International Food and Agribusiness Management Review
Volume 24, Issue 3, 2021; DOI: 10.22434/IFAMR2020.0090

Received: 3 June 2020 / Accepted: 30 November 2020

Demand analysis of peanuts and tree nuts in the United States: a micro-perspective

RESEARCH ARTICLE

Guo 'Chris' Cheng[a], Oral Capps, Jr.[b] and Senarath Dharmasena[c]

[a]*Econometrician, Department of Tech Services, StataCorp LLC, 4905 Lakeway Dr, College Station, TX 77845, USA*

[b]*Executive Professor and Regents Professor,* [c]*Associate Professor, Department of Agricultural Economics, Texas A&M University, College Station, TX 77845, USA*

Abstract

This paper examines household purchases of peanuts and tree nuts in the United States using the Nielsen Homescan Panel for calendar year 2015. Households located in different regions and from different races and ethnicities along with seasonality were important factors affecting the propensities to purchase and actual quantities purchased. The demand for pecans, almonds, and walnuts was sensitive to price changes. The reverse was true regarding the demands for cashews, macadamia nuts, pistachios, mixed nuts, and peanuts. All nuts were identified as necessities. Findings of this research provide insights for stakeholders in the nut industry, in terms of target marketing, product positioning, and pricing strategies. Moreover, we contribute to the literature by providing a micro-perspective investigation concerning the demand for nut products in the United States. In addition, we provide a more up-to-date analysis concerning factors affecting not only the likelihood of purchasing nuts but also the quantities purchased.

Keywords: peanuts, tree nuts, Nielsen Homescan Panel, pooled Heckman sample selection model
JEL code: D11, D12

Corresponding author: gchengchris@gmail.com

1. Introduction

Relatively little is known about the demand for peanuts and tree nuts in the United States. Previous studies have not analyzed the factors affecting the propensity to purchase these products, nor have provided a micro-perspective viewpoint as to how socio-demographic factors influence purchase decisions or purchased quantities of nuts.

Understanding the demand for peanuts and tree nuts in the United States is of economic importance because the nut industry is a notable component of the U.S. agricultural economy. According to the United States Department of Agriculture (USDA) Economic Research Service (ERS), the total crop values of peanuts and tree nuts were $11.74 billion and $9.02 billion respectively for the most recently available marketing year. Based on the most recent data from the USDA, per capita consumption of specific tree nuts is as follows: almonds (0.87 pounds); macadamia nuts (0.09 pounds); pecans (0.46 pounds); pistachios (0.13 pounds); walnuts (0.46 pounds); and other tree nuts defined as Brazil nuts, cashews, chestnuts, and pine nuts (0.64 pounds).

Meanwhile, the health benefits of nut products have been widely documented. The consumption of nuts was found to reduce the incidence of coronary heart disease, gallstones, diabetes, hypertension, cancer, and inflammation (Fraser *et al.*, 1992; Kris-Etherton *et al.*, 2008; Ros, 2010); and to decrease body mass index (BMI) (King *et al.*, 2008; Mattes *et al.*, 2008; Sabaté, 2003). In the latest Dietary Guidelines for Americans 2015-2020 (Dietary Guidelines Advisory Committee, 2015), nuts are included in the spectrum of healthy foods. About 50% of U.S. households purchased peanuts and 80% purchased tree nut products in calendar year 2015. However, relatively little is known not only about the demand for peanuts and tree nuts, but also the factors affecting the propensity to purchase these products.

In this paper, we examine the household-level purchase of peanuts and tree nuts in the United States using Nielsen Homescan Panel for calendar year 2015 and a pooled Heckman sample selection analysis. The specific objectives of this study include: (1) to determine the impacts of socio-demographic factors on the purchase of nuts; and (2) to develop profiles of households to assist stakeholders to strategically position nut products in the market.

Few economic studies have dealt with the national and regional demand, demand interrelationships, and consumer choices of nut products. To the best knowledge of the authors, current literature has not examined household choice to purchase peanuts and tree nuts in the United States. Lee (1950), Lerner (1959), Dhaliwal (1972), Wells *et al.* (1986) and Russo *et al.* (2008) examined the demand for several nut products. All these studies used annual time-series data at the market level and single-equation econometric models. None of these studies looked at the driving forces of demand at the household level.

The effects of socio-economic characteristics were investigated in Rimal and Fletcher (2002) on market participation and frequency of purchase of snack peanuts. Using a double-hurdle model, nutritional considerations affected household purchase frequency but not market participation. Similarly, He *et al.* (2005) utilized a multinomial logit model to explore the effects of age, education, race, and nutrient intakes on consumer preferences of six types of snack peanuts. Consideration of sodium and sugar in food choice did not affect consumer preference for salted peanuts. These respective studies however only consider snack peanuts.

By conducting surveys from June 2008 to December 2008, Leong *et al.* (2011) investigated nuts consumption, including peanuts, almonds, walnuts, hazelnuts, cashew nuts, and chestnuts, among Malaysian adults using a food frequency questionnaire from 364 adults. Differences in consumption by gender, education, and income level were not evident. Similarly, Aranceta *et al.* (2006) did not find differences in the consumption of nuts by gender in Spain, but consumption was associated with mid-aged and older households. Importantly, no studies provided a micro-perspective viewpoint as to how socio-demographic factors affect purchase decisions or the quantities purchased of nuts in the United States.

Our research differs from these previous studies by focusing on how socio-demographic factors affect the decision to purchase, and the amount purchased of peanuts and tree nuts. We also provide a look at specific tree nuts, and as such, we provide a more detailed depiction of the market. The nuts considered in our analysis are: (1) peanuts; (2) pecans; (3) almonds; (4) cashews; (5) walnuts; (6) macadamia nuts; (7) pistachios; and (8) mixed nuts. Our results show that: (1) households located in different regions and from different racial groups along with seasonality were important factors affecting quantities purchased; (2) the demands for pecans, almonds, and walnuts were sensitive to price changes. The reverse was true regarding the demands for cashews, macadamia nuts, pistachios, mixed nuts, and peanuts; (3) all nuts were identified as necessities; (4) the quantities purchased of nut products did not have a uniform seasonal pattern.

The remainder of the paper is structured as follows. In Section 2, the Heckman sample selection procedure is discussed. The derivation of marginal effects in the second stage is explained as they are usually miscalculated due to the lack of consideration of common factors used in both stages of the Heckman procedure. We provide and discuss the descriptive statistics of our sample in Section 3. Empirical findings are presented and summarized in Section 4. Major findings concerning the propensity to purchase and the factors in regard to actual quantities purchased are discussed as well in Section 4. In Section 5, we provide concluding remarks along with implications associated with this study.

2. Methods

We employ a Heckman sample selection model[1] to capture household choices and purchases regarding peanuts and tree nuts. In this section, we discuss the Heckman model, introduced by Heckman (1976), to estimate not only the decision to purchase peanuts and tree nuts but also to determine factors affecting the magnitude of the quantities purchased. As previously noted, the eight categories in this study are peanuts, pecans, almonds, cashews, walnuts, macadamia nuts, pistachios, and mixed nuts. Hence, our analysis provides a granular analysis of peanuts and tree nuts and therefore adds to the literature on this topic.

2.1 Pooled Heckman Model – two step procedure

To accommodate the censoring issue and the panel data structure, we implemented a pooled Heckman sample selection two-stage estimation procedure (Wooldridge, 1995, 2010). In the first stage, a probit model was used to capture the decision made by households to purchase nut products (Bliss, 1934a,b). Let Z_{hit} denote the purchase decision made by household h to product i in quarter t, which takes the value of one if this household purchases a nut product, and the value of zero if not purchased. The pooled probit model is given as follows:

$$\Pr[Z_{hit} = 1]\ \Phi\ (W_{ht}\beta_i + \mathcal{E}_{hit}) \tag{1}$$

where h = 1, …, 61,380 i = eight aforementioned nut categories, and quarter t = 1,2,3,4

Pr(·) indicates the probability that a purchase is made, Φ(·) is the cumulative standard normal distribution function, and W consists of the constant term and the set of explanatory variables used in the probit model in quarter t to capture household ‘participation’. A household is defined as a participant if a purchase is observed. The parameter β_i corresponds to the set of coefficients associated with the respective explanatory variables for nut product i. We consider quarterly dummies to capture seasonal patterns, own-price, household size (number of members), household income, age, education level, race, and ethnicity of the household head, the presence or absence of children in the household, and the region in which the household is located. $\mathcal{E}_{hit}$

[1] We also estimated Tobit models for the respective peanuts and tree nuts (Appendix S1 of the Supplementary Material). The empirical results from the estimation of the Tobit models are available from the authors upon request. On statistical and economics grounds, the Heckman models outperformed the Tobit models. Importantly, the Tobit model assumes that the decision to purchase is the same as the as the decision about how much to purchase. According to Haines *et al.* (1988), modeling food consumption decisions is a two-step process. ‘Ignoring the two-step nature of the decision process may hamper the understanding of true behavioral patterns, lead to erroneous conclusions, and generate incorrect policy recommendations.’

corresponds to the error term, $\mathcal{E}_{hit} \sim$ *standard normal* (0,1). The variable names and descriptions are depicted in Table 1.

