restrictions are simple to test by imposing linear restrictions on fixed parameters. Fourth, its Engel curves are consistent with household data. Although the Translog Model or Rotterdam Model possesses one or the other property, neither of them possesses all the properties at the same time. Thus, the AIDS model is chosen in this study. The next subsection gives the theoretical specification of the AIDS in details.

1.3.1 The AIDS Model

Assume the cost function takes the PIGLOG form:

(1.1) $$\log c(u,\ p) = a(p) + ub(p),$$

where p is an N-vector of prices and c(u, p) represents the minimum expenditure to achieve a specific utility u. The function forms of a(p) and b(p) are assumed to be linearly homogenous in prices, which are specified as follows:

(1.2) $$a(p) = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2}\sum_k \sum_i \gamma_{ki} \log p_k \log p_i$$
$$b(p) = \beta_0 \Pi p_k^{\beta^k},$$

where b(p) is the Cobb-Douglas price aggregator. Substituting a(p) and b(p) into the cost function, logc(u, p) becomes

(1.3) $$\log c(u,\ p) = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2}\sum_k \sum_i \gamma_{ki} \log p_k \log p_i + u\beta_0 \prod_k p_k^{\beta_k},$$

where α's, β's, and γ's are parameters. According to economic theory, c(u, p) is homogenous of degree one in prices provided that

(1.4) $$\sum_k \alpha_k = 1,\ \sum_i \gamma_{ki} = \sum_k \gamma_{ki} = 0 \text{ and } \sum_k \beta_k = 0.$$

Slutsky symmetry is satisfied when

(1.5) $$\gamma_{ik} = \gamma_{ki}.$$

By applying the logarithm version of Shephard's lemma—that is, $w_i = \frac{\partial \log(c(u,\ p))}{\partial \log p_i}$ — to the cost function logc(u, p), the share functions are derived as

(1.6) $$w_i = \alpha_i + \sum_k \gamma_{ik} \log p_k + \beta_i u \beta_0 \prod_k p_k^{\beta_k} \quad \forall i,$$

where w_i is the budget share of good i. Substituting the indirect utility function, which is obtained by inverting the cost function log(u, p), into the share function w_i, the AIDS model is derived as

(1.7)
$$w_i = \alpha_i + \sum_k \gamma_{ik} \log p_k + \beta_i \log(x / P) \quad \forall i,$$

where

(1.8)
$$\log P = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2}\sum_k \sum_i \gamma_{ki} \log p_k \log p_i,$$

where x is the total expenditure. It is straightforward to interpret the parameters in the AIDS. α_i represents the estimated budget share of good i when all logarithm prices and real expenditures are zero and can be interpreted as the subsistence budget share of good i. β_i is the coefficient of real expenditures, measuring the effects on product i's budget share. A good is a luxury if $\beta_i > 0$ and is a necessity if $\beta_i < 0$. γ_{ik} is the change in the good i's budget share with respect to a percentage change in good k's price when holding real expenditures constant. If $\gamma_{ik} > 0$, products i and k are substitutes, and if $\gamma_{ik} < 0$, then they are complements.

1.3.2 The LA/AIDS Model

Because the AIDS is a nonlinear model in parameters, the logP is often replaced by the Stone indexes (denoted as P^*) in empirical work, in which

(1.9)
$$\log P^* = \sum_i w_i \log p_i,$$

and new intercept α_i^* is $\alpha_i - \beta_i \alpha_0$. This modified model is called the LA/AIDS model. The LA/AIDS is not only easy to estimate but useful to avoid the potential multi-collinearity problems of using logP. However, there are two problems associated with the LA/AIDS. One problem is that it is unclear what the correct elasticity formulas should be. Asche and Wessells (1997) point out that the LA/AIDS has the same form of price and income elasticities as the AIDS only at the point of approximation where all prices and income are normalized at unity. The other problem, as pointed out by Moschini (1995), is that the Stone price indexes are not invariant to price scaling. Alternative solutions to this problem are suggested by Moschini, and are discussed in the next subsection.

1.3.3 Property of Closed Under Unit Scaling (CUUS) in the AIDS and LA/AIDS Estimation

A demand system is desired if the parameter estimates are invariant to the units of measurement. This feature is considered as CUUS. Sometimes data used in the estimation may not have the universal units. For example, quantities may be measured in different ways for the same good. In food demand, consider the case of oranges. One kilogram of oranges is approximately 2.20 pounds, while the corresponding prices for the "kilogram" unit is 2.2 times than the one for the "pound" unit. However, the total expenditure and expenditure shares are unchanged in both the scenarios, which may cause problems in estimation. The following discusses the cases for estimation with the AIDS and LA/AIDS models.

For the original nonlinear AIDS model, it is not consequential because only the intercepts are affected. Thus, the concerned economic effects are not affected (Moschini 1995). However, when incorporating demand shift variables in the AIDS model in a traditional way by allowing intercepts to be functions of demand shifters, the intercepts are not only changed but vary across data; hence elasticities and relevant welfare measures are no longer invariant to the scaling (Alston, Chalfant, and Piggott 2001). One alternative way to solve this scaling problem is to use the generalized AIDS (GAIDS) suggested by Alston, Chalfant and Piggott (2001), where they augment the GAIDS model with pre-committed quantities that are linear functions of the demand shifters. They show that pre-committed expenditures have not been changed after scaling, which maintains the invariance of the model to changes in the units of measurement.

As written, LA/AIDS model uses the Stone price index ($\log P_t^* = \sum_{i=1}^{n} w_{it} \log p_{it}$, where w_{it} are shares for good i at time t; p_{it} is the price of good i at time t) to approximate the original nonlinear AIDS model. Although it makes the estimation convenient, it is not consistent with the CUUS property (Moschini 1995). He shows that the problem can be solved if the price index P(p) is only affected by a constant after scaling of prices and quantities. That is, $P(p) = k * P(\tilde{p})$, where k is an arbitrary constant and $\tilde{p}$ is a set of new prices according the new units. If this relationship holds, changes in the measurement of price units only affects the intercepts, as in the case of the AIDS model. Other solutions provided by Moschini are to use other price indexes instead of the Stone indexes. These include the Tornqvist index ($\log P_t^T \equiv \frac{1}{2}\sum_{i=1}^{n}(w_{it} + w_i^0)\log(\frac{p_{it}}{p_i^0})$), the loglinear analogue of the Paasche price index ($\log P_t^S \equiv \sum_{i=1}^{n} w_{it}\log(\frac{p_{it}}{p_i^0})$), and the loglinear analogue of the Laspeyres price index ($\log P_t^L \equiv \sum_{i=1}^{n} w_i^0\log(\frac{p_{it}}{p_i^0})$, or simply, $\log P_t^L \equiv \sum_{i=1}^{n} w_i^0\log(p_{it})$).[8] The latter two retain some features of the Stone index. These alternatives are not unique in that they depend on the particular base choice. It is worth noting that, when the intercept of LA/AIDS is augmented by demand shifters, the CUUS property is not affected.[9]

The LA/AIDS model is used in the study because it enables us to reduce the burden of calculation compared to the original AIDS model. The AIDS's CUUS property is preserved by using the price index $\log P_t^L$ instead of the Stone index.

1.4 Data[10]

This chapter uses the 2002–2006 CEX Diary Survey data. The CEX data are collected by the U.S. Census Bureau for the 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 are widely used in economic research and analysis and are also used to support and maintain the CPI.

---

[8] The Paasche Price Index in share form is $P_t^S = (\sum_{i=1}^{n} w_{it}(\frac{p_{it}}{p_i^0})^{-1})^{-1}$.

[9] It is because the demand shifters do not show up in the expression of the price indexes, so there are no impacts of the change of unit measurement on the intercept.

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

The CEX is designed as a national probability sample of households representing the total civilian noninstitutional population of the United States and portion of the institutional population. The CEX 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 data for the entire year, while DS keeps records of small and frequently purchased items such as food. DS is collected over two consecutive one-week periods, so data are weekly and represent a short panel.

The FMLY (Consumer Units [CUs] characteristics and income file) and MEMB files (member characteristics and income file) include demographic and socioeconomic information of all the CUs.[11] Based on the literature review, 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, gender, and age.[12]

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.[13,14] Because 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, the data are divided into two groups: 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.[15] Federal poverty guidelines are issued each year by the Department of Health and Human

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

[12] 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.

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

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

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

Services 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 to determine financial eligibility for certain programs, such as the SNAP. There are 4,722 households from the low-income group and 12,108 households from the high-income group. Table 1.3 reports the descriptive statistics for both groups of households.

Because the CEX is a micro-level dataset, a large number of zero observations are available to construct expenditure shares. Table 1.4 shows the proportion of zeros in each group for 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 prices exist for individuals in the CEX, the CPI and Stone-Lewbel (S-L) price indexes are used to overcome the problem. 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."[16] A quarterly CPI series is used and changed to December 1997=100 base.[17,18] Due to the insufficient variation for the CPI, S-L price indexes are created for individual households following the approach proposed by Lewbel (1989). The construction method of the S-L price indexes can be found in Appendix B.[19]

1.5 Model and Methodology

Because the CEX data represent a short panel, there are two models that can be used 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 h_j (j denotes individual

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

[17] The CPI of processed fruits and vegetables does not include fruit juice, although 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.

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

[19] 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.

household here) is treated as a random variable.[20] The difference is that the fixed effects model allows h_j to have an arbitrary correlation with other observed regressors and no distribution is assumed for it. In contrast, the random effects model assumes h_j is a conditional distribution and orthogonal to other regressors.

Considering the data structure, each household has two observations: one is recorded in the first survey week and the other in the second survey week. For most households, the observed demographic variables included in the model are constant during this short two-week survey period. If the fixed effects approach is applied, time-constant demographic variables would not play a role in the estimation because they would drop out of the estimating equation after the fixed effects transformation, or, in other words, the time-demeaned explanatory variables would contain zero columns, which fails rank condition requirements. Intuitively, if the time-constant explanatory variables not included in the model are correlated with the unobserved h_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).

This study uses the random effects approach. However, the zero correlation assumption between h_j and observed regressors is not appropriate here in this context because the unobservable household effect is likely to be correlated with observed demographic variables, total expenditure, and constructed prices.[21] Hence, a correlated random-effect method (Jakubson 1988) is adopted.

Following MRS (2005), assume the cost function takes the PIGLOG form:

[20] Traditionally, h_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.
[21] 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.

(1.10)

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

where $\log c(\cdot)$ 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, h_j is unobserved household specific effects, and 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 Σ_ε, and $E(\varepsilon_{kjt} \mid \text{all independent variables}) = 0$.

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 the logarithm version of Shephard's lemma, the AIDS model (Deaton and Muellbauer 1980) is obtained as

(1.11)

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

where

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

$u_{njt} = \varepsilon_{njt} - \beta_n \sum_k \log p_{kjt} \varepsilon_{kjt}$, $\gamma_{ki} = 1/2(\tilde{\gamma}_{ki} + \tilde{\gamma}_{ik})$, 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 $\psi_k = 1$.[22] 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.

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

(1.12) $$\sum_k \gamma_{ik} = 0 \text{ and } \gamma_{ki} = \gamma_{ik}.$$ [23]

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

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

where v_j is an error term with normal distribution of mean zero and variance σ^2_v 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 h_j.

By substituting h_j into the demand function, one can get

[22] Intuitively, because the variable h is unobservable and almost has no measurement unit, it would not make sense to estimate its partial effect (Wooldridge, 2002).
[23] The adding-up restrictions are not imposed and the reason is discussed later.

$$w^*_{njt} = \alpha_n + \mu_n T_q + \sum_l (\lambda_{kl} + \eta_l) d_{lj} + \sum_i (\gamma_{ni} \log p_{ijt} + \sum_t \theta_{it} \log p_{ijt}) + \beta_n (\log y_{jt} - \log P_{jt}) + \sum_t \delta_t (\log y_{jt} - \log P_{jt}) + \tilde{u}_{njt}, \quad (1.14)$$

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.

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

$$w_{njt} = \max(0, \; w^*_{njt}), \quad (1.15)$$

where w^*_{njt} is a latent variable. This means, when w^*_{njt} is larger than zero, the observed w_{njt} equals w^*_{njt}; when w^*_{njt} is less than or equal to zero, the observed w_{njt} equals zero. Because it is almost impossible to impose additivity on a Tobit demand system, the restriction is dropped in this study. [24,25]

In the case where censoring is not a problem, the demand system can be estimated with joint 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 represent a short panel, there are in total six (N by T) demand equations that may involve larger than three dimensional integrals, which makes the estimation intractable.

MRS (2005) uses a methodology that applies QMLE to avoid evaluating high dimensional integrals. By manipulating the marginal distributions of each univariate Tobit

[24] One possible solution would be to estimate the first n – 1 equations with the nth good as a residual even though the resulting estimates are not invariant with respect to the equation excluded (Yen and Lin 2006). This will be left for future work.

[25] Because the adding-up restrictions are not imposed on the model, a two-limit Tobit model is also estimated. See the setup and results in Appendix C.

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 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 $\tilde{u}_{njt}$ in the demand equation, its variance is specified as

$$\rho_{njt} \equiv E(\tilde{u}_{njt}^2) = \sigma_n^2 \exp(s_{njt}\xi_n), \tag{1.16}$$

where s_{njt} is the vector of variables that is 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

$$\kappa_n = [\ell\alpha_n \left| \ell\mu' \right| \ell(\lambda' + \eta') \left| \ell\theta_1' + \gamma_{n1}I \;\; \cdots \;\; \ell\theta_3' + \gamma_{n3}I \right| \ell\delta' + \beta_n I \left| \ell\sigma_n^2 \;\; \right| \ell\xi_n'] \tag{1.17}$$

where $\mu' = (\mu_{n1} \cdots \mu_{n3})$, $\lambda' = (\lambda_{n1} \cdots \lambda_{n15})$, $\eta' = (\eta_{n1} \cdots \eta_{n15})$, $\theta_i' = (\theta_{i1} \;\; \theta_{i2})$ 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 three 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 $K = vec(\kappa')$.

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

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

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 reduced-form parameters estimates $\hat{K}$ and the latter involves the structural parameters ϕ. The relationship between K and ϕ is described by a function $a(\cdot)$, $K_0 = a(\phi_0)$, where $a(\cdot)$ is used to disentangle the coefficients and impose the economic restrictions, and the subscript "o" means the true values of the parameters. Because the restrictions are linear, $a(\cdot)$ is also linear with the form $A\phi_0$. The minimum distance 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,

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

which can be obtained from the univariate Tobit estimation.

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 the minimum distance estimator, the estimates of Ξ_0 and the standard error of structural parameter estimates $se(\hat{\phi})$ are obtained by bootstrapping.[26] Specifically, 1,000 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 1,000 new sets of structural parameter estimates are derived. The standard deviation of these estimates is the standard error.

However, results are more reasonable for both income groups of households when using the identity matrix as the weighting matrix. Also, difficulties in empirical applications are reported from using the efficient minimum distance estimator (e.g., Abowd and Card 1989). Because 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

[26] Meyerhoefer (2002) derives the form Ξ_0 by using asymptotic property of GMM.

recommendation. It is noteworthy that the overidentification test is valid only when one employs the optimal weighting matrix.

Elasticity

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

$$\log w(p, y) = \log p + \log q(p, y) - \log y, \tag{1.20}$$

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

$$e_{ijt} = \frac{\partial E(w_{ijt})}{\partial \log y_{jt}} \cdot \frac{1}{E(w_{ijt})} + 1 \tag{1.21}$$

and

$$e_{ik\_jt} = \frac{\partial E(w_{ijt})}{\partial \log p_{kjt}} \cdot \frac{1}{E(w_{ijt})} - \Delta_{ik\_jt}, \tag{1.22}$$

where Δ_{ik_jt} is the Kronecker delta ($\Delta_{ik\_jt} = 1$, for i=k; $\Delta_{ik\_jt} = 0$, for i≠k).[27] 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

$$E(w_{ijt}) = \Phi_{ijt}(x_{jt}K_{ijt} + \rho_{ijt}\frac{\phi_{ijt}}{\Phi_{ijt}}), \tag{1.23}$$

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 $\Phi_{ijt} \equiv \Phi(z_{ijt}) = \Phi(\frac{x_{jt}K_{ijt}}{\rho_{ijt}})$; $\phi(\cdot)$ is the probability density function of standard normal distribution with the same argument defined as above. Thus, expenditure elasticities have the form

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

(1.24)

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

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 because the heteroscedastic variance ρ_{ijt} is constructed in the way that only own-price $\log p_{ijt}$ is included. Specifically,

(1.25)
$$\frac{\partial E(w_{ijt})}{\partial \log p_{kjt}} = \Phi_{ijt}(\gamma_{ik} - \beta_i w_k^o) \quad \forall \ i \neq k.$$

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.

1.6 Results

Results are presented in Table 1.5 for both high-income households and low-income households.[28] The percentage change of parameter estimates is shown in Table 1.6.[29]

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 indicate 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, budget shares of high-income households and low-income households

[28] Definition of variables used in the model can be found in Table A2.

[29] Note that it is interesting to evaluate the effects on observed shares rather than latent shares, so the reported parameter estimates are derived from $\frac{\partial E(w_{ijt}|x_{jt})}{\partial x_{jt}} = \Phi(\frac{x_{jt}K_{ijt}}{\rho_{ijt}})K_{ijt}$.

increase by 15.58% and 11.44%, respectively, in the fourth quarter compared to the third quarter.[30] 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 that 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.

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, household heads with high school degrees demand 4.23% more processed fruits and vegetables, and those without degrees demand 4.60% 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.

Third, household heads' gender and race also influence 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, African American 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.

[30] The percentage change stated in the results only refers to the change in budget shares if there is no further information.

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. 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 variables are also of interest. High-income households in the urban areas tend to purchase 4.46% more fresh vegetables than those in the rural areas. This may be due to the fact that households in the urban areas have more access to fresh vegetables than those in the 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 (under 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 estimated 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 1.7, Table 1.8, and Table 1.9.[31] 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 one, meaning they are "necessities," and are

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

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 three goods are all negative, as expected. All the cross-price elasticities are negative, meaning that all of the goods studied are "gross complements."[32] 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 three goods are "net substitutes."[33] The own-price elasticities for fresh vegetables are a bit larger (2.413 for high-income households and 2.810 for low-income households in absolute values) than the values reported in the literature (see Table 1.1), which may be attributed 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 that low-income households are more sensitive to price changes.[34] Based on the compensated elasticities, the Slutsky matrix is derived and all the eigenvalues are negative. So the negativity condition holds at the points where elasticities are evaluated.

[32] 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.

[33] 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.

[34] 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.

1.7 Conclusion and Directions for Future Work

This study addresses the question how different fruit and vegetable consumption patterns are between low-income households and high-income households using household-level CEX Diary data from 2002 to 2006. To account for the zero observations, a censored demand system is estimated. A correlated random effects approach is used to solve 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 the economic restrictions.

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 may be due to the fact that 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.

Results indicate that seasonality has an obvious impact on consumption of fruits and vegetables for both categories of households. They consume the fewest processed fruits and vegetables and the most fresh fruits in the third quarter, and the most processed fruits and vegetables and the fewest fresh fruits in the fourth quarter.