Mathematically, we may write Equation 1 in empirical form as:

$$\begin{aligned}\Phi(W_{ht}\beta_i + \varepsilon_{hit}) &= \Pr\Bigg(\mu_{1i} + \sum_{k=1}^{3} \alpha_{ki} q_{khit} + \beta_i P_{hit} + \beta_{1i} \ln HZ_h + \beta_{2i} \ln INC_h + \beta_{3i} Age_{1h} \\ &+ \beta_{4i} Age_{2h} + \beta_{5i} Age_{3h} + \beta_{6i} Edu_{1h} + \beta_{7i} Edu_{2h} + \beta_{8i} Edu_{3h} + \beta_{9i} NE_h \\ &+ \beta_{10i} MA_h + \beta_{11i} ENC_h + \beta_{12i} WNC_h + \beta_{13i} SA_h + \beta_{14i} ESC_h + \beta_{15i} WSC_h \\ &+ \beta_{16i} Mou_h + \beta_{17i} WH_h + \beta_{18i} BL_h + \beta_{19i} AS_h + \beta_{20i} HR_h + \beta_{21i} NC_h \\ &+ \varepsilon_{hit}\Bigg)\end{aligned} \tag{2}$$

Table 1. Variable names and descriptions in the respective stages of estimation.

Notations	**Descriptions**
i	Product i, peanuts and the respective tree nuts;
h	Household h, 61,380 households;
t	Quarter t = 1,2,3,4;
Variables	
μ_{1i}/μ_{2i}	Constant terms for product i in stage 1 and in stage 2
q_{khit}	Seasonal dummy of quarter k for household h and product i at time t
P_{hit}	Unit value (a proxy for price) paid by household h for product i at time t
HZ_h	Household size for household h
INC_h	Annual income of household h
Age_{1h}	Age of household h under the age of 35
Age_{2h}	Age of household h between the age of 35 and 49
Age_{3h}	Age of household h between the age of 50 and 64
Edu_{1h}	Household h did not graduate from high school
Edu_{2h}	Household h graduated from high school but did not attend college
Edu_{3h}	Household h attended college but did not graduate
NE_h	Household h located in the New England region
MA_h	Household h located in the Middle Atlantic region
ENC_h	Household h located in the East North Central region
WNC_h	Household h located in the West North Central region
SA_h	Household h located in the South Atlantic region
ESC_h	Household h located in the East South Central region
WSC_h	Household h located in the West South Central region
Mou_h	Household h located in the Mountain region
WH_h	Race of household h White/Caucasian
BL_h	Race of household h Black
AS_h	Race of household h Asian
HR_h	Ethnicity of household h Hispanic origin
NC_h	Household h without children under the age of 18
$\mathcal{E}_{hit}$	Disturbance terms or error term in the first-stage probit specification
υ_{hit}	Disturbance term or error term in the second-stage specification

All socio-demographic factors provide insights for nut purveyors about market segmentation and targeting strategies. Hill and Lynchehaun (2002) and Dharmasena and Capps (2014) identified various cultural and socio-economic factors influencing consumer preferences including age, ethnicity, income, education, presence of children, region, and race. Hence, we hypothesize that household income, household size, and region, also are determinants of the decision to make purchases of peanuts and tree nuts. Households with higher income levels and households without children are expected to have a higher propensity for purchasing nuts. Older households are expected to purchase nut products to help mitigate health concerns (Aranceta *et al.* 2006). Given the health benefits mentioned widely in the literature, we hypothesize that college-educated households rely more on nutrition information compared to non-college-educated households (Alviola and Capps, 2010). Consequently, we expect that college-educated households are more likely to purchase peanuts and tree nuts than non-college-educated households. Seasonality, price, the location of residence as well as race and ethnicity of the households also are expected to affect the propensity to purchase peanuts and tree nuts.

In the first stage, we not only consider the factors affecting the propensity to purchase peanuts or tree nuts but also generate the inverse Mill's ratio (IMR) for each observation of non-zero purchases. IMR is the correction for the selection bias of a household not purchasing a given nut product. As noted in Equation 3, mathematically, the IMR is defined as the ratio of the standard normal probability density function (PDF) to the standard normal cumulative distribution (CDF), represented as:

$$\mathrm{IMR}_{hit} = \frac{\phi(W_{ht}\beta_i + \varepsilon_{hit})}{\Phi(W_{ht}\beta_i + \varepsilon_{hit})} \qquad \text{for} \qquad Z_{hit} = 1, \tag{3}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, $\phi(\cdot)$ is the standard normal probability density function, and IMR_{hit} is the IMR for household h and nut category i in quarter t.

In the second stage, we initially eliminate the observations of zero purchases, and to account for the possibility of sample selection bias, we add the IMR generated from the estimation of the probit model in the first stage as an additional explanatory variable. We depict this specification in Equation 4.

$$q^*_{hit} = W_{ht}\theta_i + \rho_i \mathrm{IMR}_{hit} + v_{hit} \tag{4}$$

q^*_{hit} is the observed quantity for household h and nut category i in quarter t, W_{ht} is the vector of covariates the same as in the first stage, and v_{hit} is the error term. Upon estimation of Equation 4, if the coefficient associated with IMR, ρ_i, turns out to be statistically significant, then sample selection bias is confirmed. The error terms from both stages are assumed to be normally distributed, hence the use of the probit model in the first stage. As well, these error terms are assumed to be correlated. Mathematically, we may write the empirical Equation 4 as follows:

$$\begin{aligned} q^*_{hit} = {} & \mu_{2i} + \textstyle\sum_{k=1}^{3} \sigma_{ki} q_{khit} + \theta_i P_{hit} + \theta_{1i} \ln HZ_h + \theta_{2i} \ln INC_h + \theta_{3i} Age_{1h} + \theta_{4i} Age_{2h} + \\ & \theta_{5i} Age_{3h} + \theta_{6i} Edu_{1h} + \theta_{7i} Edu_{2h} + \theta_{8i} Edu_{3h} + \theta_{9i} NE_h + \theta_{10i} MA_h + \theta_{11i} ENC_h + \\ & \theta_{12i} WNC_h + \theta_{13i} SA_h + \theta_{14i} ESC_h + \theta_{15i} WSC_h + \theta_{16i} Mou_h + \theta_{17i} WH_h + \theta_{18i} BL_h + \\ & \theta_{19i} AS_h + \theta_{20i} HR_h + \theta_{21i} NC_h + \rho_i \mathrm{IMR}_i + v_{hit} \end{aligned} \tag{5}$$

The eight nut categories are estimated separately using the pooled Heckman two-step procedure. In both stages, we assume that the impacts of household income and household size are nonlinear, hence the use of the logarithmic transformation of these variables. The base or reference categories concerning the socio-demographic variables are as follows: for age of the household head, 65 and above; for education, college graduate; for region, the Pacific; for ethnicity, not Hispanic origin; for race, other race (neither white, black, nor Asian); and the presence of children under the age of 18. Age, education, region, race, ethnicity, and the presence/absence of children in the household are dummy variables.

2.2 Marginal effects

In the first stage, the marginal effects provide insights as to how changes in the right-hand side covariate affect the probability of purchasing peanuts and tree nuts. To calculate the marginal effect for any explanatory variables or covariates, the corresponding estimated coefficient is multiplied by the standard normal density function $\Phi\ (W_{ht}\beta_i + \mathcal{E}_{hit})$. Because the marginal effects may vary from observation to observation, they are calculated at the sample means for each of the explanatory variables in the probit model.

From the previous section, we know that q^*_{hit} denotes the purchase of household h on nut product i, and if we take the expectation of Equation 4, then

$$\mathrm{E}[q^*_{hit}|Z_{hit} = 1] = W_{ht}\theta_i + \rho_i \mathrm{IMR}_{hit} \tag{6}$$

To obtain the marginal effect as discussed in Saha *et al.* (1997), we differentiate Equation 6 with respect to the set of explanatory variables (W_{hk}). The marginal effects for nut product i associated with explanatory variable k in quarter t are given by:

$$\widehat{ME}_{hkt} = \frac{\partial \mathrm{E}[q^*_{hit}|Z_{hit} = 1]}{\partial W_{hk}} = \hat{\theta}_{ik} + \hat{\rho}_i \frac{\partial \widehat{\mathrm{IMR}}_{hit}}{\partial W_{hk}} = \hat{\theta}_{ik} + \hat{\rho}_i \frac{\partial}{\partial W_{hk}} \left(\frac{\phi\left(W_{ht}\hat{\beta}_i + \varepsilon_{hit}\right)}{\Phi\left(W_{ht}\hat{\beta}_i + \varepsilon_{hit}\right)} \right)$$

With algebraic manipulation, $\widehat{ME}_{hkt}$ maybe expressed as:

$$\hat{\theta}_{ik} - \hat{\rho}_i \hat{\beta}_{ik} \left\{ W_{ht}\hat{\beta}_i \widehat{\mathrm{IMR}}_{hit} + \left(\widehat{\mathrm{IMR}}_{hit}\right)^2 \right\} \tag{7}$$

The first component $(\hat{\theta}_{ik})$ denotes the change in the kth explanatory variable associated with W_h, that affects the purchase of the ith nut category derived from the estimation of the second-stage Heckman procedure; $\hat{\rho}_i$ denotes the coefficients associated with IMR for each category; $\hat{\beta}_{ik}$ denotes the change in the k^{th} regressor associated with W_h, which affects the probability of the purchase of the i^{th} nut category derived from the estimation of the first-stage probit procedure. Based on the coefficients estimated, the calculated IMRs, and all of the explanatory variables, we obtain the marginal effects and associated standard errors in the second-stage of the Heckman procedure at the sample means for each nut product i.