Household heads' educational level and age are shown to affect only high-income households' demand decisions. Households with higher education levels demand fewer processed fruits and vegetables and more fresh fruits; each additional year of age of the household head induces the household to demand fewer processed fruits and vegetables and more fresh fruits. Moreover, households headed by males purchase more processed fruits and vegetable than those headed by females; white households purchase more fresh vegetables and fresh fruits and fewer processed fruits and vegetables than African American households. Households living in the Northeast and urban areas purchase fewer fresh vegetables than those in the West and rural areas, respectively.

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 the two income groups of households, a more detailed classification is desired. Second, different ways of grouping fruits and vegetables may lead to different results. For example, Okrent and Alston (2012) 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 consumption of fruits and vegetables prepared at home. Food away from home may also influence the results.

Table 1.1: Own-Price and Expenditure Demand Elasticities in the Previous Studies (Not Segmented by Income)

Paper (Year)	Main Data Source	Micro or Aggregate	Data Frequency	Data Years	Table # in Paper	Conditional on the Expenditures of
Okrent and Alston (2012)	CEX	Aggregate	Monthly	1998–2010	A.7	Fruits and vegetables
					7	All goods
Okrent and Alston (2011)	CEX	Aggregate	Monthly	1998–2006	24	Fruits and vegetables
	PCE		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	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, 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	Micro	Cross section	1977–1978	3	All goods

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

Table 1.1 (Continued)

Paper (Year)	Uncomp. or Comp. Price Elast.	Own-Price Elasticities: PFV – PF	Own-Price Elasticities: PFV – PV	Own-Price Elasticities: FV	Own-Price Elasticities: FF	Expenditure/Income Elasticities: PFV – PF	Expenditure/Income Elasticities: PFV – PV	Expenditure/Income Elasticities: FV	Expenditure/Income Elasticities: FF
Okrent and	Uncomp.	-0.84		-0.45~-0.98	-0.60~-1.01	0.77		0.75~1.43	0.77~1.41
Alston (2012)	Uncomp.	-0.77		-0.42~-0.94	-0.58~-1.10	(0.03)		(0.03)~(0.06)	(0.03)~(0.06)
Okrent and	Uncomp.	(-0.17)		(-0.24)~-0.85	(-0.28)~-1.25	0.81		0.69~1.41	0.83~1.66
Alston (2011)		-0.07		-0.20~-0.77	-0.28 ~-1.18	0.13		0.11~0.23	0.13~0.27
Durham and Eales (2010)	Uncomp.				-0.98 ~1.62 (store 1) -0.90~-1.68 (store 2)				
Brown and Lee (2002)	Uncomp.			-0.52~-1.11					0.40~1.75
Malaga and Williams (2002)	Uncomp.			(-0.21)~-0.53 (winter)				0.85~1.35 (winter)	
	Comp.			(-0.01)~-0.33 (winter)					
	Uncomp.			(-0.17)~-0.66 (summer)				0.74~1.71 (summer)	
	Comp.			(-0.02)~(-0.35) (summer)					
Feng and Chern	Uncomp.	-0.27	-0.56	-0.61	-0.82	0.83	0.62	0.87	0.74
(2000)	Comp.	-0.25	-0.55	-0.59	-0.80				
Henneberry et	Uncomp.			0.84~-1.65	(-0.04)~-2.10			(0.46) ~2.24	
al. (1999)	Comp.			(0.15)~-1.50	(0.06) ~-1.47				0.50~5.22
You et al. (1996)	Uncomp.	-0.35	(-0.14)	(-0.03)	-0.40	(0.34)	(0.27)	(0.29)	(0.11)
Huang (1993)	Uncomp	(-0.30)		(-0.13)	-0.20	0.43		0.41	(-0.38)
Cox and Wohlgenant (1986)	Uncomp	-0.20 (canned), -0.67 (frozen)		(-0.20)		-0.08(canned), 0.20 (frozen)		0.07	

Note: "Uncomp. or comp. price elast." stands for uncompensated or compensated price elasticities. PF denotes processed fruits. PV denotes processed vegetables. PFV denotes processed fruits and vegetables. The numbers in the parentheses are not significant.

Table 1.2: Own-Price and Expenditure Demand Elasticities 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	Uncomp. or Comp. Price Elast.	High: Own-Price Elasticities		High: Expenditure/Income Elasticities		Low: Own-Price Elasticities		Low: Expenditure/Income Elasticities	
								Veg	Fruits	Veg	Fruits	Veg	Fruits	Veg	Fruits
Dong and Lin (2009)	Nielsen Homescan Data	Micro	Weekly	2004	2	All goods	Uncomp.	-0.57	-0.58			-0.69	-0.52		
Huang and Lin (2000)	NFCS	Micro	Cross-section	1987–1988	8	Food at home	Uncomp.	-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	Uncomp.	-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: Without further distinction, the values are expenditure elasticities. The values before and after "/" are expenditure and income elasticities respectively. "Cond'l on the exp. of" stands for "Conditional on the expenditures of," "Uncomp. or comp. price elast." stands for uncompensated or compensated price elasticities. "Veg" stands for vegetables.

Table 1.3: Variables in the Model and Sample Statistics

Variable	High-Income Group (N=12,108) Mean	Std. Dev.	Min	Max	Low-Income Group (N=4,722) 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
African American***	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 Gender***								
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***	0.680	1.030	0.000	10.000	0.957	1.379	0.000	9.000
Over 64***	0.300	0.629	0.000	3.000	0.471	0.683	0.000	4.000
log Income ($)***	10.947	0.523	9.705	12.543	9.525	0.793	0.000	11.371
log Price***								
PFV	4.061	0.368	3.746	4.941	4.023	0.350	3.746	4.940
Fresh Vegetables	4.035	0.452	3.539	5.115	4.005	0.450	3.539	5.112
Fresh Fruits	3.928	0.448	3.432	4.997	3.889	0.443	3.432	4.978
Weekly Expenditure ($)***								
PFV	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								
PFV***	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, significant at 5% level and 1% level, respectively. PFV denotes processed fruits and vegetables.

Table 1.4: 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 1.5: Structural Parameter Estimates

Category n	High					
	PFV 1		Fresh Vegetables 2		Fresh Fruits 3	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.044	0.059	0.285***	0.055	0.184***	0.055
Urban	-0.008	0.008	0.014**	0.007	-0.004	0.007
QUAR1	-0.016***	0.005	-0.006	0.005	0.026***	0.005
QUAR2	-0.046***	0.005	-0.014***	0.005	0.059***	0.005
QUAR3	-0.058***	0.005	-0.003	0.005	0.066***	0.005
No Degree	0.012	0.008	0.006	0.007	-0.014**	0.006
High	0.016***	0.005	-0.004	0.005	-0.003	0.005
College	0.004	0.005	0.002	0.004	4.37E-04	0.004
Other Race	0.003	0.016	0.001	0.014	-0.010	0.015
African American	0.049***	0.007	-0.029***	0.007	-0.023***	0.007
Asian	-0.032***	0.010	0.054***	0.009	-0.023***	0.008
Gender	0.012***	0.004	-0.004	0.003	-0.011***	0.004
Age	-4.51E-04***	1.75E-04	-6.12E-05	1.65E-04	0.001***	1.66E-04
Northeast	0.003	0.006	-0.011**	0.005	-0.003	0.005
Midwest	0.007	0.005	-0.009*	0.005	0.013***	0.005
South	0.005	0.005	0.003	0.005	-0.001	0.005
Household Size	0.003	0.002	4.48E-04	0.003	-0.011***	0.002
Under 18	0.008***	0.003	-0.013***	0.003	0.010***	0.003
Over 64	0.001	0.004	9.50E-05	0.004	0.004	0.004
γ_n1	0.359***	0.004	-0.180***	0.003	-0.179***	0.003
γ_n2	-0.180***	0.003	0.330***	0.004	-0.150***	0.003
γ_n3	-0.179***	0.003	-0.150***	0.003	0.329***	0.004
θ_nt						
θ_n1	0.004***	0.001	0.007***	0.001	0.004***	0.001
θ_n2	0.005***	0.001	0.007***	0.001	0.004***	0.001
β_n	0.029***	0.003	0.015***	0.003	0.013***	0.003
σ_n	5.767***	0.453	2.593***	0.217	2.335***	0.211
δ_1	-0.005***	0.001		δ_2	-0.003***	0.001
Heteroscedastic Part in Variance						
QUAR2	0.017	0.030	-0.012	0.031	0.112***	0.033
QUAR3	0.104***	0.032	-0.017	0.031	0.128***	0.031
No Degree	-0.038	0.046	-0.066	0.048	-0.031	0.048
High	-0.048*	0.028	-0.047	0.029	-0.102***	0.031
Midwest	-0.112***	0.033	-0.162***	0.034	-0.105***	0.033
South	-0.057*	0.029	-0.096***	0.030	-0.020	0.029
Over 64	-0.077***	0.019	-0.090***	0.019	-0.054***	0.020
logp (n)	-1.690***	0.038	-1.312***	0.035	-1.310***	0.038
logy-logP (n)	-0.605***	0.017	-0.566***	0.020	-0.544***	0.018

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

Table 1.5 (Continued)

Category n	Low					
	PFV1		Fresh Vegetables 2		Fresh Fruits 3	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.099	0.072	0.283***	0.071	0.167***	0.066
Urban	1.85E-04	0.011	0.013	0.009	-0.007	0.010
QUAR1	-0.024***	0.009	-0.002	0.008	0.027***	0.007
QUAR2	-0.041***	0.010	0.009	0.008	0.032***	0.008
QUAR3	-0.044***	0.009	0.005	0.009	0.041***	0.008
No Degree	0.004	0.011	0.009	0.010	-0.004	0.010
High	0.002	0.010	0.004	0.010	-0.002	0.010
College	-0.002	0.011	0.001	0.011	-0.001	0.011
Other Race	-0.056*	0.032	0.042*	0.022	-0.002	0.022
African American	0.038***	0.009	-0.029***	0.008	-0.017**	0.008
Asian	-0.101***	0.020	0.082***	0.017	-0.010	0.014
Gender	0.013*	0.007	-0.005	0.006	-0.006	0.006
Age	-3.41E-04	2.73E-04	3.78E-04	2.33E-04	-2.51E-05	2.35E-04
Northeast	0.008	0.011	-0.017**	0.009	-0.011	0.009
Midwest	0.002	0.010	-0.020**	0.010	0.025***	0.009
South	-0.006	0.009	-0.011	0.008	0.015*	0.008
Household Size	-0.001	0.004	0.008**	0.004	-0.016***	0.004
Under 18	0.006	0.005	-0.018***	0.005	0.013***	0.005
Over 64	-0.004	0.007	-0.006	0.007	0.010	0.006
γ_n1	0.360***	0.007	-0.181***	0.005	-0.180***	0.005
γ_n2	-0.181***	0.005	0.339***	0.006	-0.158***	0.004
γ_n3	-0.180***	0.005	-0.158***	0.004	0.338***	0.006
θ_nt						
θ_n1	0.003	0.002	0.008***	0.002	0.005**	0.002
θ_n2	0.006***	0.002	0.006***	0.002	0.006***	0.002
β_n	0.053***	0.006	0.013***	0.005	0.003	0.005
σ_n	5.502***	0.747	3.065***	0.468	1.974***	0.263
δ_1	δ_1	-0.005**	0.002	δ_2	-1.98E-04	0.002
Heteroscedastic Part in Variance						
QUAR2	0.039	0.049	0.035	0.052	0.100*	0.052
QUAR3	0.059	0.048	-0.024	0.054	0.193***	0.051
No Degree	-0.087*	0.050	-0.015	0.058	-0.104**	0.049
High	-0.097**	0.049	-0.106*	0.058	-0.109**	0.053
Midwest	-0.035	0.055	-0.084	0.077	-0.076	0.053
South	-0.011	0.047	-0.102**	0.047	-0.081	0.050
Over 64	-0.029	0.031	-0.065**	0.030	-0.031	0.033
logp (n)	-1.733***	0.066	-1.424***	0.063	-1.284***	0.059
logy-logP (n)	-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 parentheses are standard errors.

Table 1.6: Structural Parameter Estimates (% Change)

Category	High			Low		
	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%	-4.60%**	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%
African American	13.04%***	-9.25%***	-7.37%***	9.95%***	-9.39%***	-5.56%**
Asian	-8.46%***	17.19%***	-7.34%***	-26.50%***	26.07%***	-3.28%
Gender	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%	2.60%**	-5.20%***
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 1.7: Conditional Uncompensated Price Elasticities

	High			Low		
	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***	-0.120***
	(0.004)	(0.116)	(0.002)	(0.005)	(0.190)	(0.004)
Fresh Fruits	-0.130***	-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 parentheses are standard errors.

Table 1.8: Conditional Compensated Price Elasticities

	High			Low		
	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***	-2.810***	0.240***
	(0.017)	(0.137)	(0.020)	(0.031)	(0.225)	(0.027)
Fresh Fruits	0.073***	0.369***	-0.364***	0.150***	0.310***	-0.501***
	(0.016)	(0.023)	(0.042)	(0.027)	(0.030)	(0.064)

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

Table 1.9: Conditional 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***	0.810***
	(0.025)	(0.010)	(0.013)	(0.029)	(0.022)	(0.028)

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

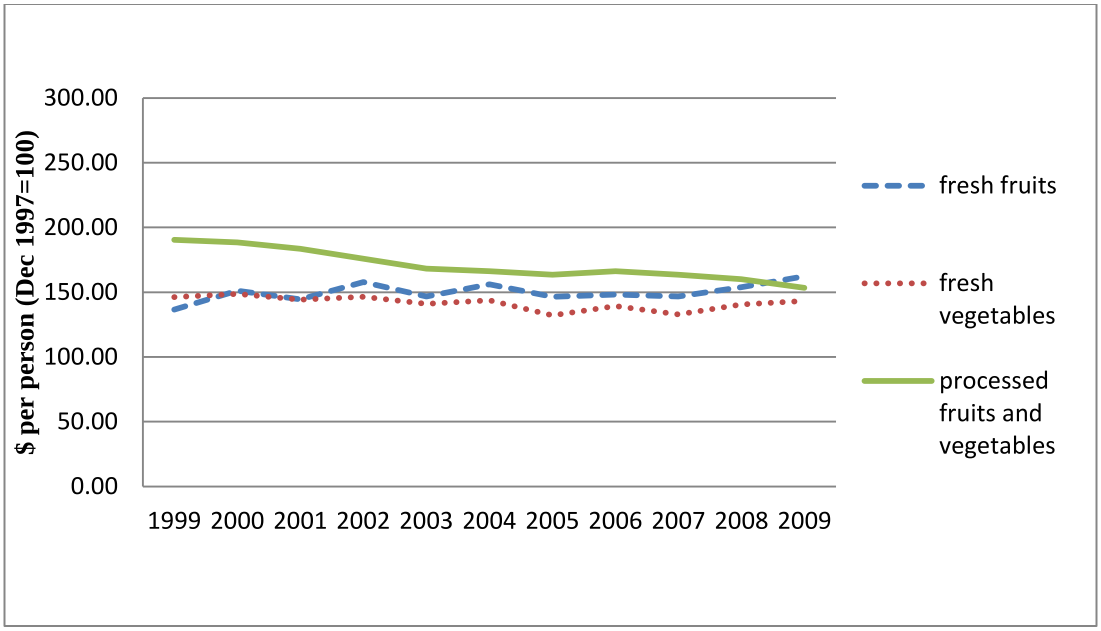

Figure 1.1: Per Capita Real Dollar Expenditures for Fruits and Vegetables, U.S., 1999–2009
Source: Consumer Expenditure Survey, Bureau Labor Statistics

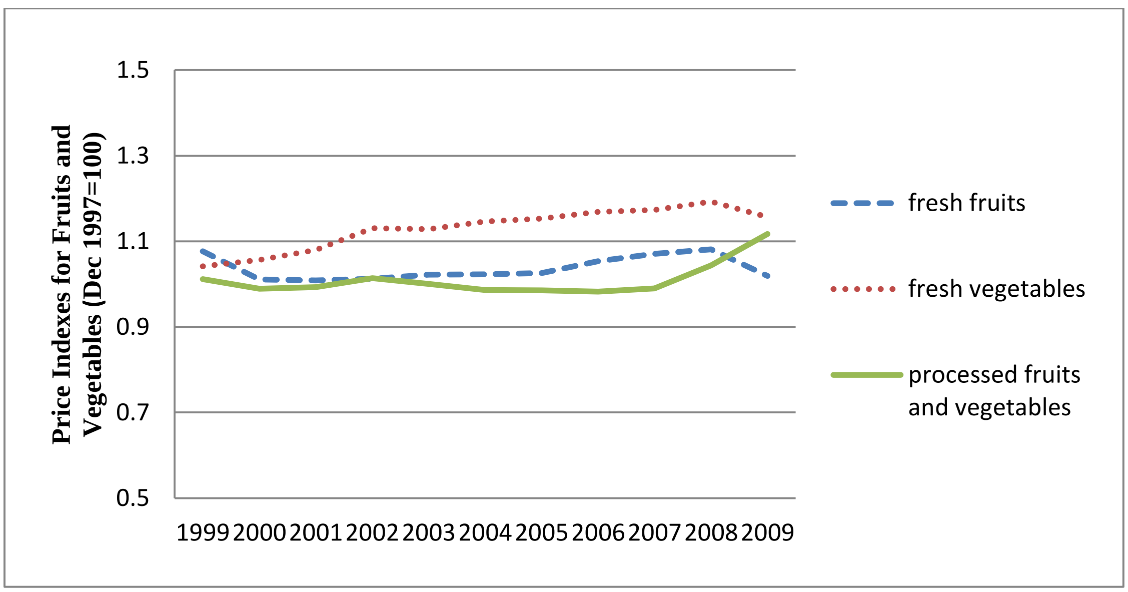

Figure 1.2: Price Indexes for Fruits and Vegetables (Dec 1997=100), U.S., 1999–2009
Source: Consumer Expenditure Survey, Bureau Labor Statistics

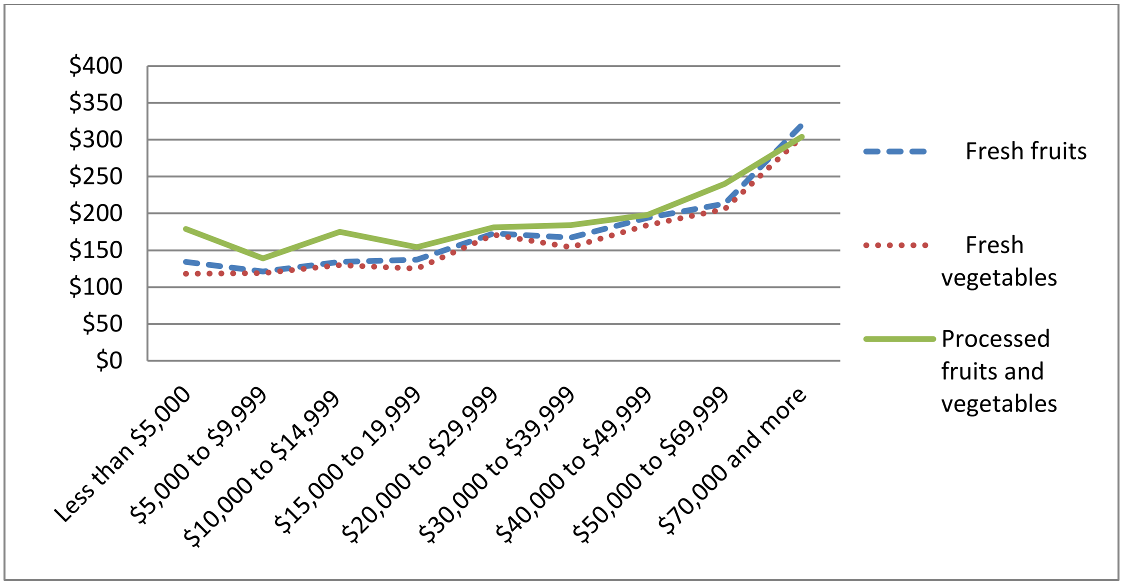

Figure 1.3: Average Annual Expenditures on Fruits and Vegetables, U.S., 2009
Source: Consumer Expenditure Survey, Bureau Labor Statistics

1.8 References

Abowd, J.M., and D. Card. 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.

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

Asche, F., and C.R. Wessells. 1997. "On Price Indexes in the Almost Ideal Demand System." *American Journal of Agricultural Economics* 79:1182-1185.

Barten, A.P. 1993. "Consumer Allocation Models: Choice of Functional Form." *Empirical Economics* 18:129-158.

Basiotis, P., A.C. Gerrior, W. Juan, and M. Lino. 2002. *The Healthy Eating Index: 1999–2000*. CNPP-12, U.S. Department of Agriculture, 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

Christenson, L.R., D.W. Jorgenson, and L.J. Lau. 1975. "Transcendental Logarithmic Utility Functions." *American Economic Review* 65:367-382.

Cox, T.L., and M.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. Available at: http://www.cnpp.usda.gov/Publications/DietaryGuidelines/2010/PolicyDoc/PolicyDoc.pdf

Dong, D., and B. Lin. 2009. *Fruit and Vegetable Consumption by Low-Income Americans: Would a Price Make a Difference?* Economic Research Report 70. U.S. Department of Agriculture, Economic Research Service.

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 Agricultural & Applied Economics Association's Annual Meeting, Tampa, FL, July 30-August 2, 2000.

Green, R., and J.M. Alston. 1991. "Elasticities in AIDS Models: A Clarification and Extension." *American Journal of Agricultural Economics* 73:874-875.

Heien, D., and C.R. Wessells. 1990. "Demand Systems Esimation with Microdata: A Censored Regression Approach." *Journal of Business & Economic Statistics* 8:365-371.

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, K.S. 1993. *A complete System of U.S. Demand for Food.* Technical Bulletin 1821. U.S. Department of Agriculture, Economic Research Service.

Huang, K.S., and B. Lin. 2000. *Estimation of Food Demand and Nutrient Elasticities from Household Survey Data.* Technical Bulletin 1887. U.S. Department of Agriculture, Economic Research Service.

Jakubson, G. 1988. "The Sensitivity of Labor-Supply Parameter Estimates to Unobserved Individual 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.

Klein, L.R., and H. Rubin. 1948. "A Constant Utility Index of the Cost of Living." *The Review of Economic Studies* 15:84-87.

Lewbel, A. 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*. Statistical Bulletin 849, U.S. Department of Agriculture, 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.” Selected Paper presented at the Southern Agricultural Economics Association’s Annual Meeting, Memphis, TN, February 1-3, 1999.

McCracken, V.A. 1992. “The U.S. Demand for Vegetables.” In R.A. Lopez, and L.C. Polopolus, eds. *Vegetable Markets in the Western Hemisphere: Trends, Policies, and Trade Linkages.* Ames, IA: Iowa State University Press. pp. 20-43.

Meyerhoefer, C.D. 2002. *The Econometrics of Non-Standard Structural Demand Modeling: Applications to Transition Country Data*. Ph.D. Dissertation. Cornell University, Ithaca NY.

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.

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.M. 1995a. “Microdata Expenditure Analysis of Disaggregate Meat Products.” *Review of Agricultural Econonomics* 17:275-285.

Nayga Jr., R.M. 1995b. “Determinants of U.S. Household Expenditures on Fruit and Vegetables: a Note and Update.” *Journal of Agricultural & Applied Economics* 27:588-594.

Okrent, A.M., and J.M. Alston. 2011. *Demand for Food in the United States: A Review of the Literature, Evaluation of Previous Estimates, and Presentation of New Estimates of Demand.* Berkeley, CA: Giannini Foundation of Agricultural Economics Monograph 48.

Okrent, A.M., and J.M. Alston. 2012. *Demand for Disaggregated Food-Away-from-Home and Food-at-Home Products in the United States*. Economic Research Report 139. U.S. Department of Agriculture, Economic Research Service.

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.

Pollack, 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., N. Blisard, and D. Jolliffe. 2003. “Do Income Constraints Inhibit Spending on Fruits and Vegetables Among Low-income Households.” *Journal of Agricultural and Resource Economics* 28:465-480.

Yen, S.T., and B. Lin. 2006. “A Sample Selection Approach to Censored Demand Systems.”*American Journal of Agricultural Economics* 88: 742-749.

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.M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

Chapter 2 Efficient Estimation of System of Demand Functions with Pseudo-panel Data: Application to Demand for Fruits and Vegetables

2.1 Introduction and Literature Overview

When we need to account for individual specific heterogeneity in the model, traditionally there are two approaches: "random effects (RE)" and "fixed effects (FE)." As mentioned in the first chapter, the RE model requires explanatory variables to be uncorrelated with unobservable individual specific effects. However, this is a relatively strong assumption that is not easily satisfied in empirical work. In contrast, the FE model relaxes this assumption and eliminates individual heterogeneity by FE transformation or first-differencing transformation. The FE model is usually applied on datasets where the same groups of individuals are traced over time, i.e., panel data. However, in reality, many datasets do not have panel features but are independent cross-sections. Examples include the British Family Expenditure Survey (BFES), India's National Sample Survey (INSS), the National Health Interview Survey (NHIS), and the National Crime Survey (NCS). Although the Consumer Expenditure Survey (CEX) used in this study is a short panel, there are only two time periods so time-constant demographic variables would not play a role in the estimation after applying FE transformation.

Deaton (1985) proposes a solution to overcome this identification problem by using a time series of repeated cross-sections. He suggests constructing a "pseudo-panel" ("synthetic panel") with a number of cohorts (or groups), where a cohort is defined as "a group with fixed membership, individuals of which can be identified as they show up in the surveys." [35] For instance, region cohorts are those consisting of all individuals in the sample who live in

[35] In this study, "cohort" and "group" are used interchangeably.

the same region of the United States, such as the South and the West. These cohorts can be tracked over time instead of the individuals within the cohorts. Other advantages to using the pseudo-panel are that it can avoid the attrition problem usually encountered in the genuine panel and can guarantee the representativeness is constantly maintained; it can also eliminate the difficulties dealing with censoring encountered in Chapter One.

Specifically, Deaton calculates the sample averages for each cohort and treats them as observations in the pseudo-panel. He argues that if it were possible to observe the cohort population, then the FE estimator would be consistent based on the population means. Nevertheless, sample means are error-ridden measurements of the true means. Deaton corrects the FE estimator for measurements error and suggests an errors-in-variables estimator. The consistency of the latter estimator requires the time dimension to approach infinity. However, it is inconsistent for small number of time periods even in the condition of large number of cohorts (Verbeek and Nijman, 1993).

Following Deaton, Verbeek and Nijman (1993) make an adjustment to the above errors-in-variables estimator and attain consistency as the number of cohorts tends to infinity given fixed time periods and the number of observations per cohort. They argue that the adjusted estimator has a smaller mean squared error than that proposed by Deaton.

Collado (1997) extends Deaton's work to a linear dynamic model. She is also concerned with the errors-in-variables problem and proposes a corrected FE estimator that is consistent as time tends to infinity. Furthermore, she uses a corrected Generalized Method of Moments (GMM) estimator and derives the asymptotic distribution when the number of cohorts approaches infinity for fixed time periods and cohort sizes. By conducting Monte Carlo simulations, she concludes that measurement error correction is not trivial to reduce the bias in the estimator.

In addition to deriving the asymptotics in the dimension of time and the number of cohorts, there are studies discussing the asymptotic property when the number of individuals per cohort—that is, cohort size—becomes large. If the cohort size is very large, the difference between sample averages and population means of the cohorts is small and the errors-in-variable estimator is asymptotically the conventional FE estimator. In empirical work, the measurement error is often ignored and the FE estimator is applied (see Browning, Deaton, and Irish, 1985; Blundell, Meghir, and Neves, 1993).

A trade-off is obvious when the groups are constructed. On one hand, if the number of cohorts is larger, then the cohort size is smaller; on the other, if the number of cohorts is smaller, then the cohort size is larger. Because the measurement error becomes small as the number of individuals per cohort goes up, more cohorts contain less precise observations while fewer cohorts consist of more precise observations. Verbeek and Nijman (1992) discuss the conditions of using an FE estimator when ignoring the measurement error in the static model and the impact of construction of cohorts on the bias introduced to the estimation. They conclude that bias is small when cohort size is large and when there is large time variation displayed in the cohort population means. At least one hundred individuals per cohort are suggested in the empirical work.

Several studies also work on dynamic models. Moffitt (1993) generalizes Deaton's discussion in the way that Instrumental Variables (IV) rather than grouping are used in the analysis. Durbin (1954) was the first to show that the procedures of taking group averages and IV estimation using group dummies are equivalent. This idea is further generalized by Moffitt (also see Griliches, 1986 and Angrist, 1991). Moffitt concentrates on a dynamic model and proposes the Two-Stage Least Squares (2SLS) estimator, where the unobserved lagged dependent variable is instrumented by a predicted value attained as the number of observations per cohort converges to infinity. However, Moffitt's estimator is inconsistent if exogenous variables are not time-varying, or there is no autocorrelation in the time-varying exogenous variables (Verbeek and Vella, 2005).

Girma (2000) obtains a consistent GMM estimator using a quasi-differenced framework in the sense that quasi-differencing has been taken across two individuals from the same cohort. This estimator allows heterogeneous parameters over different cohorts. Verbeek and Vella (2005) review the identification conditions of the above Moffitt, Collado, and Girma's estimators and propose an augmented IV estimator that is easier to compute while allowing time-invariant cohort effect to be included in the model. Mckenzie (2004) works on the asymptotic property in multidimensional limits by allowing parameters to vary by cohort. Based on appropriate assumptions, FE and IV estimators are found to yield consistency as both cohort sizes and time periods tends to infinity. The empirical example corroborates the existence of heterogeneity in the consumption growth rate across groups.

Inoue (2008) studies efficient estimation when the group FE is included in both static and dynamic models. When the group FE is in the model, he shows that the Ordinary Least Squares (OLS estimator is inconsistent and, although the FE estimator is consistent, the conventional t statistic does not have an asymptotically normal distribution, which generates an invalid inference. He proposes a robust t statistic that has an asymptotic standard normal distribution. Furthermore, he proposes an efficient GMM for which the orthogonal conditions result from grouping. The GMM overidentification test can be used to test the appropriateness of grouping.

Based on Inoue's theory, this study estimates conditional demand for fruits and vegetables in the United States using the Diary Survey of CEX from 1996 to 2010. To my knowledge, this is the first study that uses a pseudo-panel to evaluate consumption of fruits and vegetables. The group specific effect is accounted for by the economic model, which was often sidestepped by past empirical pseudo-panel analysis. Moreover, Inoue's discussions on single-equation pseudo demand are extended into a demand system where cross-equation restrictions can be imposed according to the economic theory. The demand system is estimated in two steps. In the first step, a consistent FE estimator is obtained to derive a consistent estimator of covariance matrix of error terms. In the second step, an efficient GMM estimator is estimated and a robust t statistic is derived to make valid inferences.

This chapter is organized as follows. The next section describes the CEX dataset and how the pseudo-panel data are constructed. The third section presents the economic model and the fourth section discusses the methodology used to estimate the model. Results are given in the fifth section and the final section concludes the study.

2.2 Data

This chapter uses 1996–2010 CEX Dairy Survey data, where data files include households' (or CUs') characteristics and detailed weekly expenditures on small and frequently purchased items such as food. According to the literature, the following demographic variables are included in the model: household head's urban status, educational attainment and race, region, household size, number of children under 18 in a household, and number of persons over 64 in a household. The variable region is blank for households residing in rural areas in CEX data before 1996, so only data after 1996 are used.

Data are collected over two consecutive one-week periods. To make the later derivation easy to handle, the original balanced panel data are converted to independently and identically distributed (i.i.d.) cross-sectional data. Specifically, the two observations that belong to the same household are merged to become one observation. For demographic and socio-economic variables, only values of one of the observations are used; for expenditures on fruits and vegetables, the two observations are summed to provide biweekly expenditures for each household.[36] Then, the biweekly expenditures are multiply by one-half of the number of weeks in a quarter to become quarterly data. Because no price data are available in the CEX, Stone-Lewbel price indexes are constructed based on the budget shares of subcategories of fruits and vegetables, and the quarterly Consumer Price Index (CPI) derived from monthly CPI on Bureau of Labor Statistics (BLS)'s Web site. The construction of Stone-Lewbel price indexes can be found in Chapter One. In total, there are 87,063 observations.

As mentioned in the introduction, a pseudo-panel is constructed from the mean of the grouped cross-sectional observations. The groups, which are selected according to some categorical variables, can be tracked over time although individuals within the groups cannot. The categorical variable, which is used as IV, can be written in terms of the indicator function, denoted as $1[\cdot]$. This function is unity if a set of households belong to the group g at

[36] Most values of demographic variables are not changed over two weeks' periods.

time t (i.e. $1[j \in S_{gt}]=1$), and zero otherwise, where S_{gt} denotes the set of households in the group g at time t. The total number of households is $\sum_{g=1}^{G}\sum_{t=1}^{T} J_{gt} = J$, where J_{gt} is the number of households in group g at time t.[37] So the categorical variable is assumed to be uncorrelated with the error terms in the model.

First, data are grouped by quarter, and we have 60 quarters (time groups) in total for a fifteen-year dataset. Then households in each time group are further divided by income. According to the 1996–2010 Federal Poverty Guidelines, data are grouped into five income groups: Households living below 130% of the poverty level, between 130% and 185% of the poverty level, between 185% and 350% of the poverty level, between 350% and 500% of the poverty level, and above 500% of the poverty level.[38] Note that poverty guidelines are varied by both year and household size. Because each income group is traced over sixty time periods, there are 300 (60 by 5) observations in total in the constructed pseudo-panel dataset. The average cohort size is 290.21, which satisfies the standard of "at least one hundred individuals per cohort" suggested by Verbeek and Nijman (1992). Sample statistics and group size distribution are shown in Table 2.1 and Table 2.2, respectively. From Table 2.2, number of households living below 130% of the poverty level is the largest, while the number of households living between 130% and 185% of the poverty level is the smallest. The definition of each variable refers to Appendix A.

2.3 Model

In contrast to Chapter One, Chapter Two assumes both low-income and high-income households are facing one demand model, meaning the coefficients are same for both income group households. As in Chapter One, the Linear Approximate Almost Ideal Demand System (LA/AIDS) model is used, where the original Stone price index, $\log P^S \equiv \sum_{i=1}^{n} w_{it} \log(p_{it})$, is replaced by the loglinear analogue of the Laspeyres price index, i.e.,

[37] The IV interpretation is based on the moment conditions from equation (4) in Inoue (2008).

[38] These cutoffs are chosen based on the eligibility criteria for food assistance programs such as Women, Infants & Children (WIC) Program, National School Lunch Program, and the relevant literature.

$\log P \equiv \sum_{i=1}^{n} w_i^0 \log(p_{it})$. There are two reasons for this replacement. First, this new price index is invariant to units of measurement or has Closed Under Unit Scaling (CUUS) property. Second, it can avoid the endogeneity problem introduced by the expenditure shares in the expression of the Stone price index. The model is written as

$$\text{(2.1)} \quad \begin{aligned} w_{nj} = {} & \alpha_n + \mu_n T_t + \sum_l \eta_{nl} a_{lj} + \gamma_{n1} \log p_{1j} + \gamma_{n2} \log p_{2j} + \gamma_{n3} \log p_{3j} \\ & + \beta_n (\log y_j^{FV} - \log P_j^{FV}) + \delta_g + c_j + \varepsilon_{nj} \end{aligned}$$

with $\log P_j^{FV} = w_1^o \log p_{1j} + w_2^o \log p_{2j} + w_3^o \log p_{3j}$, where w_n^o represents the mean expenditure shares of good n, and n=1, 2, 3, representing processed fruits and vegetables, fresh vegetables, and fresh fruits, for household j =1,…, J; T_t is the vector of dummies for quarter t; $a_j = (a_{lj}, ..., a_{Lj})$ denotes demographic variables for household j; $\log p_{nj}$ is logarithm commodity prices for good n and household j; y_j^{FV} is the expenditure on fruits and vegetables for household j; $\log p_j^{FV}$ is the “corrected” Stone price index for fruits and vegetables; δ_g is the group specific effect for group g =1,…, G, c_j is the household specific effect, and ε is the error term. These three variables are unobservable. Assume c_j and ε_{nj} are independent and let $u_{nj} \equiv c_j + \varepsilon_{nj}$. By taking the group time averages for equation (2.1), the pseudo-panel model are derived as

$$\text{(2.2)} \quad \begin{aligned} w_{n\_gt} = {} & \alpha_n + \mu_n T_t + \sum_l \eta_{nl} a_{l\_gt} + \gamma_{n1} \log p_{1\_gt} + \gamma_{n2} \log p_{2\_gt} + \gamma_{n3} \log p_{3\_gt} \\ & + \beta_n (\log y_{gt}^{FV} - \log P_{gt}^{FV}) + \delta_g + c_{gt} + \varepsilon_{n\_gt}, \end{aligned}$$

where $w_{n\_gt} = 1/J_{gt} \sum_{j=1}^{J_{gt}} w_{nj}$, $a_{l\_gt} = 1/J_{gt} \sum_{j=1}^{J_{gt}} a_{lj}$, $\log p_{k\_gt} = 1/J_{gt} \sum_{j=1}^{J_{gt}} \log p_{kj}$ for k=1, 2 and 3, $\log y_{gt}^{FV} = 1/J_{gt} \sum_{j=1}^{J_{gt}} \log y_j^{FV}$, $\log P_{gt}^{FV} = 1/J_{gt} \sum_{j=1}^{J_{gt}} \log P_j^{FV}$, $c_{gt} = 1/J_{gt} \sum_{j=1}^{J_{gt}} c_j$, $\varepsilon_{n\_gt} = 1/J_{gt} \sum_{j=1}^{J_{gt}} \varepsilon_{nj}$ in which J_{gt} is the number of households in the group g (g=1,…, G) at time t(t=1,…, T). So the total number of households is $\sum_g \sum_t J_{gt} = J$. Note α_n, T_t, and δ_g remain unchanged.