3. Data

The data used for this study are derived based on data from The Nielsen Company (New York, NY, USA), LLC and marketing databases provided by the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business.[2] The data correspond to records of 61,380 households for calendar year 2015, the most recently available data to us at the time of this investigation. Recorded information concerns expenditures and quantities purchased of peanuts and tree nuts. Socio-demographic factors pertaining to household size, household income, education, race, ethnicity, age of household head, and region/location are considered, along with seasonality and own-price. The eight categories considered are peanuts, pecans, almonds, cashews, walnuts, macadamia nuts, pistachios, and mixed nuts. We first aggregated nut purchases for each household for each quarter using the household identification number provided by Nielsen, including how much they paid and the total quantities (in ounces) purchased. Then we calculated the unit value paid by dividing expenditures by total quantities purchased, which was also used as a proxy for price. Because not all households make purchases at each time period (quarter), we designate all of those purchases as zeros. The panel data consist of 61,380 households across four quarters, a total of 245,520 observations.

[2] Disclaimer: Researcher(s) own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

Table 2 exhibits the descriptive statistics for unit values (US$/ounce) (used as proxy for price) and quantities purchased (ounce) including conditional and unconditional figures. Due to the nature of our data, only purchases made by households are observed (conditional). We imputed the missing unit values using the method mentioned in Appendix S2 of the Supplementary Material. We limit our discussion to the conditional quantities purchased and the unconditional unit values. Peanuts had the largest quantities purchased, averaging 50.77 ounces purchased by households per quarter. Among the tree nut products, mixed nuts, cashews, almonds, and pistachios had the largest amounts purchased. Macadamia nuts had the highest unit value (or price), 96 cents per ounce, while mixed nuts yielded the lowest value among tree nut products, averaging 41 cents per ounce. The unit values of other tree nut products varied from 45 cents to 96 cents per ounce, on average, for cashews, almonds, walnuts, pistachios, pecans, and macadamia nuts. The average unit value of peanuts was 17 cents per ounce. As exhibited in Figure 1, we calculated the percentage of purchases in the Nielsen data by dividing the number of non-zero purchases by the total number of observations. The percentage for peanuts was 23%, and the percentage for tree nuts varied from 1 to 13%.

Nielsen also provided socio-demographic information for each household. We selected eight factors, household size, household income, age, education level, race, and ethnicity of the household head, presence or absence of children, and region in which the household is located. As exhibited in Figure 2, approximately 67% of the households in the Nielsen sample had one to two members, and more than 90% of households had less than four members. There were fifteen different household income categories as shown in Figure 3. On average, household income in the Nielsen data was $58,488 (Table 3).

Table 2. Descriptive statistics of quantities (oz) and unit values (prices) ($/oz) (Nielsen Homescan Panel, 2015).[1]

	Conditional					Unconditional				
Categories	Observations	Mean	Standard deviation	Minimum	Maximum	Observations	Mean	Standard deviation	Minimum	Maximum
Quantity (Oz purchased) per household per year										
Peanuts	56,509	50.77	91.90	0.42	3,520.00	245,520	11.69	48.99	0.00	3,520.00
Pecans	18,130	20.50	31.51	0.75	2,368.00	245,520	1.51	10.10	0.00	2,368.00
Almonds	31,627	25.62	50.81	0.20	4,220.00	245,520	3.30	20.16	0.00	4,220.00
Cashews	28,855	29.57	35.16	0.75	1,216.00	245,520	3.47	15.36	0.00	1,216.00
Walnuts	21,662	25.01	35.95	1.25	1,792.00	245,520	2.21	12.82	0.00	1,792.00
Macadamia nuts	1,392	11.73	15.22	1.00	187.00	245,520	0.07	1.44	0.00	187.00
Pistachios	17,967	26.12	33.51	0.35	615.50	245,520	1.91	11.33	0.00	615.50
Mixed nuts	27,699	36.19	42.84	1.25	1,428.00	245,520	4.08	18.39	0.00	1,428.00
Unit value (US$/oz) by nuts category										
Peanuts	56,509	0.18	0.17	0.00*	11.99	245,520	0.17	0.08	0.00*	11.99
Pecans	18,130	0.66	0.27	0.01	3.33	245,520	0.63	0.08	0.01	3.33
Almonds	31,627	0.57	0.27	0.00*	8.91	245,520	0.53	0.10	0.00*	8.91
Cashews	28,855	0.47	0.16	0.00*	11.00	245,520	0.45	0.06	0.00*	11.00
Walnuts	21,662	0.57	0.23	0.00*	7.24	245,520	0.55	0.08	0.00*	7.24
Macadamia nuts	1,392	1.00	0.28	0.05	3.04	245,520	0.96	0.06	0.05	3.04
Pistachios	17,967	0.63	0.29	0.00*	19.86	245,520	0.60	0.08	0.00*	19.86
Mixed nuts	27,699	0.45	0.18	0.00*	4.37	245,520	0.41	0.07	0.00*	4.37

[1] For unconditional figures, missing unit values (a proxy for prices) were replaced with imputed prices. See Appendix S2 of the Supplementary Material for the details of this imputation process; * = numbers are less than 0.01 but greater than zero.

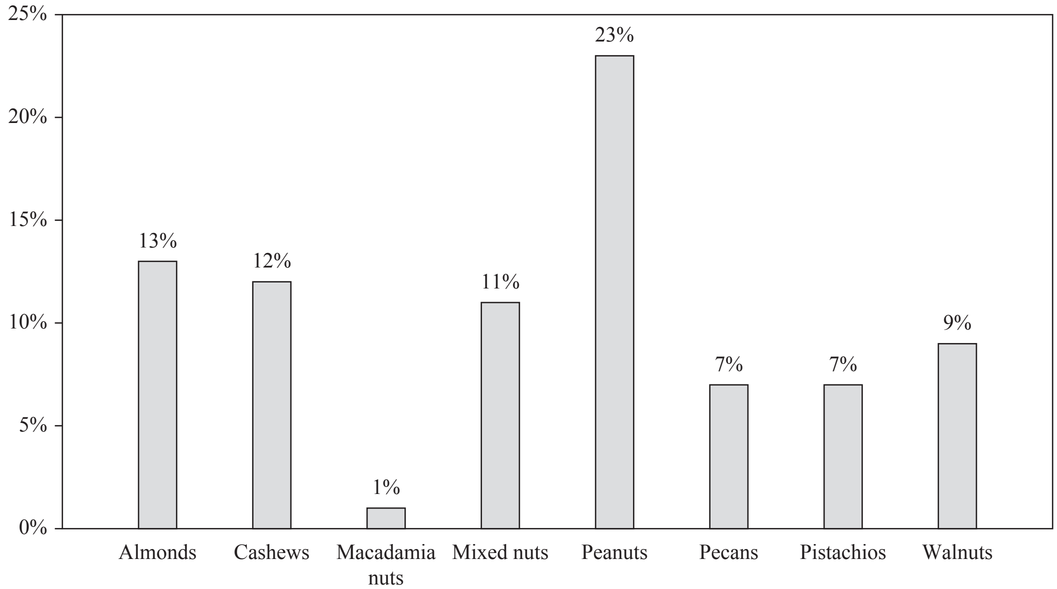

Figure 1. Percentage of non-zero quarterly purchases in calendar year 2015 (Nielsen Homescan Panel, 2015). The percentage for the respective nuts categories is calculated by counting the number of observations who had actual purchases (obs in Table 1 for each quantity) and dividing this figure by the number of total observations (245,520).

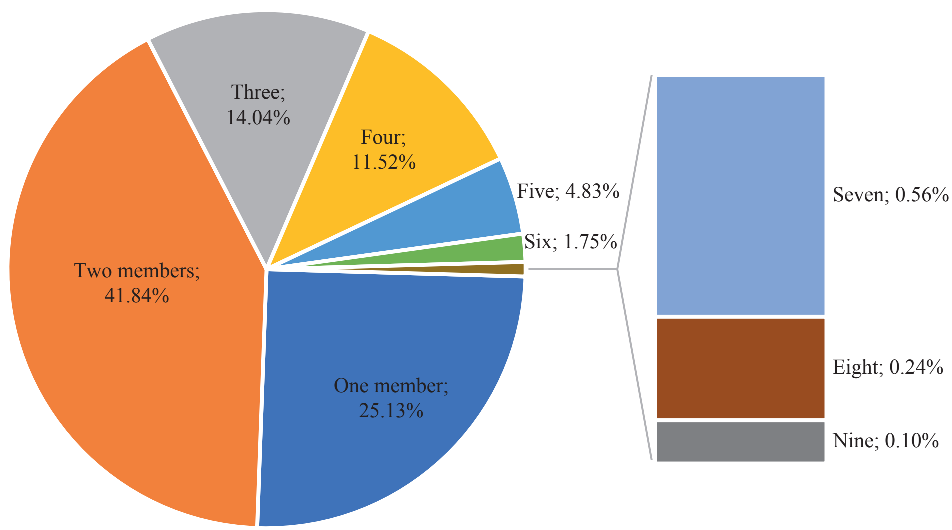

Figure 2. Household size of the panelists in calendar year 2015 (Nielsen Homescan Panel, 2015). The minimum is one household member, the maximum is nine members.

Except for household size and income, socio-demographic factors correspond to dummy variables in the regression. Consequently, the mean of these explanatory variables is the percentage of households with these characteristics in the sample, as shown in Table 3. Most of the household heads in the sample were between the ages of 50 and 64. The largest share of households was located in the South Atlantic region, around 21%, and the smallest share of households was located in the New England region, about 5%. As exhibited in Figure 4, the regional delineation follows the classification provided by the U.S. Census Bureau. About 19% of the households had at most a high-school education. Slightly more than four-fifths of households were white/Caucasian. Nearly 80% of households did not have children under the age of 18. Approximately 6% of the sample was of Hispanic origin.