To simplify equation (2.2), let X_{gt} be the 1 by k vector of all the RHS exogenous variables of equation (2.2); let u_{n_gt} be the error term, which is equal to $c_{gt} + \varepsilon_{n\_gt} = 1/J_{gt} \sum_{j=1}^{J_{gt}} (c_j + \varepsilon_{nj}) = 1/J_{gt} \sum_{j=1}^{J_{gt}} u_{nj}$, and let ϕ_n be the k by 1 vector, representing the corresponding coefficients associated with X_{gt}. So we obtain

(2.3) $$w_{n\_gt} = \alpha_n + X_{gt}\phi_n + \delta_g + u_{n\_gt}.$$

Combining all GT equations for good n, equation (2.3) becomes

(2.4) $$W_n = \alpha_n \ell + X\phi_n + \delta_g \ell + U_n,$$

where $W_n = [w_{n\_11}\ w_{n\_12}\ \cdots\ w_{n\_GT}]'$, $X = [x_{11}'\ x_{12}'\ \cdots\ x_{GT}']'$, $U_n = [u_{n\_11}\ u_{n\_12}\ \cdots\ u_{n\_GT}]'$, and ℓ is a column vector of ones. To remove the group specific effect δ_g, the fixed effects transformation is conducted by pre-multiplying the equation (2.4) by the time-demeaning matrix Q, defined as $Q = I_G \otimes (I_T - (T^{-1}\ell_T \ell_T')$, which is a GT by GT symmetric and idempotent matrix with rank GT-G. Thus, we obtain the demeaned equation for good n as

(2.5) $$QW_n = QX\phi_n + QU_n.$$

We see that α_n and δ_g have been removed. According to the economic theory, adding-up, homogeneity and symmetry are imposed on the parameters. Due to the adding-up restrictions, the covariance matrix of the errors terms is singular. One solution to this problem is to delete demand equations for an arbitrary good (Barten, 1969). Without loss of generality, the demand equation for fresh fruits is dropped from the model and the parameter estimates for fresh fruits are recovered through the above economic restrictions. So when equation (2.5) is stacked for two goods, it becomes

(2.6) $$\bar{W} = \bar{X}\phi + \bar{U},$$

where $\bar{W}$ and $\bar{U}$ are simply 2GT by 1 vectors stacking QW_n and QU_n for n=1 and 2; however, after imposing homogeneity and symmetry restrictions, $\bar{X}$ is the 2GT by $\tilde{k}$ matrix, where variables having the same coefficients are combined, and ϕ is the $\tilde{k}$ by 1 vector, representing $\bar{X}$'s corresponding coefficients. Note that $\tilde{k}$, the total number of parameters of the first two demand equations, is less than 2k due to the cross-equation restrictions. Because of the fixed effect transformation, G observations are redundant for equation (2.6), which should be deleted for each good n. One method is to delete one time period in each group and the results are invariant to the time period chosen to be deleted (Im et al. 1999). Without loss of generality, the demand system (2.6) drops the last time period for each group g. Specifically, the Tth, 2Tth,…, GTth, (G+1)Tth,…, 2GTth rows are deleted from $\bar{W}$, $\bar{X}$ and $\bar{U}$, respectively, which gives

$$\breve{W} = \breve{X}\phi + \breve{U}, \tag{2.7}$$

where $\breve{W}$ and $\breve{U}$ is a 2G(T-1) by 1 matrix, and $\breve{X}$ is a 2G(T-1) by $\tilde{k}$ matrix. The equation (2.7) is the estimating equation. The asymptotic variance of $\sqrt{J}\breve{U}$ is denoted as $\breve{\Sigma}$.

2.4 Methodology

Define $\bar{n}$ as the maximum of n; that is, n=1, 2,…, $\bar{n}$. In this context, there are three goods, so $\bar{n}$=3. The assumptions for consistency and asymptotic inference are as follows:

a. $\{(c_j, \varepsilon_{nj})$: j=1,…, J$\}$ is i.i.d. with finite fourth moments for n=1,…, $\bar{n}$.

b. $E(c_j | 1[j \in S_{gt}]) = 0$ and $E(\varepsilon_{nj} | 1[j \in S_{gt}]) = 0$ for g=1,…, G, t=1,…, T, and n=1,…, $\bar{n}$.

c. $Var(c_j + \varepsilon_{nj} | 1[j \in S_{gt}]) = Var(u_{nj} | 1[j \in S_{gt}]) = \sigma^2$ for g=1,…, G, t=1,…, T, and n=1,…, $\bar{n}$.

d. $J_{gt}/J \to r_{gt}$ as J→∞, where $r_{gt} \in (0,1)$ for g=1,…, G, and t=1,…, T.

e. $X \xrightarrow{p} A_X$ as $J_{gt}\to\infty$ for g=1,…, G, and t=1,…, T.

f. $\bar{A}_X{}'\bar{A}_X$ and $\breve{A}_X{}'\breve{\Sigma}^{-1}\breve{A}_X$ are nonsingular, where $\bar{A}_x$ is derived by stacking A_X for ($\bar{n}$-1) goods, $\breve{A}_x$ is derived by deleting the Tth, 2Tth…, 2GTth rows from $\bar{A}_x$, and $\breve{\Sigma}^{-1}$ is defined below in the text.

Here are some remarks about the above assumptions. Assumption b is the strict exogeneity assumption, which requires that the group selection variable $1[j \in S_{gt}]$ is strictly exogenous to the error term $c_j + \varepsilon_{nj}$. Note that the explanatory variables can be endogenous.

Assumption c requires that the conditional variance of the error term is constant. It is noteworthy that $\text{var}(u_{nj} | 1[j \in S_{gt}])$ may vary across groups and times. However, the White test does not reject the errors that are homoscedastic in this context. Assumption e guarantees that X converges in probability to a matrix A_X as the number of households in each group goes to infinity. Assumption f is the rank condition for identification. Assumption b and f together are considered as the conditions to ensure the validity of the group selection variable as an instrument.

The model is estimated in two stages. In the first stage, a consistent FE estimator is estimated. In the second stage, by using the estimates from the first stage, an efficient GMM estimator is obtained. In particular, the consistent FE estimator has the following form

(2.8) $$\hat{\phi}_{FE} = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{W},$$

and the efficient GMM estimator is given by

(2.9) $$\hat{\phi}_{GMM} = (\breve{X}'\hat{\breve{\Sigma}}^{-1}\breve{X})^{-1}\breve{X}'\hat{\breve{\Sigma}}^{-1}\breve{W},$$

where $\hat{\breve{\Sigma}}$ is a consistent estimator of $\breve{\Sigma}$. The following shows how $\hat{\breve{\Sigma}}$ is derived. From (2.6), we know that

(2.10) $$\sqrt{J}\bar{U} = \sqrt{J}\begin{pmatrix} QU_1 \\ QU_2 \end{pmatrix} = \sqrt{J}\begin{pmatrix} Q & 0 \\ 0 & Q \end{pmatrix}\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \sqrt{J}\bar{Q}\begin{pmatrix} u_{1\_11} \\ \vdots \\ u_{1\_GT} \\ u_{2\_11} \\ \vdots \\ u_{2\_GT} \end{pmatrix} = \sqrt{J}\bar{Q}\begin{pmatrix} \frac{1}{\sqrt{J_{11}}}\sqrt{J_{11}}\frac{1}{J_{11}}\sum_{j=1}^{J_{gt}} u_{1j} \\ \vdots \\ \frac{1}{\sqrt{J_{GT}}}\sqrt{J_{GT}}\frac{1}{J_{GT}}\sum_{j=1}^{J_{gt}} u_{1j} \\ \frac{1}{\sqrt{J_{11}}}\sqrt{J_{11}}\frac{1}{J_{11}}\sum_{j=1}^{J_{gt}} u_{2j} \\ \vdots \\ \frac{1}{\sqrt{J_{GT}}}\sqrt{J_{GT}}\frac{1}{J_{GT}}\sum_{j=1}^{J_{gt}} u_{2j} \end{pmatrix}$$

$$\xrightarrow{d} N(0,\ \bar{Q}\Xi\bar{Q}) \ \equiv\ N(0, \bar{\Sigma}),$$

where $\bar{Q} = I \otimes Q$ is a 2GT by 2GT block-diagonal matrix, and $\bar{\Xi} = \bar{R}^{-1}\bar{\Omega}$, in which $\hat{\bar{R}}$ is a 2GT by 2GT block-diagonal matrix with diagonal entries $J_{11}/J, \ldots, J_{GT}/J, J_{11}/J, \ldots, J_{GT}/J$, and $\bar{\Omega}$ is a 2GT by 2GT matrix having the form of

(2.11)

$$\bar{\Omega} = \begin{pmatrix} Var(u_{1j}|1[j \in S_{11}]) & \cdots & 0 & Cov(u_{1j}, u_{2j}|1[j \in S_{11}]) & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & Var(u_{1j}|1[j \in S_{GT}]) & 0 & \cdots & Cov(u_{1j}, u_{2j}|1[j \in S_{GT}]) \\ Cov(u_{2j}, u_{1j}|1[j \in S_{11}]) & \cdots & 0 & Var(u_{2j}|1[j \in S_{11}]) & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & Cov(u_{2j}, u_{1j}|1[j \in S_{GT}]) & 0 & \cdots & Var(u_{2j}|1[j \in S_{GT}]) \end{pmatrix},$$

where consistent estimators for variance and covariance are $\hat{\sigma}_n^2 \equiv var(u_{nj}|1[j \in S_{gt}]) =$ $(1/G)\sum_{g=1}^{G}[1/J_g \sum_{j=1}^{J_g}\hat{u}_{nj}^2 - (1/J_g \sum_{j=1}^{J_g}\hat{u}_{nj})^2]$, and $\hat{\sigma}_{12}^2 = \hat{\sigma}_{21}^2 \equiv cov(u_{1j}, u_{2j}|1[j \in S_{gt}]) =$ $(1/G)\sum_{g=1}^{G}[1/J_g \sum_{j=1}^{J_g}\hat{u}_{1j}\hat{u}_{2j} - (1/J_g \sum_{j=1}^{J_g}\hat{u}_{1j})(1/J_g \sum_{j=1}^{J_g}\hat{u}_{2j})]$, in which $\hat{u}_{nj}$ is the individual residuals from FE estimation in the first stage for good n=1,2, and household j.[39] Following (2.10)'s notation,

(2.12)
$$\sqrt{J}\ \breve{U} = \sqrt{J}\begin{pmatrix}\breve{U}_1 \\ \breve{U}_2\end{pmatrix} \xrightarrow{d} N(0,\ \breve{\Sigma}),$$

where $\breve{\Sigma}$ is the 2G(T-1) by 2G(T-1) matrix after deleting the Tth, 2Tth, …, 2GTth rows and columns of $\bar{\Sigma}$. The sampling error is known as $\hat{\phi}_{GMM} - \phi = (\breve{X}'\hat{\breve{\Sigma}}^{-1}\breve{X})^{-1}\breve{X}'\hat{\breve{\Sigma}}^{-1}\breve{U}$ from equation (2.9), so combined with equation (2.12), the asymptotic distribution of $\hat{\phi}_{GMM}$ is

(2.13)
$$\sqrt{J}(\hat{\phi}_{GMM} - \phi) \xrightarrow{d} N(0,\ (\breve{A}_x'\breve{\Sigma}^{-1}\breve{A}_x)^{-1})$$

When J converges to infinity, under the null $\hat{\phi}_{GMM,k} = \phi_{0,k}$, the t statistic is

---

39 If $var(u_{nj}|1[j \in S_{gt}])$ varies across groups and time periods, then the elements in $\bar{\Omega}$ can be estimated by $\hat{\sigma}_{n\_gt}^2 \equiv var(u_{nj}|1[j \in S_{gt}]) = 1/J_{gt}\sum_{j=1}^{J_{gt}}\hat{u}_{nj}^2 - (1/J_{gt}\sum_{j=1}^{J_{gt}}\hat{u}_{nj})^2$, and $\hat{\sigma}_{12\_gt}^2 = \hat{\sigma}_{21\_gt}^2 \equiv cov(u_{1j}, u_{2j}|1[j \in S_{gt}]) = 1/J_{gt}\sum_{j=1}^{J_{gt}}\hat{u}_{1j}\hat{u}_{2j} - (1/J_{gt}\sum_{j=1}^{J_{gt}}\hat{u}_{1j})(1/J_{gt}\sum_{j=1}^{J_{gt}}\hat{u}_{2j})$.

$$(2.14) \qquad t_{\hat{\phi}_{GMM,k}} = \frac{\sqrt{J}(\hat{\phi}_{GMM,k} - \phi_{0,k})}{\sqrt{[(\breve{X}'\hat{\hat{\Sigma}}^{-1}\breve{X})^{-1}]_{kk}}} \xrightarrow{d} N(0,1).$$

Based on above specification in this study, the Hansen's (1982) overidentification test statistic is $J\,\hat{\breve{U}}'\hat{\hat{\Sigma}}\,\hat{\breve{U}}' \sim \chi^2_{G(T-1)-\tilde{k}}$. Recall that J is the total number of households in the sample. As Inoue stated, the Hansen's "J test" can be used to test the validity of grouping. In this study, the p-value of test statistic is 0.409, which does not reject the null of overidentifying restrictions in the entire system. Thus, the group selection is valid here.

2.5 Results

The coefficient estimates are reported in Table 2.3. Compared to Chapter One, there are more insignificant estimates. This may be due to the small sample size used in the estimation.[40] Even so, the parameters still shed light on consumption structure.

Seasonal effects are found to have a statistically significant effect on fruit and vegetable consumption. Results indicate that, compared to the fourth quarter, households demand fewer processed fruits and vegetables and fresh vegetables but more fresh fruits in other quarters, which is consistent with what was found in Chapter One. In the second and third quarters, households demand the least processed fruits and vegetables and most fresh fruits.

Some demographic characteristics significantly affect demand for fruits and vegetables. Compared to rural households, urban households demand fewer processed fruits and vegetables and more fresh fruits. Compared to households residing in the West, households residing in the Midwest and South demand fewer processed fruits and vegetables. Also households residing in the South demand more fresh fruits. As in Chapter One, the more children there are in a household, the more processed fruits and vegetables and fewer fresh vegetables are demanded. In contrast, the more seniors there are in a household, the fewer processed fruits and vegetables and more fresh fruits are demanded. However, the household

[40] In this study, there are only 300 demeaned observations, far fewer than the observations used for estimation in Chapter One, in which there are 4,722 low-income households and 12,108 high-income households.

head's educational attainment and race are not found to have a significant effect on households' demand for fruits and vegetables.

The results of estimated conditional price and expenditure elasticities evaluated at the sample mean are presented in Table 2.4, Table 2.5 and Table 2.6. Uncompensated and compensated own-price elasticities are all negative, as expected. Uncompensated cross-price elasticities are all negative, showing all three goods are "gross complements," which is consistent with the results in Chapter One. Furthermore, after accounting for the expenditure effect, compensated cross-price elasticities show that all the fruits and vegetables are found to be "net substitutes" except that processed fruits and vegetables and fresh vegetables are estimated to be "net complements," but are not statistically significant. Based on the matrix of compensated elasticities, all eigenvalues of Slutsky matrix (evaluated at the sample mean) are obtained and found to be negative. This shows that the concavity conditions hold for observed demand behavior. What is more, all three conditional expenditure elasticities are significantly positive, with all estimates around one. Because a conditional demand model is used, where households take expenditures on fruits and vegetables rather than total income constant, conditional expenditure elasticities are expected to be much larger than unconditional expenditure elasticities.

2.6 Conclusion

This study estimates demand for three products of fruits and vegetables (processed fruits and vegetables, fresh vegetables, and fresh fruits) in the United States as a weekly separable group using a 1996 to 2010 pseudo-panel constructed from the CEX and controls for a number of demographic, seasonal, regional, and unobserved group effects. This study provides useful information to the current empirical literature. First, this is the first study using pseudo-panel approach to estimating demand for fruits and vegetables. Second, this study accounts for the time-invariant group effects that are often neglected in past studies and estimates an efficient GMM estimator for valid hypothesis testing. Third, this study extends the cohort analysis to a system of equations instead of a single equation that most empirical studies did.

The results indicate that non-price factors, primarily season, region, urban status, number of children under 18, and number of persons over 64 significantly affect the demand for fruits and vegetables. Household head's educational attainment and race do not have significant effects on demand for fruits and vegetables. All fruits and vegetables are found to be "gross complements;" all fruits and vegetables except for processed fruits and vegetables and fresh vegetables are found to be "net substitutes." Results are consistent with what was found in the first chapter, even though different assumptions and data were used in two studies.

Note that the demand elasticities derived from Chapter One and Chapter Two are conditional on fruit and vegetable expenditures, where households only demand the three fruit and vegetable products. Conditional elasticities provide information on the relationships among goods within the fruit and vegetable group. This is the second stage in a two-stage budgeting framework. However, if other goods are available to consume conditional on the total income, households may change their allocation by substituting fruits and vegetables with goods outside the group. Thus, unconditional elasticities for ordinary demand functions are needed to account for total effect of price changes in consumption. By combining the results from this chapter, Chapter Three further addresses this topic through modeling the first stage of a two-stage budgeting, the results of which are used in policy analysis to sheds light on relevant policy purposes.

Table 2.1: Variables in the Model and Sample Statistics

Variable	Mean	Std. Dev.	Min	Max
Urban	0.918	0.274	0.000	1.000
Seasonality				
Quar1	0.245	0.430	0.000	1.000
Quar2	0.249	0.432	0.000	1.000
Quar3	0.247	0.431	0.000	1.000
Reference Person's Education				
No Degree	0.142	0.349	0.000	1.000
High	0.276	0.447	0.000	1.000
College	0.476	0.499	0.000	1.000
Reference Person's Race				
Black	0.104	0.305	0.000	1.000
Orace	0.047	0.211	0.000	1.000
Region				
Northeast	0.187	0.390	0.000	1.000
Midwest	0.248	0.432	0.000	1.000
South	0.326	0.469	0.000	1.000
Household Size	2.628	1.492	1.000	24.000
Under 18	0.713	1.117	0.000	12.000
Over 64	0.322	0.627	0.000	5.000
log Price				
PFV	4.169	0.409	3.553	5.089
Fresh Vegetables	4.103	0.496	3.217	5.157
Fresh Fruits	4.020	0.488	3.219	5.054
Quarterly Expenditure ($)				
PFV	55.678	60.802	0.000	1545.740
Fresh Vegetables	49.198	60.866	0.000	3463.940
Fresh Fruits	51.118	67.651	0.000	2833.740
Budget Share				
PFV	0.383	0.273	0.000	1.000
Fresh Vegetables	0.308	0.238	0.000	1.000
Fresh Fruits	0.309	0.243	0.000	1.000

Note: PFV denotes processed fruits and vegetables.

Table 2.2: Size Distribution of Groups

Time Group (T)	Income Group: < 130 % PG	130%–185% PG	185%–350% PG	350%–500% PG	>500% PG	Time Group (T)	Income Group: < 130 % PG	130%–185% PG	185%–350% PG	350%–500% PG	>500% PG
T=1	290	108	219	99	125	T=31	620	159	392	205	284
T=2	326	108	260	139	124	T=32	615	163	370	220	276
T=3	328	122	276	141	147	T=33	228	145	407	292	364
T=4	465	141	365	200	206	T=34	229	174	425	296	462
T=5	363	131	282	149	156	T=35	238	182	442	274	420
T=6	366	116	273	144	151	T=36	224	152	425	281	430
T=7	325	119	264	132	152	T=37	224	152	407	306	426
T=8	493	132	352	185	206	T=38	235	166	416	300	427
T=9	395	118	259	147	160	T=39	253	178	422	289	410
T=10	433	115	267	168	142	T=40	240	141	429	274	437
T=11	478	111	263	138	148	T=41	418	145	347	199	306
T=12	609	166	347	199	228	T=42	434	161	341	219	318
T=13	632	149	390	209	233	T=43	429	145	367	235	363
T=14	637	175	353	197	233	T=44	416	160	345	233	331
T=15	637	160	322	215	216	T=45	389	122	321	202	320
T=16	614	160	340	184	234	T=46	364	142	337	230	303
T=17	608	167	353	209	248	T=47	395	117	335	209	301
T=18	573	162	329	226	249	T=48	432	137	323	213	312
T=19	620	161	327	225	234	T=49	433	127	359	197	320
T=20	605	175	382	209	237	T=50	444	151	336	235	323
T=21	626	179	379	215	249	T=51	444	143	316	211	308
T=22	668	140	373	228	221	T=52	480	133	341	199	292
T=23	652	177	341	202	265	T=53	461	147	350	214	281
T=24	646	154	348	189	259	T=54	457	162	344	231	323
T=25	630	150	386	229	310	T=55	453	146	361	214	310
T=26	624	188	362	197	268	T=56	486	140	384	203	306
T=27	602	164	367	215	298	T=57	479	167	371	212	265
T=28	615	143	369	198	299	T=58	492	147	349	201	289
T=29	658	167	336	224	280	T=59	473	142	313	167	307
T=30	578	161	405	208	304	T=60	456	146	345	174	255

Note: PG stands for poverty guidelines.

Table 2.3: Parameter Estimates

Category n	PFV 1		Fresh Vegetable 2		Fresh Fruits 3	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
QUAR1	-0.008***	0.003	-0.005**	0.002	0.013***	0.002
QUAR2	-0.040***	0.003	-0.008***	0.003	0.048***	0.003
QUAR3	-0.039***	0.003	-0.005**	0.002	0.045***	0.003
Urban	-0.163***	0.046	0.016	0.039	0.147***	0.039
No Degree	0.028	0.058	0.049	0.049	-0.077	0.051
High	0.037	0.052	-0.027	0.043	-0.010	0.045
College	-0.020	0.050	0.014	0.042	0.006	0.043
Black	0.022	0.050	-0.028	0.042	0.005	0.043
Other Race	-0.084	0.071	0.050	0.059	0.034	0.062
Northeast	0.057	0.044	-0.002	0.037	-0.055	0.038
Midwest	-0.078*	0.041	0.039	0.034	0.039	0.036
South	-0.088**	0.037	0.030	0.031	0.058*	0.032
Household Size	-0.013	0.017	0.019	0.014	-0.006	0.015
Under 18	0.050**	0.023	-0.049***	0.019	-0.001	0.020
Over 64	-0.054**	0.023	0.005	0.019	0.048**	0.020
γ_nn						
γ_n1	0.178***	0.024	-0.126***	0.014	-0.053***	0.018
γ_n2	-0.126***	0.015	0.146***	0.016	-0.020	0.013
γ_n3	-0.053***	0.018	-0.020	0.013	0.073***	0.021
β_n	-0.008	0.017	-0.005	0.014	0.014	0.015

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

Table 2.4: Conditional Uncompensated Price Elasticities

	PFV	Fresh Vegetable	Fresh Fruits
PFV	-0.526***	-0.321***	-0.131***
	(0.062)	(0.037)	(0.050)
Fresh Vegetables	-0.401***	-0.521***	-0.061
	(0.052)	(0.047)	(0.047)
Fresh Fruits	-0.189***	-0.080*	-0.777***
	(0.062)	(0.045)	(0.071)

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

Table 2.5: Conditional Compensated Price Elasticities

	PFV	Fresh Vegetable	Fresh Fruits
PFV	-0.151***	-0.020	0.171***
	(0.058)	(0.037)	(0.048)
Fresh Vegetables	-0.025	-0.218***	0.243***
	(0.046)	(0.049)	(0.043)
Fresh Fruits	0.212***	0.242***	-0.454***
	(0.059)	(0.043)	(0.068)

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

Table 2.6: Conditional Expenditure Elasticities

	PFV	Fresh Vegetable	Fresh Fruits
Elasticities	0.978***	0.982***	1.045***
	(0.041)	(0.043)	(0.047)

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

2.7 References

Angrist, J. 1991. "Grouped-Data Estimation and Testing in Simple Labor-Supply Models." *Journal of Econometrics* 47:243-266.

Barten, A.P. 1969. "Maximum Likelihood Estimation of a Complete System of Demand Equations." *European Economic Review* 1: 7-73.

Blundell, R., C. Meghir, and P. Neves. 1993. "Labour Supply and Intertemporal Substitutions." *Journal of Econometrics* 59:137-160.

Browning, M., A. Deaton, and M. Irish. 1985. "A Profitable Approach to Labor Supply and Commodity Demands over Life-Cycle." *Econometrica* 53:503-544.

Collado, M. 1997. "Estimating Dynamic Models from Time Series of Independent Cross-Sections." *Journal of Econometrics* 83:37-62.

Deaton, A. 1985. "Panel Data form Time Series of Cross-Sections." *Journal of Econometrics* 30: 109-126.

Durbin, J. 1954. "Errors in Variables." *Review of the International Statistical Institute* 22:23-32.

Girma, S. 2000. "A Quasi-Differencing Approach to Dynamic Modelling from a Time Series of Independent Cross Sections." *Journal of Econometrics* 98:365-383.

Griliches, Z. 1986. "Economic Data Issues." In Z. Griliches and M.D. Intriligator, eds. *Handbook of econometrics , Volume 3.* Amsterdam: Elsevier Science. pp. 1465–1514.

Hansen, L.P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50:1029-1054.

Im, K.S., S.C. Ahn, P. Schmidt, and J.M. Wooldridge. 1999. "Efficient Estimation of Panel Data Models with Strictly Exogenous Explanatory Variables." *Journal of Econometrics* 93:177-201.

Inoue, A. 2008. "Efficient Estimation and Inference in Linear Pseudo-Panel Data Models." *Journal of Econometrics* 142:449-466.

McCracken, V.A. 1992. "The U.S. Demand for Vegetables." In R.A. Lopez, and L.C. Polopolus, eds. *Vegetable Markets in the Western Hemisphere: Trends, Policies, and Trade Linkages.* Ames, IA: Iowa State University Press. pp. 20-43.

McKenzie, D.J. 2004. “Asymptotic Theory for Heterogeneous Dynamic Pseudo-Panels.” *Journal of Econometrics* 120:235-262.

Moffitt, R. 1993. “Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-Sections.” *Journal of Econometrics* 59:99-123.

Verbeek, M., and F. Vella. 2005. “Estimating Dynamic Models from Repeated Cross-Sections.” *Journal of Econometrics* 127:83-102

Verbeek, M., and T. Nijman. 1992. “Can Cohort Data be Treated as Genuine Panel Data?” *Empirical Economics* 17:9-23.

Verbeek, M., and T. Nijman. 1993. “Minimum MSE Estimation of a Regression Model with Fixed Effects from a Series of Cross-Sections.” *Journal of Econometrics* 59:125-136.

Chapter 3 Subsidizing Fruits and Vegetables by Income Group: A Two-Stage Budgeting Approach

3.1 Introduction

Average Americans eat fewer fruits and vegetables compared to dietary recommended intakes. Intakes of fruits and vegetables are only 42% and 59% of the recommended goal, although the intake of solid fats and added sugars is nearly three times the recommended limit (2010 Dietary Guidelines for Americans). Lacking sufficient intake of fruits and vegetables may increase the risk of having many diseases such as heart disease, diabetes, high blood pressure, and obesity (Bazzano 2006 and Tohill et al. 2004), which are likely to cause huge economic burdens to both individuals and society. Cawley and Meyerhoefer (2012) show that obesity is estimated to raise annual medical costs by $2,741 (in 2005 dollars), while higher intake of fruits and vegetables is found to save up to more than $2,000 in annual and cumulative Medicare charges per person (Daviglus et al., 2005). [41]

A variety of policy questions are considered for improving individual's eating habits, including taxing high-calorie foods such as sugar-sweetened beverages (Zhen et al 2011) and subsidizing low-calorie and high-nutrient foods such as fruits and vegetables. Many food policies and food assistance programs are aimed at encouraging low-income households to buy healthier food, where per capita fruit and vegetable consumption for low-income households is the lowest (Lin, 2005; Dong and Lin 2009). The Farm Bill provides policy makers with the opportunity to address agricultural policy and food issues. The last Farm Bill was passed in 2008 and expired in 2012.The nutrition assistance program comprises 65% of the Bill's funding, where the Supplemental Nutrition Assistance Program (SNAP, formerly

[41] In Cawley and Meyerhoefer (2012), the authors limit the sample to adults between the ages of 20 and 64 with biological children between the ages of 11 years (132 months) and 20 years (240 months) and exclude pregnant women.

known as the Food Stamp Program) is targeted at assisting low-income households to buy healthier food (Elliott and Raziano 2012). Moreover, the Farm Bill includes the Healthy Incentives Pilot (HIP) project that authorized $20 million to evaluate incentives provided to SNAP recipients at the point of sale to increase the purchase of fruits, vegetables, or other healthful food (U.S. Department of Agriculture, Food and Nutrition Service).

Several studies have already worked on the relevant policy issues. Okrent (2012) uses an equilibrium displacement model to estimate and compare a range of obesity policies including taxing unhealthy foods such as fat and sugar and subsidizing fruits and vegetables at both farm and retail levels. Under the assumption that prices are exogenous, she found that a 10% subsidy on fruits and vegetables at the retail level increases an average adult's calorie consumption of these goods by 343 kcal per year, while the increase is only 16 kcal per year under the upward-sloping supply assumption. By Comparison, subsidizing fruits and vegetables at the farm level would lead to a larger increase in calorie consumption. However, a tax on calories is shown to be the most efficient obesity policy.

Dong and Lin (2009) estimate the effects of a 10% price discount on purchases of fruits and vegetables at the retail level for low-income households using 2004 Nielsen Homescan data.[42] Under the case of exogenous prices, they show that at-home fruit consumption increases from 0.72 cups to 0.74–0.77 cups (that is, increase by 2.8–6.9%); at-home vegetable consumption increases from 1 cup to 1.03–1.07 cups (that is, increase by 3.0–7.0%). In comparison, total fruit consumption (including the consumption of food away from home) increases from 0. 96 cups to 0.98–1.01 cups (that is, increase by 2.1–5.2%) and total vegetable consumption increases from 1.43 cups to 1.46–1.50 cups (that is, increase by 2.1–4.9%). The same dataset is also used in Dong and Leibtag (2010) to estimate the effects of two methods, coupons (10% off) and price discounts (10%), on lowering the cost of fruits and vegetables to promote fruit and vegetable consumption. They found that the effect from using coupons are larger than that from a pure price discount if consumers use coupons more than 30% of the time.

[42] A low-income household in Dong and Lin (2009) is defined as the one in which the annual income is below or equal to 130% of the federal poverty guidelines.

Lin et al. (2010) use three survey datasets to evaluate two strategies on diet improvements for food stamp recipients: subsidizing healthy food and increasing food stamp benefits. A 10% price subsidy on fruits increases at-home fruit consumption from 0.38 to 0.42 cup (10.5%) and a 10% price subsidy on vegetables increases at-home vegetable consumption from 0.94 to 1 cup (6.4%). Total consumption would increase from 0.89 to 0.97 cup (8.9%) for fruits and from 1.26 to 1.33 (5.6%) cups for vegetables. By comparison, a 10% increase in food expenditure increases at-home fruit consumption from 0.38 to 0.43 cup (13.2%) and at-home vegetable consumption from 0.94 to 1.04 cups (10.6%); it increases total fruit consumption from 0.89 to 0.97 cup (8.9%) and total vegetable consumption from 1.26 to 1.36 cup (7.9%).

The objective of this chapter is to estimate unconditional price and total expenditure elasticities and use them to examine how a price subsidy on fruits and vegetables affects demand for three products of fruits and vegetables (processed fruits and vegetables, fresh vegetables, and fresh fruits) for low-income households. Comparisons are also made between low-income and high-income households to evaluate the differences in demand for fruits and vegetables by income group. Compared to the literature, this chapter uses a new methodology to estimate demand elasticities and investigates the effects of a price subsidy on the consumption of the three fruit and vegetable products instead of aggregate fruit and vegetable consumption previously studied. This chapter also investigates how a price subsidy on fruits and vegetables affects consumption of all other goods excluding fruits and vegetables.

This chapter develops a two-stage budgeting approach to estimate a complete demand for fruits and vegetables using 1986–2010 quarterly Consumer Expenditure Survey (CEX) data. Under the assumption of rational random behavior theory (Theil 1975; Theil 1976; Theil 1980), the relative price version of the Rotterdam Model (Theil 1965; Theil 1975; Barten 1966) is applied to estimate the composite demand for a group of fruits and vegetables by using conditional elasticity estimates derived from Chapter Two. The unconditional demand elasticities are estimated by combining composite and conditional demand elasticities together. Precise standard errors are estimated by bootstrapping the entire two-stage estimation procedure.

Results show that low-income households have larger unconditional expenditure elasticities but smaller unconditional price elasticities than high-income households. For both income groups, all the fruits and vegetables are found to be net substitutes; fruits and vegetables and all other goods are also found to be net substitutes. Assuming that the supplies for fruits and vegetables are perfectly elastic, a 10% price subsidy increases consumption of processed fruits and vegetables, fresh vegetables, and fresh fruits by 3.27% (10.68%), 3.29% (10.73%), and 3.50% (11.42%), respectively, for low-income (high-income) households, and only causes a small change in consumption of all other goods.

The remainder of the chapter is organized as follows. The next section describes the CEX data, how the variables are constructed, and the data's time series properties. The third section gives a detailed introduction to the Rotterdam Model, including both the absolute and relative versions as well as the Rotterdam Model for composite goods. This section also explains the methodology used in the model estimation and gives the derivations of unconditional elasticities. Results of the demand estimation in addition to a policy application are shown in the fourth section. The final section summarizes and concludes the whole chapter and presents some issues for further study.

3.2 Data

The chapter uses the Diary Survey of CEX data from 1986 to 2010. The CEX is conducted by the Bureau Labor of Statistics (BLS) and used to maintain and support the Consumer Price Index (CPI). Although the CEX is available from 1980, nonfood composition has changed a lot from 1986. To make the data consistent, data between 1980 and 1985 are not used in this study.

The CEX represents a short panel. There are two observations for each household: One is collected in the first survey week and the other is collected in the second survey week.[43] The weekly expenditure series becomes a quarterly series by multiplying by the number of weeks in each quarter. Per capita quarterly data are calculated as the weighted average of all the observations in the same quarter, where weights, provided in the CEX for each observation, are used. It is noteworthy that weights may be different for the two observations of the same household.

To study consumption patterns for different income groups of households, data are divided into two income groups: high-income and low-income. High-income (low-income) households are ones with annual disposable income larger than (lower than or equal to) 185% of the federal poverty guidelines (PG). PG is varied by household size. Figure 3.1 shows the relationship between the two. Starting in 2004, CEX provides imputed income values, which allows not-reported income values to be estimated. Specifically, CEX includes five derived imputations and their means in addition to the original income data. In 2004 and 2005, CEX deleted the original income data and recovered them from 2006. The not-imputed data are used from 1986 to 2010 except that means of the five imputed income values are used in 2004 and 2005 due to the deletion of original income data.[44] Income could be negative for self-employed households if they reported a business loss that was greater than the income they brought in. Thus, households with zero or negative income are grouped into the low-income group.

Because the purpose of this chapter is to estimate the unconditional elasticities for the fruit and vegetable group, total expenditure is simply allocated over two broad groups: one is "fruits and vegetables" and the other is "all other goods." Given that there are no prices provided in the CEX, the nationwide CPI for fruits and vegetables is used. The quarterly CPI is derived from the monthly CPI reported on the BLS Web site. The aggregate CPI is available for fresh fruits and fresh vegetables during the whole sample period, but it is not

[43] As in previous chapters, "Consumer Units (CUs)" and "households" are used interchangeably.

[44] For more information about the imputation process and how to use these data, one can refer to "User's Guide to Income Imputation in the CE."

available for processed fruits and vegetables before the year of 1998. Thus, the 1986–1997 CPI of processed fruits and vegetables needs to be derived. We know that

(3.1) $$P_{FFV} w^*_{FFV} + P_{PFV} w^*_{PFV} = P_{FV}$$

where w^*_{FFV} and w^*_{PFV} are the expenditure shares of fresh fruits and vegetables and processed fruits and vegetables at base period, respectively; P_{FFV}, P_{PFV}, and P_{FV} are price indexes for fresh fruits and vegetables, processed fruits and vegetables, and all fruits and vegetables, respectively. Because the base period for P_{FFV} and P_{FV} is 1982–1984=100, w^*_{FFV} and w^*_{PFV} should be also in this period. However, no data are available for w^*_{FFV} and w^*_{PFV} because data from 1982 to 1984 are not used in this chapter. Considering the range of the dataset, the base period is changed to 1998. Thus, P_{PFV} can be calculated from the above equation as

(3.2) $$P_{PFV} = \frac{P_{FV} - P_{FFV} w^*_{FFV}}{w^*_{PFV}}$$

Table 3.1 shows the descriptive sample statistics for both income groups of households. From the table we can see that, on average, the fruit and vegetable expenditures and other goods expenditures for high-income households are larger than low-income households. The budget shares of fruits and vegetables are larger and those of all other goods are lower for low-income households than those for high-income households. These comparisons show that high-income households allocate a smaller portion of expenditures on fruits and vegetables compared to low-income households, which is consistent with the Engel's Law that expenditures on food falls as income increases.

The group quantities are created by dividing the current expenditure by the group CPI. Because the CPI is a close approximation to the implicit price deflator, the quantities approximate constant dollar expenditures (Nelson 1991).

Test unit roots

Before proceeding to the model, the data's time series properties should be verified. First, the data are investigated to test if there are unit roots. The Augmented Dickey-Fuller (ADF) test (Fuller, 1976; Dickey and Fuller, 1981) is applied on the following variables: expenditure shares, price indexes, and quantity indexes for each commodity group. The latter two variables are in logarithm form.

Table 3.2 reports the ADF test statistics for the null hypothesis H_0: ρ=1 in the model

$$X_t = \delta_0 + \delta_1 t + \rho X_{t-1} + \rho_1 \Delta X_{t-1} + ... + \rho_p \Delta X_{t-p} + \varepsilon_t , \tag{3.3}$$

where X_t is the variable of interest, t is the time trend and ε_t is a white noise process. The test results show that the CPI of all other goods contains a unit root; both quantity indexes and total expenditure contain a unit root for low-income households. Given the data structure, the level-form demand system should not be used, although the differential demand model is appropriate for consistent estimating the demand systems.[45]

3.3 Model

In the literature, two demand systems are widely used by agricultural economists: the Almost Ideal Demand System (AIDS) (Deaton and Muellbauer 1980a) and the Rotterdam Model. Because the AIDS model is nonlinear in parameters, a linear approximate version of AIDS (LA/AIDS) is often used in empirical work. There are many similar advantages between the LA/AIDS and the Rotterdam Model. They are both second-order locally flexible functional forms (Mountain 1988) and are both linear in parameters so are easily estimated and interpreted. Moreover, both models can be used to test economic restrictions with only linear restrictions on parameters. Brown, Lee, and Seale (1994) show that the two models are approximately equivalent in first difference form. By conducting a Monte Carlo study, Barnett and Seck (2008) compare the full AIDS model, the LA/AIDS model, and the Rotterdam Model in terms of the ability to recover the true elasticities. They conclude that the Rotterdam Model and AIDS perform much better than the LA/AIDS. The Rotterdam Model performs as well as the AIDS and often better when implementing exact aggregation within weakly separable utility function and building consistent aggregates. Alston and Chalfant (1993) use a statistical test and found the first-difference LA/AIDS model is

[45] If the dependent variables, budget shares in this case, have a unit root and all the variables are cointegrated, then the error-corrected level-form demand system may be used.

rejected but Rotterdam Model is not in an application to the meat demand. Thus, the Rotterdam Model is chosen for demand for fruits and vegetables as the first stage in a two-stage budgeting framework.