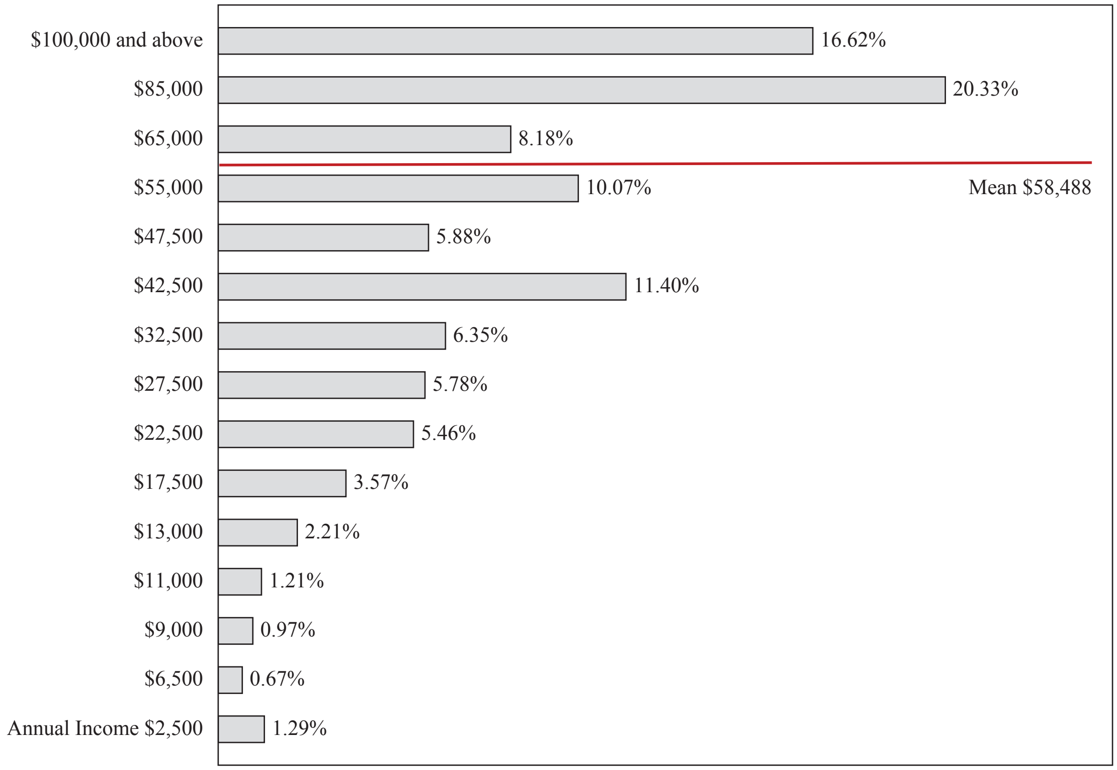

Figure 3. Income levels of households in the Nielsen Homescan Panel for calendar year 2015 (Nielsen Homescan Panel, 2015).

To demonstrate the representativeness of our sample to the U.S. population, we compare the socio-demographic characteristics of our sample with population statistics provided by the U.S. Census Bureau for 2010 (DeNavas-Walt *et al.*, 2011; U.S. Census Bureau, 2010, 2020). As exhibited in Table 4, according to the 2010 Census, average household income was $49,445, less than the average income of our sample, $58,488; household size, according to the 2010 Census was 2.58, in line with our average household size of 2.38. Further, similar percentages of race, region, and the presence/absence of children were evident. The average age of our sample (58) was higher compared to the average age reported by the 2010 Census (48). As well, the percentage of Hispanic households (6.13%) was lower in our sample compared to the percentage of Hispanic households reported by the 2010 Census (11.51%). Finally, in our sample, the percentage of households whose heads received at least some college education was 80.51%, compared to 55.24% from the 2010 Census. Aside from these differences in household income, age, ethnicity, and education of the household head, our sample of households matched up to the U.S. population as represented by the 2010 Census.

Table 3. Descriptive statistics associated with the set of socio-demographic variables (Nielsen Homescan Panel, 2015).

Variable/label	**Description (n=245,520)**	**Mean**	**Std. Dev.**
Continuous variables			
Household size			
hsize	Number of household members	2.38	1.30
Income (US$)			
income	Household income	58,488	29,235
Dummy variables			
Season			
q1	First quarter of calendar year 2015	0.25	0.43
q2	Second quarter of calendar year 2015	0.25	0.43
q3	Third quarter of calendar year 2015	0.25	0.43
q4[1]	Fourth quarter of calendar year 2015	0.25	0.43
Age of household head			
agehh_under35	Under the age of 35	0.08	0.26
agehh_35to49	Between 35 and 49	0.24	0.43
agehh_50to64	Between 50 and 64	0.43	0.50
agehh_65andabove[1]	65 and above[1]	0.25	0.43
Education level of household head			
eduhh_lesshigh	Less than high school	0.01	0.11
eduhh_highschool	High school	0.18	0.39
eduhh_somecollege	Some college	0.29	0.45
eduhh_grad[1]	At least college degree[1]	0.52	0.50
Race			
White	White	0.82	0.39
Black	Black	0.11	0.31
Asian	Asian	0.03	0.18
Other[1]	Other[1]	0.05	0.21
Ethnicity			
hispanic_reg	Hispanic Origin	0.06	0.24
Region			
NewEngland	New England	0.05	0.21
MiddleAtlantic	Middle Atlantic	0.13	0.34
ENCentral	East North Central	0.18	0.38
WNCentral	West North Central	0.08	0.28
SouthAtlantic	South Atlantic	0.21	0.40
ESCentral	East South Central	0.06	0.24
WSCentral	West South Central	0.10	0.31
Mountain	Mountain	0.07	0.26
Pacific[1]	Pacific[1]	0.12	0.33
Presence of children under 18			
no_child	No presence of children	0.78	0.42

[1] Denotes the base or reference category of the socio-demographic characteristics. The list of regions is from the U.S. Census Bureau, see Figure 3 for more details.

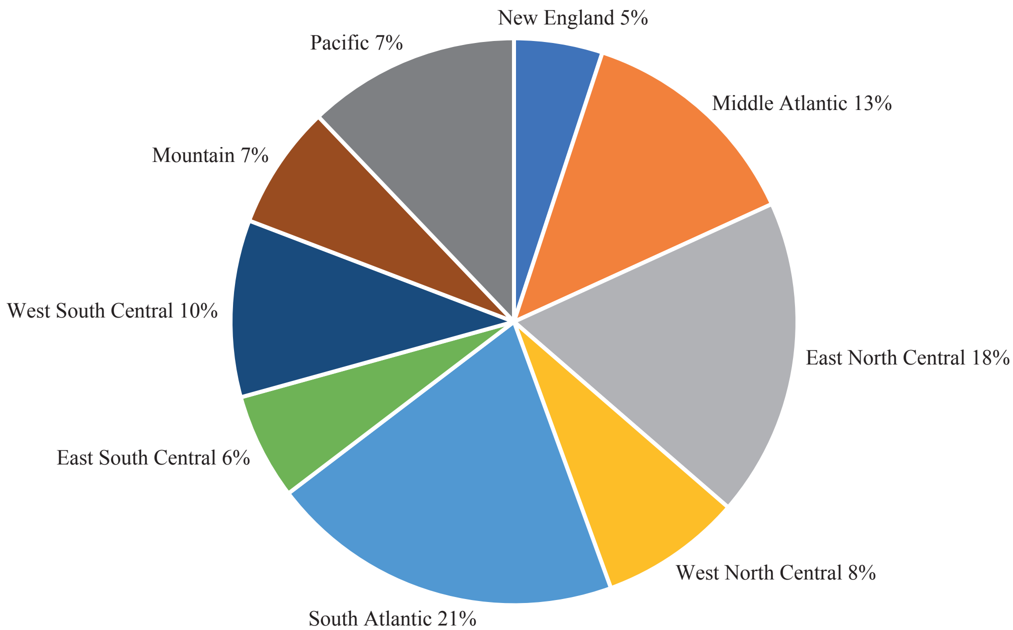

Figure 4. Location of households in the Nielsen Panel in the United States for calendar year 2015 (Nielsen Homescan Panel, 2015). New England: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont; Middle Atlantic: New Jersey, New York, and Pennsylvania; East North Central: Indiana, Illinois, Michigan, Ohio, and Wisconsin; West North Central: Iowa, Nebraska, Kansas, North Dakota, Minnesota, South Dakota, and Missouri; South Atlantic: Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, and West Virginia; East South Central: Alabama, Kentucky, Mississippi, and Tennessee; West South Central: Arkansas, Louisiana, Oklahoma, and Texas; Mountain: Arizona, Montana, Colorado, Utah, Idaho, Nevada, New Mexico, and Wyoming; Pacific: Alaska, California, Hawaii, Oregon, and Washington.

4. Empirical results

4.1 Coefficient estimates and joint tests

The pooled Heckman selection models[3] were estimated using Stata, Version 16 (StataCorp LLC, College Station, TX, USA). In Table 5, we provide the empirical results of the estimation of the first-stage pooled probit models for peanuts and tree nuts. Each model rests on the use of 245,520 pooled observations. The goodness-of-fit metric, McFadden R^2, ranged from 0.026 to 0.071.

All coefficients associated with own-price variables[4] were statistically significant and positive, meaning that household choices align with more expensive products of peanuts and tree nuts. Perhaps, higher prices translate to higher quality, which translates to a greater likelihood of purchases. Trajtenberg (1989) and

[3] As a check on robustness, we also estimated the respective models for peanuts and tree nuts using ordinary least squares (OLS) owing to a comment provided by one reviewer. As discussed in Puhani (2000), in the presence of multicollinearity, OLS provides more robust estimates than either maximum likelihood estimators or Heckman estimators. Variance inflation factors, condition indices and variable proportions were used to examine potentially degrading collinearity issues (Belsley *et al.*, 1980). No degrading collinearity issues were evident upon this examination (Appendix S3 of the Supplementary Material). This justifies our use of Heckman sample selection procedure in this study. Owing to another comment, the exclusion restrictions (Puhani, 2000) require that at least one variable that appears with a non-zero coefficient in the selection equation (first stage) not appear in the second stage equation. There is evidence (Appendix S4 of the Supplementary Material) that at least one variable is statistically significant in the first stage, however insignificant in the second stage. That is, we do include the factors that have an impact on the propensity to purchase but they do not have a statistically significant impact on the actual quantity purchased. It is common to have the same set of covariates in both stages because the factor affecting the propensity to purchase would also be likely to influence the actual quantity purchased by a household.

[4] Owing to the data-censoring issue, we imputed the missing price variables that were included as explanatory variables. We regressed the non-missing prices on selected socio-demographic factors and used the estimated coefficients to impute missing prices. Then the missing prices were replaced with these imputations. In Appendix S2 of the Supplementary Material, details are provided concerning the imputations of missing prices or unit values.

Table 4. Representativeness of the sample to the U.S. population based on the 2010 U.S. Census (Nielsen Homescan Panel, 2015; U.S. Census Bureau, 2020).

Variable/label	Description	Our sample (mean)	US Census (2010)
Household size			
hsize	Number of household members	2.38	2.58
Income (US$)			
income	Household income	58,488	49,445
Education level of household head			
eduhh_lesshigh	Less than high school	0.01	0.14
eduhh_highschool	High school	0.18	0.31
eduhh_somecollege	Some college	0.29	0.28
eduhh_grad[1]	At least college degree[1]	0.52	0.27
Race			
White	White	0.82	0.81
Black	Black	0.11	0.13
Asian	Asian	0.03	0.04
Other[1]	Other[1]	0.05	0.02
Ethnicity			
hispanic_reg	Hispanic Origin	0.06	0.12
Region			
NewEngland	New England	0.05	0.05
MiddleAtlantic	Middle Atlantic	0.13	0.13
ENCentral	East North Central	0.18	0.15
WNCentral	West North Central	0.08	0.07
SouthAtlantic	South Atlantic	0.21	0.19
ESCentral	East South Central	0.06	0.06
WSCentral	West South Central	0.10	0.12
Mountain	Mountain	0.07	0.07
Pacific[1]	Pacific[1]	0.12	0.16

[1] Denotes the base or reference category of the socio-demographic characteristics.

Spence (1973) suggest price and quality are positively associated. However, once the decision to purchase has been made, households are likely to limit quantities purchased due to price in accord with the law of demand. Seasonality was evident in the decision to purchase peanuts and tree nuts. However, the seasonal nature of the propensity to purchase was not uniform across the respective nut products. Household size and household income were positively related to the decision to purchase peanuts and tree nuts. Most of the socio-demographic factors influence the decision to purchase nuts and as such were driving forces of household-level choices in the decision to purchase.

In Table 6, we report the empirical results associated with the estimation of the second-stage Heckman selection models. The goodness-of-fit metric, pseudo R^2, for each nut category was obtained following Veall and Zimmermann (1996). The pseudo R^2 ranged from 0.000[5] to 0.063. All of the coefficients associated with the inverse Mill's ratio, except for macadamia nuts, were statistically significant, indicative of the presence of sample selection bias. As common covariates were used in both stages, we calculated the respective marginal effects based on Equation 7, following Saha *et al.* (1997). The discussion of the marginal effects is provided in the next section.

[5] The pseudo R^2 for macadamia nuts is 0.00006.

Table 5. Summary of parameter estimates of the coefficients in the respective first-stage probit models for peanuts and tree nuts (Nielsen Homescan Panel, 2015).[1]

	(1) Peanuts	(2) Pecans	(3) Almonds	(4) Cashews	(5) Walnuts	(6) Macadamia nuts	(7) Pistachios	(8) Mixed nuts
up_peanuts	**4.221**							
up_pecans		**1.560**						
up_almonds			**1.736**					
up_cashews				**2.245**				
up_walnuts					**1.756**			
up_macadamia						**1.745**		
up_pistachios							**1.598**	
up_mixed								**2.315**
q2	**-0.037**	**-0.056**	**-0.082**	**-0.032**	**-0.079**	0.030	**-0.146**	**-0.081**
q3	**-0.049**	**-0.087**	**-0.177**	**-0.103**	-0.016	**0.047**	**-0.193**	**-0.096**
q4	**-0.026**	**0.521**	**-0.179**	-0.010	**0.478**	**0.184**	**-0.034**	**0.055**
lhsize	**0.253**	**0.174**	**0.080**	**0.147**	**0.147**	**0.092**	**0.107**	**0.176**
lincome	**0.036**	**0.130**	**0.168**	**0.060**	**0.102**	**0.070**	**0.124**	**0.060**
agehh_under35	**-0.489**	**-0.346**	**-0.092**	**-0.428**	**-0.534**	-0.064	**-0.335**	**-0.555**
agehh_35to49	**-0.268**	**-0.276**	**-0.021**	**-0.249**	**-0.415**	-0.016	**-0.165**	**-0.363**
agehh_50to64	**-0.098**	**-0.098**	**0.030**	**-0.070**	**-0.188**	0.023	**-0.028**	**-0.112**
eduhh_lesshigh	**-0.111**	**-0.190**	**-0.317**	**-0.139**	**-0.164**	**-0.311**	**-0.204**	**-0.147**
eduhh_highschool	**-0.029**	**-0.068**	**-0.194**	**-0.067**	**-0.070**	**-0.104**	**-0.054**	**-0.091**
eduhh_somecollege	**-0.014**	**-0.034**	**-0.103**	**-0.029**	**-0.062**	-0.025	-0.005	**-0.050**
NewEngland	**0.115**	-0.033	0.019	-0.019	**0.172**	**-0.112**	**0.129**	-0.005
MiddleAtlantic	**0.033**	**-0.203**	**-0.031**	-0.018	**0.124**	**0.062**	**0.084**	**-0.061**
ENCentral	**0.165**	**0.119**	**0.023**	**0.117**	**0.186**	**-0.294**	**0.085**	**0.032**
WNCentral	**0.144**	**0.136**	**0.034**	**0.071**	-0.020	**-0.318**	0.001	**0.129**
SouthAtlantic	**0.140**	**0.150**	**0.044**	**0.090**	**0.066**	**-0.197**	**0.062**	**0.151**
ESCentral	**0.164**	**0.198**	-0.026	**0.065**	0.002	**-0.101**	**-0.040**	**0.196**
WSCentral	**0.091**	**0.281**	0.007	0.001	**0.040**	**-0.255**	**-0.047**	**0.112**
Mountain	**0.035**	**0.183**	**-0.050**	**0.089**	-0.002	**-0.089**	**0.076**	**0.078**
White	0.014	**0.036**	**0.054**	**-0.051**	**0.037**	**-0.090**	0.000	-0.003
Black	-0.018	**0.107**	**-0.106**	0.006	-0.037	**-0.372**	**0.039**	-0.012
Asian	0.004	**-0.208**	0.003	-0.010	0.005	-0.031	**0.095**	**0.065**
hispanic_reg	**-0.033**	-0.001	0.007	**-0.082**	0.004	-0.072	**0.058**	-0.012
no_child	**0.182**	**0.127**	**0.080**	**0.150**	**0.188**	**0.107**	**0.051**	**0.191**
Constant	**-2.108**	**-4.194**	**-3.871**	**-2.910**	**-3.674**	**-4.911**	**-3.756**	**-2.960**
McFadden R^2	0.029	0.071	0.034	0.026	0.053	0.034	0.027	0.043

[1] Bold numbers indicate significance level at 10% or lower. See Appendix S5 of the Supplementary Material for the detailed table with standard errors and significance level.