3.3.1 Rotterdam Model[46]

The differential demand system was introduced by Theil (1965, 1975) and Barten (1966). Suppose a representative consumer maximizes his or her utility u=u(q) subject to a budget constraint y=p'q, where $p = (p_1, \ldots, p_n)'$ and $q = (q_1, \ldots, q_n)'$ are vectors of prices and quantities for n goods and y is the total expenditure. The solution to this problem is the set of demand equations

(3.4) $$q_i = g_i(y,\ p),\ \text{for } i=1,\ldots,n.$$

The total differential of the equation (3.4) is

(3.5) $$dq_i = \frac{\partial g_i}{\partial y} dy + \sum_{k=1}^{n} \frac{\partial g_i}{\partial p_k} dp_k\ .$$

Using the identities $dq_i \equiv q_i d\log q_i$, $dy \equiv y d\log y$, and $dp_k \equiv p_k d\log p_k$, equation (3.5) can be written as

(3.6) $$q_i d\log q_i = \frac{\partial g_i}{\partial y} y d\log y + \sum_{k=1}^{n} \frac{\partial g_i}{\partial p_k} p_k d\log p_k\ .$$

Dividing both sides by q_i yields

(3.7) $$d\log q_i = e_i d\log y + \sum_{k=1}^{n} e_{ik} d\log p_k\ ,$$

where e_i and e_{ik} are the expenditure and Marshallian price elasticities, respectively. The generalized Slutsky equation in terms of elasticities is $e_{ik} = e_{ik}^* - w_k e_i$, where e_{ik}^* is the Hicksian price elasticity and w_k is the budget share for good k. Substituting e_{ik} for (3.7), yields

(3.8) $$d\log q_i = e_i (d\log y - \sum_{k=1}^{n} w_k d\log p_k) + \sum_{k=1}^{n} e_{ik}^* d\log p_k\ .$$

---

[46] This section is based on Theil (1975, 1976, 1980), Clements and Johnson(1983), Clements and Selvanathan (1988), and Barnett and Serletis (2009).

Recall that the budget constraint is y=p'q. The differential of the budget constraint can be written in logarithmic form as

$$(3.9) \qquad d\log y = \sum_{k=1}^{n} w_k d\log p_k + \sum_{k=1}^{n} w_k d\log q_k$$

where the second term of the right-hand side of (3.9) is the Divisia quantity index, denoted as dlogQ, representing approximate change in real total expenditure. Substituting (3.9) for (3.8) and multiplying w_i on both sides, the equation (3.8) becomes

$$(3.10) \qquad w_i d\log q_i = b_i d\log Q + \sum_{k=1}^{n} c_{ik} d\log p_k \ ,$$

where $b_i \equiv w_i e_i = p_i \frac{\partial x_i}{\partial y}$ is the marginal budget share of good i, meaning the allocation of additional expenditure on good i. If $b_i > 0$, good i is a normal good; if $b_i < 0$, good i is an inferior good; $c_{ik} \equiv w_i e_{ik}^* = (\frac{p_i p_k}{y})(\frac{\partial x_i^*}{\partial p_k})$ is the Slutsky coefficient, with $\frac{\partial x_i^*}{\partial p_k}$ the ikth element of the Slutsky matrix, representing the total substitution effect of price changes. Thus, the expenditure and compensated price elasticities are $e_i = \frac{b_i}{w_i}$ and $e_{ik}^* = \frac{c_{ik}}{w_i}$, respectively. The equation (3.10) is a differential demand system in absolute prices. The economic restrictions (adding-up, homogeneity and symmetry) in term of elasticities are

$$(3.11) \qquad \sum_{i=1}^{n} w_i e_i = 1, \ \sum_{k=1}^{n} e_{ik}^* = 0, \text{ and } w_i e_{ik}^* = w_k e_{ki}^*,$$

which implies the restrictions on the model are

$$(3.12) \qquad \sum_{i=1}^{n} b_i = 1, \ \sum_{i=1}^{n} c_{ik} = 0, \ \sum_{k=1}^{n} c_{ik} = 0, \text{ and } c_{ik} = c_{ki},$$

where $[c_{ik}]$ is negative semi-definite with rank n-1.

3.3.2 The Absolute Price Version of the Rotterdam Model

The absolute price version of the Rotterdam Model is the finite-change form of the model (3.10) with the assumption that b_i and c_{ik} are constants over the sample period. The Rotterdam Model is obtained by replacing w_i with the arithmetic average of the budget shares in t and t-1; that is, $\bar{w}_{it} = \frac{1}{2}(w_{it} + w_{it-1})$. The infinitesimal-change notation "d" in (3.10) is replaced with finite-change notation "Δ," for example, $d\log q_{it} \approx \Delta\log q_{it} = \log q_{it} - \log q_{it-1}$. Hence, the absolute price version of the Rotterdam Model is

(3.13) $$\bar{w}_{it}\Delta\log q_{it} = b_i\Delta\log Q_t + \sum_{k=1}^{n} c_{ik}\Delta\log p_{kt} + \varepsilon_{it},$$

where $\Delta\log Q_t = \sum_{k=1}^{n} \bar{w}_{kt}\Delta\log q_{kt}$ and ε_{it} is the error term.

The above model can be applied to demand for all the n goods. In applied work, we often need a demand model for groups of goods. On the one hand, the number of parameters to be estimated in a demand system can be reduced while conserving the degrees of freedom. On the other hand, data are often available for composite commodities, such as *apple* instead of Gala and Red Delicious. So the above Rotterdam Model needs to be adjusted for consistency with the data structure.

3.3.3 Composite Demand Model for the Absolute Price Version of the Rotterdam Model

Let the n goods be divided into N_G groups, where a group G has n_G observations such that $\sum_{G=1}^{N} n_G = n$. Sum (3.13) over i∈G, and this yields

(3.14) $$\sum_{i\in G} \bar{w}_{it}\Delta\log q_{it} = \sum_{i\in G} b_i\Delta\log Q_t + \sum_{i\in G}\sum_{k=1}^{n} c_{ik}\Delta\log p_{kt} + \sum_{i\in G} \varepsilon_{it},$$

Define the group quantity $\Delta\log Q_{Gt} \equiv \sum_{i\in G} \frac{\bar{w}_{it}}{\bar{w}_{Gt}}\Delta\log q_{it}$, where $\bar{w}_{Gt} = \frac{1}{2}(w_{Gt} + w_{Gt-1})$ with $w_{Gt} = \sum_{i\in G} w_{it}$; the group price $\Delta\log p_{Gt} \equiv \sum_{i\in G} \frac{\bar{w}_{it}}{\bar{w}_{Gt}}\Delta\log p_{it}$; $b_G \equiv \sum_{i\in G} b_i$, $c_{GH} \equiv \sum_{i\in G}\sum_{k\in H} c_{ik}$, $c_{Gk} \equiv \sum_{i\in G} c_{ik}$, $c_{iH} \equiv \sum_{k\in H} c_{ik}$, and $\varepsilon_G \equiv \sum_{i\in G} \varepsilon_i$. So equation (3.14) becomes

(3.15) $\bar{w}_{Gt}\Delta\log Q_{Gt} = b_G\Delta\log Q_t + \sum_H c_{GH}\Delta\log p_{Ht} + \varepsilon_G + \sum_{i\in G}\sum_{k=1}^{n} c_{ik}\Delta\log p_{kt} - \sum_H c_{GH}\Delta\log p_{Ht}$.

If the following assumptions—the covariance between c_{ik} and $\Delta\log p_{kt}$ is zero, or $c_{ik} = c_{iH}\frac{\bar{w}_{kt}}{\bar{w}_{Ht}}$, or $\Delta\log p_{kt} = \Delta\log p_{Ht}$—hold, $\sum_{k\in H}\left(c_{ik} - c_{iH}\frac{\bar{w}_{kt}}{\bar{w}_{Ht}}\right)(\Delta\log p_{kt} - \Delta\log p_{Ht})$ would be zero, and the last two expressions on the right side of (3.14) would also be zero (Barten 1967).[47] As long as the assumptions are approximately true, an error can be added to the model. By adding an intercept to (3.15), the following estimating equation is derived.

(3.16) $$\bar{w}_{Gt}\Delta\log Q_{Gt} = a_G + b_G\Delta\log Q_t + \sum_H c_{GH}\Delta\log p_{Ht} + \varepsilon_G.$$

Recall that the Divisia quantity index $d\log Q_t = \sum_k w_{kt} d\log q_{kt} = \sum_G w_{Gt} d\log Q_{Gt}$ following the budget constraint, so $\Delta\log Q_t$ is approximately equal to $\sum_G \bar{w}_{Gt}\Delta\log Q_{Gt}$. Summing over G on both sides of (3.16) implies the adding-up restrictions for the composite demand model are

(3.17) $$\sum_G a_G = 0,\ \sum_G b_G = 1,\ \sum_G c_{GH} = 0 \text{ and } \sum_G \varepsilon_G = 0;$$

and from the above definitions, homogeneity and symmetry restrictions are straightforward:

(3.18) $$\sum_H c_{HG} = 0 \text{ and } c_{GH} = c_{HG}.$$

3.3.4 The Relative Price Version of the Rotterdam Model

From "Barten's fundamental matrix equation," the Slutsky equation can be written in another form (Barten 1964; Philips 1974):

(3.19) $$\frac{\partial q_i}{\partial p_k} = \lambda U^{ik} - \left(\frac{\lambda}{\lambda_y}\right)\frac{\partial q_i}{\partial y}\frac{\partial q_k}{\partial y} - q_k\frac{\partial q_i}{\partial y}$$

where λ is the Lagrange multiplier, representing the marginal utility of income, and $\lambda_y = \frac{\partial\lambda}{\partial y}$; U^{ik} is the ikth element of the inverse of the Hessian matrix of utility function $\frac{\partial^2 U}{\partial q_i\,\partial q_k}$. Obviously, the total effect of a change in price of good k on the quantity of good i can be

---

47 In fact, $\sum_{i\in G}\sum_{k=1}^{n} c_{ik}\Delta\log p_{kt} - \sum_H c_{GH}\Delta\log p_{Ht} = \sum_{i\in G}\sum_H\sum_{k\in H}\left(c_{ik} - c_{iH}\frac{\bar{w}_{kt}}{\bar{w}_{Ht}}\right)(\Delta\log p_{kt} - \Delta\log p_{Ht})$ (Statement 1). To see why the two expressions are equivalent, please see the derivations in Appendix D.

decomposed into two effects: the total substitution effect ($\lambda U^{ik}-(\frac{\lambda}{\lambda_y})\frac{\partial q_i}{\partial y}\frac{\partial q_k}{\partial y}$) and the income effect ($-q_k\frac{\partial q_i}{\partial y}$). The total substitution effect can also be decomposed into two effects: the specific substitution effect (λU^{ik}) and the general substitution effect ($-(\frac{\lambda}{\lambda_y})\frac{\partial q_i}{\partial y}\frac{\partial q_k}{\partial y}$) (Houthakker 1960). Multiplying $\frac{p_k}{q_i}$ on both sides of the above equation, (3.19) can be written in elasticity form

(3.20) $$e_{ik} = e_{ik}^F - \phi w_k e_i e_k - w_k e_i ,$$

where $e_{ik}^F \equiv \frac{p_k}{q_i}\lambda U^{ik}$ is the Frisch elasticity that holds the marginal utility of money constant; $\phi \equiv \frac{\lambda}{\lambda_y y} = (\frac{\partial \log\lambda}{\partial \log y})^{-1}$ is the money flexibility, the inverse of the income elasticity of the marginal utility of income.[48] It is straightforward to show that the Hicksian price elasticity is

(3.21) $$e_{ik}^* = e_{ik}^F - \varphi w_k e_i e_{k.}$$

Recall that $c_{ik} \equiv w_i e_{ik}^*$ and $b_i \equiv w_i e_i$. Multiplying both sides of (3.21) by w_i, we obtain

(3.22) $$c_{ik} = \pi_{ik} - \phi b_i b_k$$

where $\pi_{ik} \equiv w_i e_{ik}^F = w_i \frac{p_k}{q_i}\lambda U^{ik} = \frac{\lambda}{y} p_i U^{ik} p_k$. Because $[U^{ik}]$ is symmetric negative definite, $[\pi_{ik}]$ is also symmetric negative definite. Combining homogeneity ($\sum_{k=1}^n c_{ik}=0$) and additivity ($\sum_{k=1}^n b_k = 1$), sum π_{ik} over k, we obtain

(3.23) $$\sum\nolimits_{k=1}^n \pi_{ik} = \sum\nolimits_{k=1}^n (c_{ik} + \phi b_i b_k) = \phi b_i .$$

Combining (3.22) and (3.23), equation (3.10) can be also expressed as

(3.24) $$w_i d\log q_i = b_i d\log Q + \sum\nolimits_{k=1}^n \pi_{ik}(d\log p_k - d\log p') ,$$

where $d\log p' = \sum_{k=1}^n b_k d\log p_k$ is the Frisch price index, used as the price deflator to transform the absolute prices into relative prices. As the absolute version of the Rotterdam Model, the relative version of the Rotterdam Model is the finite change form of (3.24) with

[48] This definition is consistent with Deaton and Muellbauer (1980b). Note that in some literature, the definition of money flexibility is the inverse of the above definition (e.g., George and King 1971; and Frisch 1959).

marginal budget share b_i and price coefficients π_{ik} treated as constants. By making the same adjustments as (3.13), equation (3.24) becomes

(3.25) $$\bar{w}_{it}\Delta\log q_{it} = b_i\Delta\log Q_t + \sum_{k=1}^{n}\pi_{ik}(\Delta\log p_{kt} - \Delta\log p_t^{'})$$

where $\Delta\log p_t' = \sum_{k=1}^{n} b_k \Delta\log p_{kt}$.

Suppose the consumer's preferences are block-independent; that is, $v(q) = \sum_{G=1}^{N_G} v_G(q_G)$, where q_G is a vector of quantities of individual goods in group G. In particular, all the n goods are divided into N_G groups such that each good belongs only to one group. Consistent with the definitions above, a group G has n_G observations where $\sum_{G=1}^{N} n_G = n$. Under block-independent preferences, marginal utility of a good only depends on the quantities of goods within the same group, so the Hessian matrix of utility function and its inverse are both block-diagonal. Thus, $\pi_{jk} = 0$ for j ∈H and k ∈G, where H and G are two different groups. Under this situation, equation (3.23) becomes

(3.26) $$\sum_{k=1}^{n}\pi_{jk} = \sum_{k\in G}\pi_{jk} = \phi b_j \text{ for } j \in H,\ k \in G$$

and equation (3.25) can be rewritten as

(3.27) $$\bar{w}_{it}\Delta\log q_{it} = b_i\Delta\log Q_t + \sum_{k\in G}\pi_{ik}(\Delta\log p_{kt} - \Delta\log p_t^{'}) .$$

3.3.5 Composite Demand Model under Block-Independent Preferences

Combining with (3.26), sum (3.27) over i∈G, we can obtain the composite demand equations as

(3.28) $$\bar{w}_{Gt}\Delta\log Q_{Gt} = b_G\Delta\log Q_t + \phi b_G(\Delta\log p_{Gt}^{'} - \Delta\log p_t^{'})$$

where $\bar{w}_{Gt,}$ $\Delta\log Q_{Gt}$ and b_G have the same definitions as in (3.14) and recall that $\sum_G b_G = 1$; $\Delta\log p_{Gt}' = \sum_{k\in G} b_k^G \Delta\log p_{kt}$ is the composite price index for group G; $b_k^G = \frac{b_k}{b_G}$ is the conditional marginal budget share of good k given expenditures on group G with $\sum_{k\in G} b_k^G = 1$. Hence, the expenditure and own-price elasticities for the group are $e_G = \frac{b_G}{\bar{w}_G}$ and $e_{GG}^* = \frac{\phi b_G}{\bar{w}_G}$.

3.3.6 Conditional Demand Model under Block-Independent Preferences

Substituting $\Delta \log Q_t$ derived from (3.28) to (3.27), conditional demand equations can be obtained as

(3.29) $$\bar{w}_{it}\Delta \log q_{it} = b_i^G \bar{w}_{Gt}\Delta \log Q_{Gt} + \sum_{k\in G} \pi_{ik}(\Delta \log p_k - \Delta \log p'_{Gt})$$

for i, k∈G, in relative price version, or in the absolute price version as

(3.30) $$\bar{w}_{it}\Delta \log q_{it} = b_i^G \bar{w}_{Gt}\Delta \log Q_{Gt} + \sum_{k\in G}^{n} c_{ik}^G \Delta \log p_{kt} ,$$

where c_{ik}^G is the conditional Slutsky coefficient. We can see that (3.30) describes demand for good i given the group expenditure of G. Thus, the conditional expenditure and compensated price elasticities are $e_i^G = \frac{b_i^G \bar{w}_{Gt}}{\bar{w}_{it}} = \frac{e_i}{e_G}$ and $e_{ik}^{G*} = \frac{c_{ik}^G}{\bar{w}_{it}}$, respectively. The economic restrictions for (3.30) are

(3.31) $$\sum_{k\in G}^{n} c_{ik}^G = 0,\ c_{ik}^G = c_{ki}^G.$$ [49]

The conditional Slutsky matrix, [c_{ik}^G], is negative semi-definite and symmetric with maximum rank $n_G - 1$.

3.3.7 Methodology for Estimating Unconditional Elasticities for Fruits and Vegetables

The overall goal of this chapter is to estimate the unconditional demand elasticities for fruits and vegetables. Because the conditional demand were already estimated in the previous chapters, they represent the second-stage demand estimates under two-stage budgeting. This chapter focuses on estimating the first-stage composite demand model. In the first stage, the consumer allocates total expenditure over two broader groups: fruits and vegetables and all other goods. In the second stage, the group expenditures of fruits and vegetables are allocated over the three products of fruits and vegetables: fresh fruits, fresh vegetables, and processed fruits and vegetables. Gorman (1959) shows that either strong separability or homothetic

[49] Remember $\sum_{k\in G} b_k^G = 1$ holds automatically.

separability in addition to weak separability guarantees the consistency of two-stage budgeting and single-stage maximization, where strong separability is the least restrictive condition.[50] Recall that the relative price version of the Rotterdam Model under block-independent preferences imposes strong separability implicitly on its functional form. Thus, this model (i.e., equation (3.28)) is used to estimate the composite demand for fruits and vegetables.

Under the assumption of rational random behavior theory (Theil 1975; Theil 1976; Theil 1980), the error terms of the composite demand model and conditional demand model are independent. Hence, the estimated conditional elasticities from previous chapters are consistent with ones derived by estimating two-stage demand models simultaneously.[51] Hence, the estimated conditional elasticities from Chapter Two are chosen to combine with the results from the first-stage demand estimation to calculate the unconditional demand responses to prices and total expenditure.