Table 6. Summary of parameter estimates of the coefficients in the respective second-stage Heckman models for peanuts and tree nuts (Nielsen Homescan Panel, 2015).[1]

	(1) Peanuts	(2) Pecans	(3) Almonds	(4) Cashews	(5) Walnuts	(6) Macadamia nuts	(7) Pistachios	(8) Mixed nuts
up_peanuts	**15.060**							
up_pecans		**66.157**						
up_almonds			**11.271**					
up_cashews				**7.947**				
up_walnuts					**27.243**			
up_macadamia						-23.279		
up_pistachios							**4.272**	
up_mixed								-7.022
q2	**-4.706**	**-3.684**	**-2.554**	0.838	**-3.188**	-0.534	**-2.956**	-0.563
q3	**-4.341**	**-5.193**	**-4.825**	**-1.912**	0.429	-1.787	**-3.710**	0.500
q4	**-3.366**	**37.212**	**-5.353**	0.792	**20.639**	-1.508	-0.911	-0.249
lhsize	**23.205**	**14.657**	**5.931**	**6.474**	**8.255**	-1.175	**4.327**	**5.656**
lincome	**5.999**	**10.830**	**9.383**	**5.117**	**7.094**	0.094	**6.502**	**6.720**
agehh_under35	**-52.660**	**-29.679**	**-10.236**	**-19.525**	**-33.295**	-1.807	**-16.645**	**-15.247**
agehh_35to49	**-30.267**	**-23.186**	**-5.679**	**-11.708**	**-25.151**	-0.940	**-8.638**	**-10.494**
agehh_50to64	**-11.745**	**-7.671**	**-1.883**	**-3.853**	**-11.334**	0.445	**-1.491**	**-2.643**
eduhh_lesshigh	**-8.373**	**-13.332**	**-9.587**	**-11.137**	**-13.650**	1.863	**-9.443**	**-11.001**
eduhh_highschool	**-3.026**	**-5.378**	**-10.916**	**-5.910**	**-6.208**	-1.682	**-3.202**	**-6.410**
eduhh_somecollege	**-2.609**	**-3.935**	**-6.580**	**-4.105**	**-5.772**	-0.253	-0.829	**-6.070**
NewEngland	2.019	**-7.468**	**-3.360**	**-8.906**	**3.277**	-2.041	-2.164	**-12.492**
MiddleAtlantic	1.248	**-17.391**	**-2.868**	**-7.737**	**4.424**	-2.955	**-2.779**	**-13.338**
ENCentral	0.979	**4.941**	**-3.239**	**-2.963**	**3.540**	0.221	**-2.732**	**-10.328**
WNCentral	-2.020	**5.967**	**-2.696**	**-4.769**	**-7.299**	-2.115	-1.803	**-6.280**
SouthAtlantic	-1.330	**7.841**	**-4.280**	**-4.449**	-0.828	-3.071	**-3.567**	**-9.026**
ESCentral	-2.979	**12.220**	**-6.537**	**-6.487**	**-4.431**	-3.642	**-7.491**	**-5.379**
WSCentral	**-5.900**	**17.450**	**-5.672**	**-5.732**	**-3.876**	-3.843	**-3.929**	**-7.271**
Mountain	**-7.934**	**12.984**	**-3.152**	1.448	-0.620	**-6.584**	0.489	-1.759
White	**5.413**	3.540	**3.240**	-0.906	-0.237	-0.296	**-2.248**	-1.564
Black	-1.347	**11.140**	1.056	-1.986	-0.746	3.616	-1.467	**-6.183**
Asian	**5.505**	**-14.483**	**4.920**	**6.177**	**4.816**	**5.506**	**8.366**	**4.058**
hispanic_reg	**-5.353**	0.376	-1.347	**-3.442**	-1.163	2.728	**-2.041**	-1.592
no_child	**22.375**	**11.253**	**7.047**	**8.083**	**12.785**	1.004	**4.372**	**9.064**
IMR	**91.077**	**85.747**	**48.863**	**35.517**	**59.274**	-9.035	**33.255**	**17.413**
Constant	**-153.154**	**-326.571**	**-160.674**	**-86.660**	**-180.160**	62.429	**-105.187**	**-57.871**
Pseudo R^2	0.006	0.019	0.000*	0.015	0.000*	0.063	0.010	0.037

[1] * the pseudo R^2 for walnuts is less than 0.001; bold numbers indicate significance level at 10% or lower. See Appendix S6 of the Supplementary Material for the detailed table with standard errors and significance level. Pseudo R^2 was calculated following Veall and Zimmermann (1996).

As shown in Table 7, joint tests were conducted for each demographic variable to examine significance for age of household head, education of household head, region, and race for both stages. For almonds, cashews, mixed nuts, peanuts, pecans, pistachios, and walnuts, these socio-demographic factors were jointly statistically significant concerning the propensity to purchase and the number of purchases. For macadamia nuts, age of household head was not a determinant of the propensity to purchase and the number of purchases. The education of household head also was not a determinant of the amount of macadamia nuts.

4.2 Marginal effects and elasticities

The marginal effects associated with the probit model from the first-stage are not reported for brevity, which is available from the authors upon request.

The marginal effects associated with the second-stage were calculated using Equation 7 and the PREDICTNL commands in Stata (version 16) for the various nut products. The standard errors were obtained using the delta-method. As shown in Table 8, all of the marginal effects associated with unit values were negative and statistically significant. Unlike the first stage, no seasonal patterns were found for the purchase of peanuts, pecans, and macadamia nuts. In the third quarter of year 2015, households purchased more almonds, pistachios, and mixed nuts. As household size increased, households tended to purchase more peanuts and tree nuts, except for walnuts and macadamia nuts. Income, as well as age of household head, were not found to affect the purchase of macadamia nuts.

Table 7. Joint tests associated with socio-demographic factors in the respective first-stage and second-stage models (Nielsen Homescan Panel, 2015).[1,2]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Peanuts	**Pecans**	**Almonds**	**Cashews**	**Walnuts**	**Macadamia nuts**	**Pistachios**	**Mixed nuts**
Joint tests – first stage								
Age of household head	1,556.89***	566.95***	86.11***	898.07***	1,440.01***	5.04	400.33***	1,512.05***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.17)	(0.00)	(0.00)
Education of household head	25.85***	52.45***	478.19***	60.18***	79.81***	16.12***	43.40***	102.33***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Region	429.80***	1,053.08***	82.32***	232.51***	380.74***	150.32***	174.93***	513.64***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Race	12.48***	124.09***	206.35***	36.75***	38.19***	51.83***	27.79***	13.57***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Joint tests – second stage								
Age of household head	611.08***	45.61***	60.89***	198.29***	121.10***	1.32	162.72***	74.55***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.72)	(0.00)	(0.00)
Education of household head	12.06***	22.34***	134.73***	115.87***	79.30***	1.15	24.07***	122.71***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.77)	(0.00)	(0.00)
Region	43.50***	46.43***	28.77***	150.47***	88.61***	24.35***	45.45***	236.87***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)***	(0.00)	(0.00)	(0.00)
Race	27.14***	33.66***	8.41**	33.68***	8.53**	11.29**	57.36***	48.79***
	(0.00)	(0.00)	(0.04)	(0.00)	(0.04)	(0.01)	(0.00)	(0.00)

[1] *P*-values in parentheses: * $P<0.1$, ** $P<0.05$, *** $P<0.01$.

[2] Joint tests are used to test socio-demographic coefficients jointly. The null hypotheses are: *agehh_under3* = *agehh_35to49* = *agehh_50to64* = 0; *eduhh_lesshigh* = *eduhh_highschool* = *eduhh_somecollege* = 0; *NewEngland* = *MiddleAtlantic* = *ENCentral* = *WNCentral* = *SouthAtlantic* = *ESCentral* = *WSCentral* = *Mountain* = 0; *White* = *Black* = *Asian* = 0.

Table 8. Marginal effects of unit values, seasonality, and socio-demographic factors associated with second-stage Heckman models for peanuts and tree nuts (Nielsen Homescan Panel, 2015).[1,2]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Peanuts	**Pecans**	**Almonds**	**Cashews**	**Walnuts**	**Macadamia nuts**	**Pistachios**	**Mixed nuts**
own unit value (price)	**-275.379**	**-44.836**	**-57.055**	**-56.982**	**-58.592**	**-8.970**	**-40.346**	**-39.777**
q2	-2.139	0.304	0.663	**1.751**	0.650	-0.285	1.108	0.578
q3	-0.946	1.006	**2.135**	1.064	1.235	-1.401	**1.682**	**1.855**
q4	-1.578	0.106	**1.684**	1.091	**-2.716**	0.002	0.052	-1.022
hsize	**2.404**	**0.936**	**1.130**	**0.953**	0.472	-0.172	**0.553**	**1.368**
Income	**0.059e-03**	**0.025e-03**	**0.042e-03**	**0.055e-03**	**0.034e-03**	0.010e-03	**0.048e-03**	**0.094e-03**
agehh_under35	**-18.997**	**-5.028**	**-6.603**	**-7.157**	**-7.170**	-2.329	**-7.303**	**-7.391**
agehh_35to49	**-11.854**	**-3.534**	**-4.864**	**-4.500**	**-4.863**	-1.070	**-4.039**	**-5.360**
agehh_50to64	**-5.022**	-0.711	**-3.068**	**-1.831**	**-2.139**	0.633	-0.699	**-1.062**
eduhh_lesshigh	-0.730	0.169	2.894	**-7.128**	-5.646	-0.690	-3.740	**-8.916**
eduhh_highschool	-1.031	-0.511	**-3.280**	**-3.975**	**-2.772**	**-2.537**	**-1.705**	**-5.126**
eduhh_somecollege	-1.669	-1.515	**-2.527**	**-3.252**	**-2.744**	-0.461	-0.687	**-5.366**
NewEngland	**-5.867**	**-5.132**	**-4.116**	**-8.356**	**-5.145**	-2.957	**-5.768**	**-12.419**
MiddleAtlantic	-0.998	-2.912	-1.629	**-7.209**	-1.663	-2.445	**-5.121**	**-12.472**
ENCentral	**-10.405**	**-3.550**	**-4.138**	**-6.353**	**-5.546**	-2.187	**-5.116**	**-10.779**
WNCentral	**-11.944**	**-3.712**	**-4.028**	**-6.817**	**-6.345**	**-4.726**	-1.833	**-8.109**
SouthAtlantic	**-10.977**	-2.820	**-6.011**	**-7.051**	**-4.065**	**-4.689**	**-5.293**	**-11.164**
ESCentral	**-14.271**	-1.879	**-5.529**	**-8.353**	**-4.534**	**-4.474**	**-6.371**	**-8.149**
WSCentral	**-12.153**	-2.529	**-5.961**	**-5.769**	**-5.825**	**-5.935**	**-2.605**	**-8.858**
Mountain	**-10.356**	-0.023	-1.165	-1.112	-0.518	**-7.316**	-1.639	**-2.868**
White	**4.439**	0.974	1.107	0.573	-2.066	-1.038	-2.261	-1.521
Black	-0.076	3.557	**5.238**	-2.161	1.073	0.566	-2.570	**-6.008**
Asian	5.219	0.325	**4.792**	**6.455**	**4.547**	**5.252**	**5.724**	3.133
hispanic_reg	-3.085	0.455	-1.637	-1.068	-1.374	2.136	**-3.673**	-1.423
no_child	**9.833**	2.246	**3.912**	**3.731**	**3.571**	1.879	**2.935**	**6.366**

[1] Numbers indicate significance level at 10% or lower. See Appendix S7 of the Supplementary Material for the detailed table with standard errors and significance level.