Rewriting equation (3.28) in the application on fruits and vegetables, the equation becomes

$$\bar{w}_{Gt}\Delta\log Q_{Gt} = b_G \Delta\log Q_t + \phi b_G(\Delta\log p'_{Gt} - \Delta\log p'_t) \qquad (3.32)$$

where $\Delta\log p'_{Gt} = \sum_{k=1}^{3} b_k^G \Delta\log p_{kt}$ and $\Delta\log p'_t = \sum_{k=1}^{n} b_k \Delta\log p_{kt}$.

Using the definition in the model section, the conditional expenditure elasticity derived in Chapter Two is e_k^G; that is, it is equal to $\frac{b_k^G \bar{w}_{Gt}}{\bar{w}_{kt}}$ from equation (3.30). Given $\bar{w}_{Gt}$ and $\bar{w}_{kt}$, we can calculate b_k^G, and further $\Delta\log p'_{Gt}$ used in equation (3.32). Moreover, to identify this model, $\Delta\log p'_t$ also needs to be known. That is, all the weights (all b_k's in its expression) need to be known. Because $b_k^G = \frac{b_k}{b_G}$, which implies $b_k = b_k^G b_G$, only b_k for k=1,2, 3 in $\Delta\log p'_t$ can be estimated. However, we cannot estimate the b_k's outside the group G.

[50] Recall that the concept of the strong separability is the following: The preference $v(\mathbf{q})$ is strongly separable if and only if $v(\mathbf{q}) = F[v_1(q_1) + \cdots + v_G(q_G)]$, where F function is monotonic.

[51] For the details on the procedures of estimating two demand systems simultaneously, see Theil (1980) and Clements and Johnson (1983).

Following Clements and Johnson (1983), the Frisch price index $\Delta \log p'_t$ can be approximated as

(3.33) $$\Delta \log p'_t \approx b_G \sum_{k=1}^{3} b_k^G \Delta \log p_{kt} + (1 - b_G)\Delta \log p_{ot}$$

where $\Delta \log p_{ot}$ is the CPI for all other goods excluding fruits and vegetables and can be calculated as

(3.34) $$\Delta \log p_{ot} = \frac{\Delta \log CPI_t - \sum_{k \in G} \bar{w}_{kt} \Delta \log p_{kt}}{1 - \bar{w}_{Gt}}$$

Substitute (3.33) in (3.32) and add an intercept and an error term, the estimating equation is the following

(3.35) $$\bar{w}_{Gt} \Delta \log Q_{Gt} = a_G + b_G \Delta \log Q_t + \phi b_G (1 - b_G)(\Delta \log p'_{Gt} - \Delta \log p_{ot}) + v_{Gt},$$

where a_G is the intercept to address the trend-related changes and v_G is the error term with a zero mean and a constant variance. To account for the serial correlation, first assume v_G follows an AR(1) process $v_{Gt} = \rho v_{Gt-1} + \varepsilon_{Gt}$, where ρ is the unknown parameters and is assumed to be same across equations (Berndt and Savin 1975).[52] The quarterly seasonal dummies are also put into the model to account for any seasonal effects on demand for fruits and vegetables. The equation for all other goods is dropped to account for the singularity problem due to the restriction of additivity and its parameters are retrieved from the economic restrictions.

Substitute (3.35) (ignoring the intercept and error term at this moment) into (3.30) to eliminate $\bar{w}_{Gt} \Delta \log Q_{Gt}$ (which is approximately the change of real expenditures of group G). We then obtain

(3.36) $$\bar{w}_{it} \Delta \log q_{it} = b_i \Delta \log Q_t + \sum_{k=1}^{3} c_{ik} \Delta \log p_{kt} + c_{io} \Delta \log p_{ot}$$

where

(3.37) $$c_{ik} = c_{ik}^G + \phi b_G (1 - b_G) b_i^G b_k^G \text{ and } c_{io} = -\phi b_G (1 - b_G) b_i^G$$

[52] During estimation, it is found that AR(2) is a preferred model. This issue will be revisited in the results section.

are unconditional Slutsky coefficients. Recall $b_k^G = \frac{e_k^G \bar{w}_{Gt}}{\bar{w}_{kt}}$, and $c_{ik}^G = e_{ik}^G \bar{w}_{it}$. Thus, the unconditional expenditure elasticity for fruits and vegetables is $e_i = \frac{b_i}{\bar{w}_{it}}$, for $i \in G$; the unconditional compensated price elasticities are $e_{ik}^* = \frac{c_{ik}}{\bar{w}_{it}}$ for fruits and vegetables with i, k∈G, and $e_{io}^* = \frac{c_{io}}{\bar{w}_{it}}$ for the relationship between fruits and vegetables and all other goods. The unconditional uncompensated price elasticities can be derived using the Slutsky equation.

Recall that b_k^G's are "generated regressors" in the model, by which the variance of the parameter estimates are affected. There are many discussions on the methods of consistently estimating the variance in econometrics literature (e.g., Murphy and Topel (1985)). In this chapter, a bootstrapping method is applied. Recall the data structures used in Chapter Two and Chapter Three. Both samples are CEX data. However, the sample in Chapter Two is from 1986 to 2010, while the sample in Chapter Three is from 1996 to 2010. So the samples overlap each other. This overlap needs to be handled during the bootstrapping so that the correlation among the parameters and error terms in the two models are addressed. That is also the reason why the variance derived only from the composite demand is not correct. In particular, 1,000 new samples are randomly drawn from the original sample (allowing repeated sampling and keeping the same number of households). In each draw, one new sample (called Sample A) is created and then used to repeat the whole estimation procedure in Chapter Three, from which the 1996–2010 subsample (called Sample B) is selected to first repeat the whole estimation procedure in Chapter Two. The latter procedure using Sample B produces a new set of parameters, which can further generate a new set of elasticities (evaluated at sample mean in each new sample) and b_k^G's. Combining the new b_k^G's with Sample A, a new set of parameters for a composite demand model in this chapter is derived. In total, 1,000 new sets of parameter estimates are generated, and the variance (standard deviation) of these estimates is the variance (standard error) of the parameter estimates in question.

3.4 Results

3.4.1 Model Estimates and Elasticities

Table 3.3 reports the conditional Slutsky coefficient c_{ik}^{G} and marginal budget share b_{k}^{G}, which are estimated based on the estimated conditional demand elasticities derived from Chapter Two. We can see that the estimates are very similar across income groups. Taking low-income households as an example, given group expenditures of fruits and vegetables, the allocation of an additional dollar spent on the three fruit and vegetable products (b_{k}^{G}) are approximately 37 cents, 30 cents, and 33 cents for processed fruits and vegetables, fresh vegetables, and fresh fruits, respectively.

Table 3.4 shows the maximum likelihood estimates from estimating the composite demand equation (3.35). The intercept a_{G} is around zero meaning there is no evidence of trend-related changes such as taste changes in the model. Seasonal effects play an important role in demand for fruits and vegetables. Compared to the fourth quarter, households demand more fruits and vegetables in the other three quarters, where both high-income and low-income households demand the most in the second quarter. The estimates of the group marginal budget share (b_{G}) are positive as expected and different between the two income groups. The results indicate that when the total expenditure increases by \$100, the expenditure on the group of fruits and vegetables increases by 24 cents for high-income households and by 45 cents for low-income households, respectively. Following the notation in the model section, the marginal budget shares of processed fruits and vegetables, fresh vegetables, and fresh fruits are represented as b_{1}, b_{2}, and b_{3}, respectively. By the relationship $b_{i} = b_{i}^{G} b_{G}$, b_{1}, b_{2}, and b_{3} are estimated as 8.63×10^{-4}, 7.05×10^{-4}, and 7.95×10^{-4} for high-income households, and 1.63×10^{-3}, 1.36×10^{-3}, and 1.46×10^{-3} for low-income households. This means that for high-income households, the allocation of an additional dollar income are 9 cents, 7 cents and 8 cents for processed fruits and vegetables, fresh vegetables, and fresh fruits respectively; for low-income households, the allocation of an additional dollar income are 16 cents, 14 cents and 15 cents, respectively. Thus, we know that when income increases, low-income households would increase expenditure on fruits and vegetables more than high-income households. The money flexibility (i.e., the inverse of the income elasticity of the marginal utility of income) ϕ is negative as expected. High-income households have a higher ϕ (-7.728) in absolute value than low-income households (-1.862). This is consistent with the results found in Frisch (1959). The adjusted R^2 is 0.626 for high-income households and 0.383 for low-income households. Although the error term in (3.35) is assumed to follow an AR(1) process, the problem of serial correlation still exists in the model. So an AR(2) is used to ensure that no further serial correlation exists.

All the elasticities are evaluated at the mean budget shares. The group expenditure elasticities for fruits and vegetables are positive and smaller than one, indicating fruits and vegetables are "normal goods" and "necessities," while the expenditure elasticities of all other goods are larger than one, indicating all other goods are "luxuries." Low-income households have larger group expenditure elasticities for fruits and vegetables (0.178) and all other goods (1.021) than high-income households (0.141 and 1.015, correspondingly) as expected, meaning low-income households are more responsive to total expenditure changes. By contrast, the group own-price compensated elasticities are negative as expected, and high-income households are more responsive to own-price changes than low-income households. The $[c_{ik}]$ matrix is also obtained using equation (3.37) and verified to satisfy the negativity condition.

Table 3.5 and Table 3.6 report the estimated unconditional expenditure and price elasticities for the three fruit and vegetable products and all other goods. All unconditional expenditure elasticities for fruits and vegetables are between zero and one, as expected. Low-income households have larger total expenditure elasticities than high-income households. As for price elasticities, no significant differences are found between uncompensated and compensated elasticities due to the small income effects. For high-income households, all fruits and vegetables are found to be unconditional gross and net complements. For low-income households, fresh fruits are found to be unconditional gross and net substitutes for processed fruits and vegetables and fresh vegetables, although the corresponding elasticities are not significant. High-income households have larger own-price elasticities than low-income households. Moreover, fruits and vegetables are found to be net substitutes for all other goods.

3.4.2 A Price Subsidy on Consumption of Fruits and Vegetables

As mentioned in the introduction, the aim of this chapter is to use unconditional elasticities to evaluate the total effects of a price subsidy on consumption of the three products of fruits and vegetables, especially for low-income households. Because data used here are at-home fruit and vegetable consumption, a price discount is considered to apply only to fruits and vegetables consumed at home. Assume the supplies of fruits and vegetables are perfectly elastic and other goods' prices remain unchanged. This assumption may overstate the consumption response to price changes.

Rewriting equation (3.36), we obtain

$$(3.38) \qquad \Delta \log q_{it} = e_i \Delta \log Q_t + \sum_{k=1}^{3} e_{ik}^{*} \Delta \log p_{kt} + e_{io}^{*} \Delta \log p_{ot} \, .$$

Recall that $\Delta \log Q_t$ represents the change in real total expenditure; e_i is unconditional expenditure elasticities of good i; and e_{ik}^{*} and e_{io}^{*} are unconditional compensated price elasticities of demand for fruits and vegetables with respect to their own-prices and the price of all other goods, respectively. For simplicity, equation (3.38) is written in terms of unconditional uncompensated price elasticities as

$$(3.39) \qquad \Delta \log q_{it} = e_i \Delta \log y_t + \sum_{k=1}^{3} e_{ik} \Delta \log p_{kt} + e_{io} \Delta \log p_{ot} \, ,$$

where $\Delta \log y_t$ is the change in total expenditure and e_{ik} and e_{io} are unconditional uncompensated price elasticities corresponding to the compensated definitions. Given a mean zero error term, Table 3.8 shows that a 10% price subsidy would increase low-income (high-income) households' consumption by 3.27% (10.68%) for processed fruits and vegetables, 3.29% (10.73%) for fresh vegetables, and 3.50% (11.42%) for fresh fruits. Because the demand elasticities of all other goods with respect to prices of fruits and vegetables are negative for low-income households and positive (not significant) for high-income households, and their values are small, the quantity change in all other goods are only 0.17% and -0.02% for low-income and high-income households, respectively.

3.5 Conclusion and Directions for Future Work

This chapter develops a two-stage budgeting methodology using the Rotterdam Model to estimate unconditional demand elasticities for fruits and vegetables using 1986–2010 CEX data. The demand system is separately estimated for low-income and high-income households. The estimates derived here are consistent with those derived in the literature. The estimated unconditional own-price elasticities are between -0.269 and -0.838, which are within the range of -0.07 to -2.10 estimated from previous studies (See Table 1.1); the estimated unconditional total expenditure elasticities are between 0.138 and 0.186, which are within the range of 0.07 to 5.22 from previous studies. Compared to high-income households, low-income households have larger total expenditure elasticities and smaller own-price elasticities. Fruits and vegetables and all other goods are found to be net substitutes for both the income groups of households.

The price subsidy on fruits and vegetables would increase consumption of fruits and vegetables by 3.27% to 3.50% for low-income households and by 10.68% to 11.42% for high-income households, which are also consistent with the literature. However, the subsidy's effects on consumption of all other goods are found to be very small.

Some related questions are worthy of exploring for further study. In particular, both composite demand system and conditional demand system (equations (3.35) and (3.29)) can be estimated simultaneously, allowing the correlation between the error terms of the two demand systems. The results can be considered as a test to the theory of rational random behavior (Clements and Johnson, 1983). In addition, the assumption of perfectly elastic supply curves of fruits and vegetables can be relaxed by allowing for upward-sloping supply curves for fruits and vegetables. Also, it would be interesting to disaggregate the group of "all other goods" and further explore what the effects are on both consumption of fruits and vegetables and unhealthy foods, such as sugar-sweetened beverages, if the revenue from taxing the unhealthy food is used to subsidize fruits and vegetables.

Table 3.1: Variables in the Model and Sample Statistics

	High-Income Group (N=100)				Low-Income Group (N=100)			
Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Group Frequency	1491.120	423.754	833.000	2299.000	1596.470	365.047	1070.000	2353.000
Quarterly Expenditure ($/per household)								
Total Expenditure	7861.760	1942.950	4411.790	11052.630	4285.450	943.971	2355.000	6537.780
Fruits and Vegetables	128.993	26.677	80.040	192.586	103.412	17.076	64.362	153.974
Other Goods	7732.760	1918.340	4327.530	10876.710	4182.040	929.704	2290.640	6398.900
Constant Dollar Expenditure ($/per household)								
PFV	0.476	0.046	0.395	0.589	0.388	0.055	0.257	0.470
FV	0.414	0.040	0.341	0.519	0.345	0.061	0.230	0.505
FF	0.437	0.067	0.307	0.593	0.349	0.069	0.221	0.512
Fruits and Vegetables	1.323	0.107	1.084	1.566	1.077	0.159	0.746	1.440
Other Goods	75.745	6.719	60.290	98.545	41.477	5.624	29.921	51.736
Price (1998=100)								
PFV	101.612	21.706	65.336	143.269	-	-	-	-
FV	97.851	28.481	48.764	146.691	-	-	-	-
FF	97.057	26.693	45.882	142.160	-	-	-	-
Fruits and Vegetables	99.046	25.214	54.558	143.099	-	-	-	-
Other Goods	101.116	19.945	67.083	134.381	-	-	-	-
Budget Share								
PFV	0.006	0.001	0.004	0.008	0.009	0.001	0.005	0.012
FV	0.005	0.000	0.004	0.007	0.008	0.001	0.006	0.012
FF	0.005	0.001	0.004	0.007	0.008	0.001	0.006	0.011
Fruits and Vegetables	0.017	0.002	0.013	0.021	0.025	0.003	0.018	0.035
Other Goods	0.983	0.002	0.979	0.987	0.975	0.003	0.965	0.982

Note: Prices are same for both income groups. PFV, FV, and FF denote processed fruits and vegetables, fresh vegetables, and fresh fruits, respectively.

Table 3.2: Test for Unit Roots

<table>
<tr><th rowspan="2">Group Variable</th><th colspan="2" rowspan="2">Price Indexes</th><th colspan="6">High</th><th colspan="6">Low</th></tr>
<tr><th colspan="2">Expenditure Shares</th><th colspan="2">Quantity Indexes</th><th colspan="2">Total Expenditure</th><th colspan="2">Expenditure Shares</th><th colspan="2">Quantity Indexes</th><th colspan="2">Total Expenditure</th></tr>
<tr><td></td><td>No. of Lags</td><td>ADF</td><td>No. of Lags</td><td>ADF</td><td>No. of Lags</td><td>ADF</td><td>No. of Lags</td><td>ADF</td><td>No. of Lags</td><td>ADF</td><td>No. of Lags</td><td>ADF</td><td>No. of Lags</td><td>ADF</td></tr>
<tr><td>Fruits and Vegetables</td><td>0</td><td>-3.71**</td><td>0</td><td>-5.60***</td><td>1</td><td>-5.15***</td><td rowspan="2">1</td><td rowspan="2">-3.85**</td><td>0</td><td>-5.80***</td><td>0</td><td>-5.86***</td><td rowspan="2">2</td><td rowspan="2">-2.06</td></tr>
<tr><td>Other Goods</td><td>2</td><td>-1.73</td><td>0</td><td>-5.25***</td><td>0</td><td>-7.04***</td><td>0</td><td>-7.04***</td><td>2</td><td>-2.57</td></tr>
</table>

Note: ADF stands for Augmented Dickey-Fuller test. Price indexes and total expenditure are in logarithm form and deflated by CPI of all item. The critical values at 1%, 5%, and 10% significance levels (when sample size =100) are -4.04, -3.45, and -3.15 for ADF(t), respectively. ***and ** represents significance at the 1% and 5% level. The Lag Length is determined by the sequential t rule and Akaike information criterion (AIC); its upper bound is 12 using the rules in Schwert (1989).

Table 3.3: Conditional Slutsky Coefficients and Marginal Budget Shares

	High			Low		
	PFV	Fresh Vegetables	Fresh Fruits	PFV	Fresh Vegetables	Fresh Fruits
Conditional Slutsky Coefficient (c_{ik}^{G})						
PFV	-9.456E-04***	-1.236E-04	1.069E-03***	-1.415E-03***	-1.850E-04	1.600E-03***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Fresh Vegetables	-1.253E-04	-1.113E-03***	1.238E-03***	-1.901E-04	-1.689E-03***	1.879E-03***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Fresh Fruits	1.143E-03***	1.307E-03***	-2.450E-03***	1.662E-03***	1.900E-03***	-3.562E-03***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Conditional Marginal Budget Share (b_{k}^{G})						
	0.367***	0.300***	0.338***	0.366***	0.304***	0.328***
	(0.015)	(0.013)	(0.015)	(0.015)	(0.013)	(0.015)

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

Table 3.4: Composite Demand Estimates for Both Income Groups

	High		Low	
	Fruits and Vegetables G		Fruits and Vegetables G	
Variable	Coef.	Std. Err.	Coef.	Std. Err.
a_G	2.257E-05**	0.000	-1.811E-05	0.000
b_G	2.351E-03**	0.001	4.458E-03***	0.002
ΔQ1	3.795E-05	0.000	1.501E-04	0.000
ΔQ2	6.927E-04***	0.000	1.291E-03***	0.000
ΔQ3	3.706E-04**	0.000	9.742E-04***	0.000
Phi	-7.728	15.762	-1.862	4.699
ρ_1	-0.606***	0.096	-0.399***	0.092
ρ_2	-0.193**	0.086	-0.223***	0.071
e_G	0.141**	0.060	0.178***	0.069
e^*_{GG}	-1.088***	0.183	-0.332	0.292
Adj. R^2	0.626	-	0.383	-

Note: ** and *** denote significance at 5% and 1% level, respectively. PFV denotes processed fruits and vegetables. Model is corrected for second-order autocorrelation in error term, where the parameter estimates are ρ_1 and ρ_2.