[2] Marginal effects are calculated by following Saha *et al.* (1997) and using the sample means of the respective covariates. We also account for the logarithmic transformation of household size and income in reporting their respective marginal effects.

Household heads who were college-educated purchased more almonds, cashews, walnuts, macadamia nuts, pistachios, and mixed nuts relative to household heads who were not college-educated. White households purchased more peanuts relative to other racial groups, while Black households purchased more almonds. Asian households purchased more cashews, walnuts, macadamia nuts, and pistachios relative to other racial groups. Ethnicity, Hispanic origin, had no impact on nut purchases. The presence of children under the age of 18 was found to be positively associated with the purchases of all nut products, except for macadamia nuts.

In general, older households purchased more nuts. Relative to the Pacific area, households located in the Mountain area purchased the least of peanuts, households located in the West South Central area had purchased the most of pecans, the least purchase of almonds was in East South Central, while the least purchase of cashews was in New England. Households in Middle Atlantic purchased the most walnuts, while households in East South Central purchased the least pistachios. Finally, the least purchase of mixed nuts was in the Middle Atlantic.

Based on the marginal effects associated with price and income, the own-price elasticities and income elasticities for the various nuts were calculated at the sample means. As shown in Table 9, the demands for pecans, almonds, and walnuts were estimated to be elastic, ranging from -1.26 to -1.45; the demands for cashews, macadamia nuts, pistachios, mixed nuts, and peanuts were estimated to be inelastic, ranging from -0.49 to -0.99. All nuts were found to be necessities with income elasticities less than 1, ranging from 0.06 to 0.16. The estimated household size elasticities varied from 0.05 (pistachios) to 0.11 (peanuts).

4.3 Evaluation of prediction success

We also provide evaluations of prediction success, also known as classification tables, for all first stage-probit models. These evaluations gauge how well the probit models predict the decision to purchase nuts. We categorize the predictions of all observations in the first-stage probit models into four scenarios in a 2×2 matrix defined as follows:

		Observed outcome (Z=0 or 1 buy or not)	
		1	0
Predicted outcome	1	true positive	false positive
(cut-off value)	0	false negative	true negative

The use of classifications of correct and incorrect responses are commonly used with the estimation of probit (or logit) models. To implement these evaluations, a cutoff value must be established initially. Subsequently, the predicted probability obtained from the probit model is compared to the aforementioned cutoff value. If the predicted probability is larger than the cutoff value, then we predict that the household purchased the respective nut product. Similarly, if the predicted probability is less than the cutoff value, then we predict that the household did not purchase the respective nut product.

Table 9. Conditional own-price elasticities, income elasticities, and household size elasticities calculated at the sample means (Nielsen Homescan Panel, 2015).[1,2]

	Own price elasticity	**Income elasticity**	**Household size elasticity**
Peanuts	-0.995***	0.070***	0.114***
	(0.0301)	(0.015)	(0.025)
Pecans	-1.453***	0.077*	0.111*
	(0.112)	(0.040)	(0.064)
Almonds	-1.261***	0.109***	0.108***
	(0.050)	(0.023)	(0.036)
Cashews	-0.900***	0.114***	0.076***
	(0.037)	(0.014)	(0.022)
Walnuts	-1.341***	0.083***	0.044
	(0.063)	(0.023)	(0.037)
Macadamia nuts	-0.765***	0.057	-0.036
	(0.164)	(0.064)	(0.094)
Pistachios	-0.973***	0.116***	0.052*
	(0.043)	(0.018)	(0.029)
Mixed nuts	-0.490***	0.162***	0.088***
	(0.0212)	(0.013)	(0.020)

[1] Standard errors given in parentheses were obtained using the delta-method; * P<0.1, ** P<0.05, *** P<0.01.

[2] Elasticities were derived from the product of the marginal effects of price, income, and household size depicted in Table 8 times the ratio of price, income, or household size at the sample means to the conditional mean of the dependent variable.

The cutoff values vary for each nut category, and these cut-off values correspond to the number of households who purchased a given nut product out of the total number of households (Figure 1). For example, 23% of households purchased peanuts. Hence, in the derivation of the prediction-success (Table 10), the cutoff probability for classification purposes is 0.23. That is, we predict that a given household will purchase peanuts if the probability of doing so exceeds 0.23. In agreement with Greene (2012: 658), 'in general any prediction rule will make two types of errors; it will incorrectly classify zeros as one and ones as zeros.' For binary choice models, to the best of our knowledge, no benchmark exists regarding the percentage of correct classifications.

Table 10. Prediction-success evaluations for the respective pooled probit models (Nielsen Homescan Panel, 2015).[1,2]

	Outcome (Z=1)	**Outcome (Z=0)**		
	Peanuts (cut-off = 0.23)		Pecans (cut-off = 0.07)	
Predicted prob. > cut-off	30,580 (true positive)	83,489 (false positive)	11,758	83,208
Predicted prob. < cut-off	25,929 (false negative)	105,522 (true negative)	6,372	144,182
Sensitivity	54.12%		64.85%	
Specificity	55.83%		63.41%	
Correctly classified	55.43%		63.51%	
	Almonds (cut-off = 0.13)		Cashews (cut-off = 0.12)	
Predicted prob. > cut-off	17,073	96,862	14,652	95,903
Predicted prob. < cut-off	14,554	117,031	14,203	120,762
Sensitivity	53.98%		50.78%	
Specificity	54.71%		55.74%	
Correctly classified	54.62%		55.15%	
	Walnuts (cut-off = 0.09)		Macadamia nuts (cut-off = 0.01)	
Predicted prob. > cut-off	12,240	81,053	556	12,609
Predicted prob. < cut-off	9,422	142,805	836	231,519
Sensitivity	56.50%		39.94%	
Specificity	63.79%		94.84%	
Correctly classified	63.15%		94.52%	
	Pistachios (cut-off = 0.07)		Mixed nuts (cut-off = 0.01)	
Predicted prob. > cut-off	9,594	117,913	16,454	101,563
Predicted prob. < cut-off	8,373	109,640	11,245	116,258
Sensitivity	53.40%		59.40%	
Specificity	48.18%		53.37%	
Correctly classified	48.56%		54.05%	

[1] The number of observations is 245,520. The cut-offs for respective nuts are equivalent to the number of households who purchased a given nut product out of the total number of households in the sample (also known as 'market penetration'). The columns are actual outcomes, 0 or 1, and the rows are predicted outcomes. The numbers in each entry correspond to the count of observations that fall into each scenario.

[2] Sensitivity is defined as the proportion of observed purchases that were correctly predicted to occur. Specificity is defined as the proportion of observed non-purchases that were correctly predicted not to occur.

We also calculated the sensitivity, defined as the proportion of observations where purchases occurred that were predicted to occur, as well as the specificity, defined as the proportion of observed non-purchases that were predicted not to occur. To illustrate, within-sample the probit model correctly classified the decision to not make purchases of peanuts with 55.83% accuracy (105,522 out of 189,011), which is the specificity. Within-sample, the probit model correctly classified the decision to make purchases of peanuts with 54.12% accuracy (30,580 out of 56,509), which is the sensitivity. Overall, within-sample, the model correctly classified all decisions 136,102 out of 245,520 times, with 55.43% accuracy (prediction-success rate).

As shown in Table 10, the prediction-success rate (correct classification of all decisions) ranged from 48.56% (pistachios) to 94.52% (macadamia nuts); the sensitivity varied from 39.94% (macadamia nuts) to 64.85% (pecans); and the specificity ranged from 48.18% (pistachios) to 94.84% (macadamia nuts).

5. Conclusions and implications

This study aims to understand the purchase decision of U.S. households concerning peanuts and tree nuts as well as their purchase behavior of the respective nut categories. Pooled probit models were estimated to determine the factors affecting the decision to purchase or not to purchase various nuts. We subsequently estimate the key determinants of the quantities purchased based on the pooled Heckman sample selection models.[6] From this investigation, we estimate own-price elasticities and income elasticities as well as the impacts of socio-demographic characteristics concerning the demand for peanuts and tree nuts.

Older households, well-educated households, more wealthy households, and households without children were most likely to purchase peanuts and tree nuts. The propensity to purchase nut products was different across regions, race, and ethnicity. For the most part, the propensity to purchase peanuts and tree nuts was higher in the fourth quarter of the year.

However, the quantities purchased of nut products did not have a uniform seasonal pattern. Households with higher income levels, no children under the age of 18, older household heads, and college-educated household heads generally purchased more nut products. Households located in different regions and from different racial groups also were important factors concerning the amount purchased of the respective nut products.