Table 3.5: Uncompensated Price Elasticities

	High				Low			
	PFV	Fresh Vegetables	Fresh Fruits	Other Goods	PFV	Fresh Vegetables	Fresh Fruits	Other Goods
PFV	-0.541***	-0.339***	-0.189**	0.926***	-0.271**	-0.119	0.063	0.153
	(0.099)	(0.067)	(0.078)	(0.197)	(0.123)	(0.094)	(0.105)	(0.295)
Fresh Vegetables	-0.416***	-0.539***	-0.118*	0.930***	-0.145	-0.318***	0.135	0.154
	(0.082)	(0.070)	(0.071)	(0.188)	(0.114)	(0.099)	(0.104)	(0.296)
Fresh Fruits	-0.205**	-0.099	-0.838***	0.990***	0.083	0.135	-0.569***	0.164
	(0.093)	(0.068)	(0.098)	(0.200)	(0.126)	(0.103)	(0.112)	(0.305)
Other Goods	0.000	0.000	0.001	-8.839	-0.006**	-0.005**	-0.005**	-2.897
	(0.001)	(0.001)	(0.001)	(13.197)	(0.003)	(0.002)	(0.002)	(4.350)

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

Table 3.6: Compensated Price Elasticities

	High				Low			
	PFV	Fresh Vegetables	Fresh Fruits	Other Goods	PFV	Fresh Vegetables	Fresh Fruits	Other Goods
PFV	-0.540***	-0.338***	-0.188**	1.061***	-0.269**	-0.118	0.065	0.323
	(0.099)	(0.067)	(0.078)	(0.191)	(0.123)	(0.094)	(0.105)	(0.288)
Fresh Vegetables	-0.415***	-0.538***	-0.118*	1.066***	-0.144	-0.317***	0.136	0.325
	(0.082)	(0.070)	(0.071)	(0.180)	(0.114)	(0.099)	(0.104)	(0.289)
Fresh Fruits	-0.204**	-0.098	-0.837***	1.134***	0.085	0.137	-0.567***	0.346
	(0.093)	(0.068)	(0.098)	(0.194)	(0.126)	(0.103)	(0.112)	(0.298)
Other Goods	0.007***	0.006***	0.006***	-7.841	0.003	0.003	0.003	-1.901
	(0.001)	(0.001)	(0.001)	(16.030)	(0.003)	(0.002)	(0.002)	(4.817)

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

Table 3.7: Total Expenditure Elasticities

	High				Low			
	PFV	Fresh Vegetables	Fresh Fruits	Other Goods	PFV	Fresh Vegetables	Fresh Fruits	Other Goods
Elasticities	0.138**	0.138***	0.147**	1.015***	0.174**	0.175**	0.186***	1.021***
	(0.060)	(0.060)	(0.063)	(0.001)	(0.068)	(0.069)	(0.071)	(0.002)

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

Table 3.8: Average Effects of a 10% Price Subsidy on Consumption of Fruits and Vegetables and Other Goods

	High				Low			
	PFV	Fresh Vegetables	Fresh Fruits	Other Goods	PFV	Fresh Vegetables	Fresh Fruits	Other Goods
Percentage Change on Consumption	10.683%	10.731%	11.419%	-0.015%	3.272%	3.287%	3.497%	0.170%

Note: PFV denotes processed fruits and vegetables.

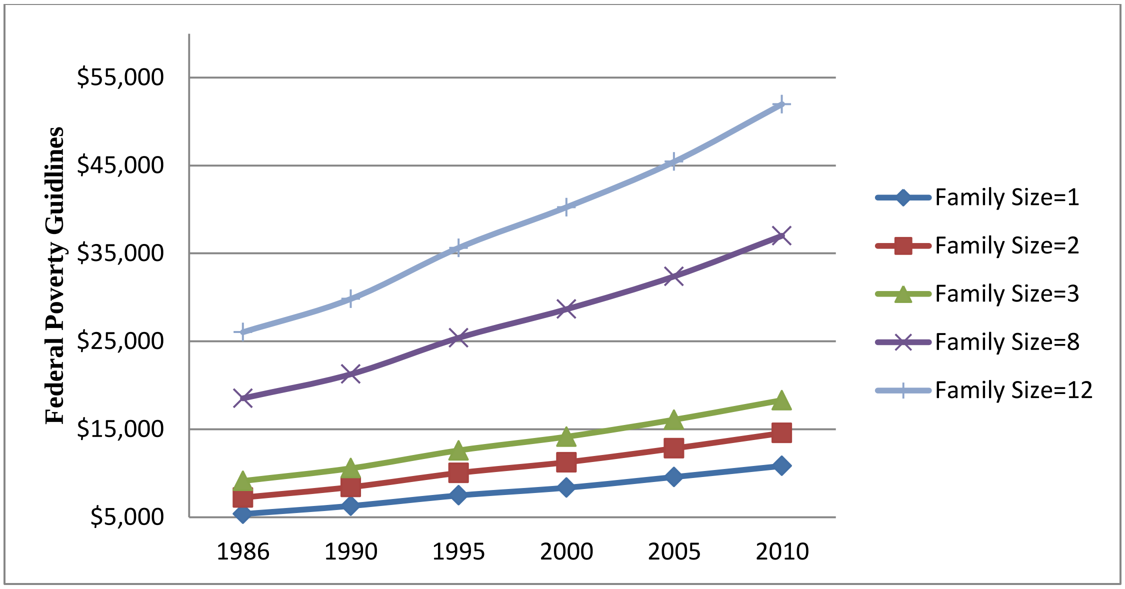

Figure 3.1: Family Size and Federal Poverty Guidelines, U.S., 1986–2010
Source: Calculations are based on data from the U.S. Department of Health and Human Services.

3.6 References

Alston, J.M., and J.A. Chalfant. 1993. "The Silence of the Lambdas: A Test of the Almost Ideal and Rotterdam Models." *American Journal of Agricultural Economics* 75:304-313.

Barnett, W.A., and A. Serletis. 2009. "Chapter 3 The Differential Approach to Demand Analysis and the Rotterdam Model." In D.J. Slottje, eds. *Quantifying Consumer Preferences (Contributions to Economic Analysis, Volume 288).* Emerald Group Publishing Limited. pp. 61-81.

Barnett, W.A., and O. Seck. 2008. "Rotterdam Model versus Almost Ideal Demand System: Will the Best Specification Please Stand Up?" *Journal of Applied Econometrics* 23:795-824.

Barten, A.P. 1964. "Comsumer Demand Functions under Conditions of Almost Additive Preferences." *Econometrica* 32:1-38

Barten, A.P. 1966. *Theorie en empirie van een volledig stelsel van vraagvergelijkingen (Theory and Empirics of a Complete System of Demand Equations)*. Ph.D. Dissertation. Netherlands School of Economics, Rotterdam.

Barten, A.P. 1967. "Evidence on the Slutsky Conditions for Demand Equations." *The Review of Economics and Statistics* 49:77-84

Bazzano, L.A. 2006. "The High Cost of Not Consuming Fruits and Vegetables." *Journal of the American Dietetic Association* 106:1364-1368.

Berndt, E.R., and N.E. Savin. 1975. "Estimation and Hypothesis Testing in Singular Equation Systems with Autoregressive Disturbances." *Econometrica* 43:937-958.

Brown, M.G., J. Lee, and R.M. Behr. 1994. "Conditional Demand and Endogeneity? A Case Study of Demand for Juice Products." *Journal of Agricultural and Resource Economics* 19:129-140.

Cawley, J., and C. Meyerhoefer. 2012. "The Medical Care Costs of Obesity: an Instrumental Variables Approach." *Journal of Health Economics* 31:219-230.

Clements, K.W., and E.A. Selvanathan. 1988. "The Rotterdam Demand Model and Its Application in Marketing." *Marketing Science* 7:60-75.

Clements, K.W., and L.W. Johnson. 1983. "The Demand for Beer, Wine, and Spirits: A Systemwide Analysis." *The Jounal of Business* 56:273-304.

Daviglus, M.L., K. Liu, A. Pirzada, L.L. Yan, D.B. Garside, R. Wang, L. Van Horn, W.G. Manning, L.M. Manheim, A.R. Dyer, P. Greenland, and J. Stamler. 2005. "Relationship of Fruit and Vegetable Consumption in Middle-aged Men to Medicare Expenditures in Older Age: The Chicago Western Electric Study." *Journal of the American Dietetic Association* 105:1735–1744.

Deaton, A., and J. Muellbauer. 1980a. "An Almost Ideal Demand System." *The American Economic Review* 70:312-326.

Deaton, A., and J. Muellbauer. 1980b. *Economics and Consumer Behavior*. New York, NY: Cambridge University Press.

Dickey, D.A., and W.A. Fuller. 1981. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica* 49:1057-1072.

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

Dong, D., and B. Lin. 2009. *Fruit and Vegetable Consumption by Low-Income Americans: Would a Price Reduction Make a Difference?* Economic Research Report 70, U.S. Department of Agriculture, Economic Research Service.

Dong, D., and E.S. Leibtag. 2010. *Promoting Fruit and Vegetable Consumption: Are Coupons More Effective Than Pure Price Discounts?* Economic Research Report 96, U.S. Department of Agriculture, Economic Research Service.

Elliott, P., and A. Raziano. 2012. *The Farm Bill and Public Health: A Primer for Public Health Professionals*. Issue Brief 6, American Public Health Association.

Frisch, R. 1959. "A Complete Scheme for Computing All Direct and Cross Demand Elasticities in a Model with Many Sectors." *Econometrica* 27:177-196.

Fuller, W. 1976. *Introduction to Statistical Time Series*. Hoboken, NJ: John Wiley & Sons Inc.

George, P.S., and G.A. King. 1971. *Consumer Demand for Food Commodities in the United States with Projections for 1980*. Davis, CA: Giannini Foundation of Agricultural Economics Monograph 25.

Gorman, W.M. 1959. "Separable Utility and Aggregation." *Econometrica* 27:469-481.

Houthakker, H.S. 1960. "Additive Preferences." *Econometrica* 28:244-257.

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

Lin, B., S.T. Yen, D. Dong, and D.M. Smallwood. 2010. "Economic Incentives for Dietary Improvement among Food Stamp Recipients." *Contemporary Economic Policy*, 28:524-536.

Mountain, D.C. 1988. "The Rotterdam Model: An Approximation in Variable Space." *Economietrica* 56:477-484

Murphy, K.M., and R.H. Topel. 1985. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business and Economic Statistics* 3:370-379.

Nelson, J.A. 1991. "Quality Variation and Quantity Aggregation in Demand for Food." *American Journal of Agricultural Economics* 73:1204-1212.

Okrent, A.M., and J.M. Alston. 2012. "The Effects of Farm Commodity and Retail Food Policies on Obesity and Economic Welfare in the United States." *American Journal of Agricultural Economics* 94:611-646.

Philips, L. 1974. *Applied Consumption Analysis*. North-Holland Publishing Co., Amsterdam.

Schwert, W. 1989. "Tests for Unit Roots: A Monte Carlo Investigation." *Journal of Business and Economic Statistics* 20:5-17.

Theil, H. 1965. "The Information Approach to Demand Analysis." *Econometrica* 33:67-87

Theil, H. 1975. *Theory and Measurement of Consumer Demand, Volume 1.* Amsterdam: North-Holland Publishing Co.

Theil, H. 1976. *Theory and Measurement of Consumer Demand, Volume 2*. Amsterdam: North-Holland Publishing Co.

Theil, H. 1980. *The System-Wide Approach to Microeconomics.* Chicago, IL: University of Chicago Press.

Tohill B.C., J. Seymour, M. Serdula, L. Kettel-Khan, and B.J. Rolls. 2004. "What Epidemiologic Studies Tell Us about the Relationship between Fruit and Vegetable Consumption and Body Weight." *Nutrition Reviews* 62:365-374.

U.S. Department of Agriculture, Food and Nutrition Service. Supplemental Nutrition Assistance Program, Healthy Incentives Pilot Project webpage. Available at: http://www.fns.usda.gov/snap/HIP/default.htm.

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

APPENDICES

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

Table A.1: 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 A.2: Definitions of Selected 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 gender	
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 Construction of S-L Price Indexes

One difficulty in estimating consumer demand is that there are no price data in the CEX. Although the CPI is used to overcome the problem, there is lack of price variation in aggregate prices compared to the demand variation by individuals. 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 the 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 because 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 modeled using the LA/AIDS model in this study.

Appendix C The Two-Limit Tobit Model

As mentioned in the text, additivity is not imposed on the model. As it is known, observed budget shares are bounded by 0 and 1. 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 $w^*_{njt} \le 0$, $w_{njt} = 1$ if $w^*_{njt} \ge 1$, $w_{njt} = w^*_{njt}$ 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 1.7, 1.8, and 1.9, Tables C1, C2, and C3 indicate 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 Tables 1.7 and 1.8, the own-price elasticities of fresh vegetables are the largest among all the own-price elasticities; in contrast, those elasticities become the smallest in the new model (see Tables C1 and C2). By comparing both income groups, low-income households have significantly smaller own-price elasticities for processed fruits and vegetables than high-income households; low-income households also have significantly larger cross-price elasticities. Moreover, the expenditure elasticities are larger for all 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 expenditure elasticity of processed fruits and vegetables is significantly larger for low-income households compared to high-income households.

Table C.1: Uncompensated Price Elasticities

	High			Low		
	PFV	Fresh vegetables	Fresh fruits	PFV	Fresh vegetables	Fresh fruits
PFV	-1.449***	-0.253***	-0.250***	-1.222***	-0.150***	-0.150***
	(0.059)	(0.008)	(0.008)	(0.054)	(0.009)	(0.009)
Fresh vegetables	-0.212***	-1.003***	-0.176***	-0.160***	-1.039***	-0.140***
	(0.007)	(0.024)	(0.006)	(0.010)	(0.045)	(0.010)
Fresh fruits	-0.227***	-0.191***	-1.128***	-0.165***	-0.144***	-1.133***
	(0.007)	(0.006)	(0.027)	(0.010)	(0.009)	(0.038)

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

Table C.2: Compensated Price Elasticities

	High			Low		
	PFV	Fresh vegetables	Fresh fruits	PFV	Fresh vegetables	Fresh fruits
PFV	-1.210***	0.131***	0.066***	-0.879***	0.264***	0.213***
	(0.080)	(0.017)	(0.016)	(0.087)	(0.022)	(0.022)
Fresh vegetables	0.092***	-0.513***	0.226***	0.219***	-0.580***	0.262***
	(0.015)	(0.051)	(0.017)	(0.020)	(0.083)	(0.023)
Fresh fruits	0.056***	0.265***	-0.753***	0.185***	0.279***	-0.762***
	(0.015)	(0.019)	(0.050)	(0.023)	(0.026)	(0.068)

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

Table C.3: 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 parentheses are standard errors.

Appendix D Proof of Statement 1

To proof $\sum_{i\in G}\sum_{k=1}^{n} c_{ik}\Delta\log p_{kt} - \sum_{H} c_{GH}\Delta\log p_{Ht} = \sum_{i\in G}\sum_{H}\sum_{k\in H}\left(c_{ik} - c_{iH}\frac{\bar{w}_{kt}}{\bar{w}_{Ht}}\right)(\Delta\log p_{kt} - \Delta\log p_{Ht})$, recall that the definition in the main text, $c_{GH} \equiv \sum_{i\in G}\sum_{k\in H} c_{ik}$, $c_{Gk} \equiv \sum_{i\in G} c_{ik}$ and $c_{iH} \equiv \sum_{k\in H} c_{ik}$. Then the left-hand side of the equation can be simplified as

$$\sum_{i\in G}\sum_{k=1}^{n} c_{ik}\Delta\log p_{kt} - \sum_{H} c_{GH}\Delta\log p_{Ht} = \sum_{i\in G}\left(\sum_{k=1}^{n} c_{ik}\Delta\log p_{kt} - \sum_{H} c_{iH}\Delta\log p_{Ht}\right)$$

$$= \sum_{i\in G}\left(\sum_{H}\sum_{k\in H} c_{ik}\Delta\log p_{kt} - \sum_{H}\sum_{k\in H} c_{ik}\Delta\log p_{Ht}\right)$$

$$= \sum_{i\in G}\sum_{H}\sum_{k\in H} c_{ik}(\Delta\log p_{kt} - \Delta\log p_{Ht}).$$

By using the definition $\Delta\log p_{Ht} \equiv \sum_{k\in H}\frac{\bar{w}_{kt}}{\bar{w}_{Ht}}\Delta\log p_{kt}$ and the fact $\sum_{k\in H}\frac{\bar{w}_{kt}}{\bar{w}_{Ht}} = 1$, the right-hand side of the equation becomes

$$\sum_{i\in G}\sum_{H}\sum_{k\in H}\left(c_{ik} - c_{iH}\frac{\bar{w}_{kt}}{\bar{w}_{Ht}}\right)(\Delta\log p_{kt} - \Delta\log p_{Ht})$$

$$= \sum_{i\in G}\sum_{H}\sum_{k\in H} c_{ik}(\Delta\log p_{kt} - \Delta\log p_{Ht}) - \sum_{i\in G}\sum_{H}\sum_{k\in H} c_{iH}\frac{\bar{w}_{kt}}{\bar{w}_{Ht}}(\Delta\log p_{kt} - \Delta\log p_{Ht})$$

$$= \sum_{i\in G}\sum_{H}\sum_{k\in H} c_{ik}(\Delta\log p_{kt} - \Delta\log p_{Ht}) - \sum_{i\in G}\sum_{H}\sum_{k\in H} c_{iH}\frac{\bar{w}_{kt}}{\bar{w}_{Ht}}\Delta\log p_{kt}$$

$$+ \sum_{i\in G}\sum_{H}\sum_{k\in H} c_{iH}\frac{\bar{w}_{kt}}{\bar{w}_{Ht}}\Delta\log p_{Ht}$$

$$= \sum_{i\in G}\sum_{H}\sum_{k\in H} c_{ik}(\Delta\log p_{kt} - \Delta\log p_{Ht}) - \sum_{i\in G}\sum_{H} c_{iH}\left(\sum_{k\in H}\frac{\bar{w}_{kt}}{\bar{w}_{Ht}}\Delta\log p_{kt}\right)$$

$$+ \sum_{i\in G}\sum_{H} c_{iH}\,\Delta\log p_{Ht}\left(\sum_{k\in H}\frac{\bar{w}_{kt}}{\bar{w}_{Ht}}\right)$$

$$= \sum_{i\in G}\sum_{H}\sum_{k\in H} c_{ik}(\Delta\log p_{kt} - \Delta\log p_{Ht}) - \sum_{i\in G}\sum_{H} c_{iH}\Delta\log p_{Ht} + \sum_{i\in G}\sum_{H} c_{iH}\,\Delta\log p_{Ht}$$

$= \sum_{i\in G}\sum_{H}\sum_{k\in H} c_{ik}(\Delta\log p_{kt} - \Delta\log p_{Ht})$, which is exactly the left-hand side of the equation.