Based on the own-price elasticities, the demands for pecans, almonds, and walnuts were elastic varying from -1.26 to -1.45; consequently, to increase revenue in the short term, holding all other factors constant, nut purveyors should lower the prices of pecans, almonds, and walnuts. On the other hand, the demands for cashews, macadamia nuts, pistachios, mixed nuts, and peanuts were inelastic, ranging from -0.49 to -0.99; hence, to increase revenue in the short run, holding all other factors constant, nut purveyors should raise the prices of cashews, macadamia nuts, pistachios, mixed nuts, and peanuts. All nut products were necessities with income elasticities less than 1, ranging from 0.06 to 0.16. Therefore, a 10% change in household income in either direction would generate a 0.6% change to a 1.6% change in quantities purchased of peanuts and tree nuts.

These findings contribute to the literature by providing more up-to-date information concerning factors affecting not only the likelihood of purchasing nuts but also the amounts purchased. As well, our research provides a more granular analysis, concerning nuts. From this analysis, stakeholders in the nut industry are in a position to develop profiles of U.S. households who purchase peanuts and various tree nuts. These profiles are instrumental for stakeholders in the industry to initiate target marketing and product positioning strategies to increase nut sales. To illustrate, profiles of households for selected tree nut products are exhibited in Figure 5. Simply put, stakeholders associated with the nuts industry should target elderly households, more educated households, more wealthy households, and households without children. As well, differences in the propensity to purchase peanuts and tree nuts are evident according to race, ethnicity, and region.

[6] The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

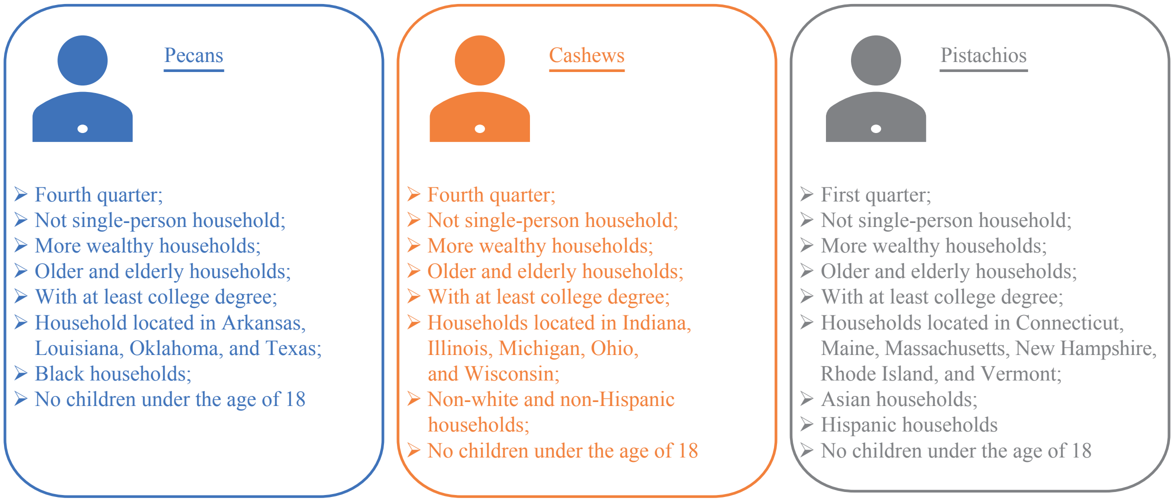

Figure 5. Household profiles for selected tree nuts in the United States (Nielsen Homescan Panel, 2015).

Federal price and income support programs, aside from Federal food purchase and donation programs, do not directly cover tree nuts. Nevertheless, our research can assist industry stakeholders in developing a coordinated program of marketing and outreach efforts not only to maintain but also to increase market exposure for tree nuts.

Limitations of our research are threefold. First, our study does not consider away-from-home purchases. The Nielsen data pertain exclusively to at-home purchases made by households from grocery stores, convenience stores, supercenters, drugstores, and mass merchandisers. Finally, the data correspond to calendar year 2015. Hence, to check on the robustness of the results, this study should be updated with more recent data. This analysis then in essence serves as a baseline or benchmark for future research. Finally, the research fails to capture the impacts of advertising and promotion on the likelihood of purchasing nuts as well as on the quantities purchased. Despite these limitations, our study nevertheless adds to the literature by providing a micro-perspective analysis concerning the propensity to purchase peanuts and tree nuts as well as the household demand for peanuts and tree nuts in the United States.

Supplementary material

Supplementary material can be found online at https://doi.org/10.22434/IFAMR2020.0090

Appendix S1. Robustness check #1 Heckman vs Tobit.
Appendix S2. Price imputation.
Appendix S3. Robustness check #2 Multicollinearity.
Appendix S4. Robustness check #3.
Appendix S5. Complete table for Table 5 in the manuscript with standard errors.
Appendix S6. Complete table for Table 6 in the manuscript with standard errors.
Appendix S7. Complete table for Table 8 in the manuscript with standard errors.

References

Alviola IV, P.A. and O. Capps Jr. 2010. Household demand analysis of organic and conventional fluid milk in the United States based on the 2004 Nielsen Homescan panel. *Agribusiness* 26(3): 369-388.

Aranceta, J., C.P. Rodrigo, A. Naska, V.R. Vadillo and A. Trichopoulou. 2006. Nut consumption in Spain and other countries. *British Journal of Nutrition* 96(S2): S3-S11.

Belsley, D.A., E. Kuh and R.E. Welsh. 1980. *Regression diagnostics: identifying influential data and sources of collinearity*. John Wiley & Sons, New York, NY, USA.

Bliss, C.I. 1934a. The method of probits. *Science* 79: 38-39.

Bliss, C.I. 1934b. The method of probits – a correction. *Science* 79: 409-410.

Capps Jr., O., R. Tsai, R. Kirby and G.W. Williams. 1994. A comparison of demands for meat products in the Pacific Rim region. *Journal of Agricultural and Resource Economics* 19(1): 210-224.

DeNavas-Walt, C., B.D. Proctor and J.C. Smith. 2011. *U.S. Census Bureau, current population reports, P60-239, income, poverty, and health insurance coverage in the United States: 2010.* U.S. Government Printing Office, Washington, DC.

Dhaliwal, H.S. 1972. *An econometric investigation of demand interrelationships among tree nuts and peanuts.* PhD-thesis, Oregon State University, Corvallis, OR, USA.

Dharmasena, S. and O. Capps, Jr. 2014. Unraveling demand for dairy-alternative beverages in the United States: the case of soymilk. *Agricultural and Resource Economics Review* 43(1): 140-157.

Dietary Guidelines Advisory Committee, 2015. *Dietary guidelines for Americans 2015-2020*. Government Printing Office, Washington, DC, USA.

Fraser, G.E., J. Sabate, W.L. Beeson and T.M. Strahan. 1992. A possible protective effect of nut consumption on risk of coronary heart disease: the Adventist health study. *Archives of Internal Medicine* 152(7): 1416-1424.

Greene, W.H. 2012. *Econometric analysis*, 7th edition. Pearson, New York, NY, USA.

Haines, P.S., D.K. Guilkey and B.M. Popkin. 1988. Modeling food consumption decisions as a two-step process. *American Journal of Agricultural Economics* 70(3): 543-552.

He, S., S.M. Fletcher and A.P. Rimal. 2005. Snack peanut consumption: type preference and consumption manners. *Journal of Food Distribution Research* 36(1): 79-85.

Heckman, J.J. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5(4): 475-492.

Hill, H. and F. Lynchehaun. 2002. Organic milk: attitudes and consumption patterns. *British Food Journal* 104(7): 526-542.

King, J.C., J. Blumberg, L. Ingwersen, M. Jenab, and K.L. Tucker. 2008. Tree nuts and peanuts as components of a healthy diet. *The Journal of Nutrition* 138(9): 1736S-1740S.

Kris-Etherton, P.M., F.B. Hu, E. Ros and J. Sabaté. 2008. The role of tree nuts and peanuts in the prevention of coronary heart disease: multiple potential mechanisms. *The Journal of Nutrition* 138(9): 1746S-1751S.

Kyureghian, G., O. Capps, Jr. and R.M. Nayga, Jr. 2011. A missing variable imputation methodology with an empirical application. *Advances in Econometrics* 27(1): 313-337.

Lee, I.M. 1950. *Statistical analysis of annual average returns to growers of almonds, 1924/25 to 1948/49.* College of Agriculture, University of California, Davis, CA, USA.

Leong, Y.H., N. Ismail, A.A. Latiff, A. Nurul Izzah, M.Y. Narazah and A.B. Nurul Ain. 2011. Nuts consumption pattern among Malaysian adults: a socio-demographic and dietary behaviour perspective. *International Food Research Journal* 18(1): 319-328.

Lerner, E.B. 1959. *An econometric analysis of the demand for pecans with special reference to the demand interrelationships among domestic tree nuts*. PhD-thesis, Oklahoma State University, Stillwater, OK, USA.

Mattes, R.D., P.M. Kris-Etherton and G.D. Foster. 2008. Impact of peanuts and tree nuts on body weight and healthy weight loss in adults. *The Journal of Nutrition* 138(9): 1741S-1745S.

Nielsen Homescan Panel. 2015. *Homescan consumer panel data [data set].* Kilts Center for Marketing, Chicago, IL, USA.

Puhani, P., 2000. The Heckman correction for sample selection and its critique. *Journal of Economic Surveys* 14(1): 53-68.

Rimal, A.P. and S.M. Fletcher. 2002. Snack peanuts purchase pattern: effects of nutritional considerations and household characteristics. *Journal of Agricultural and Applied Economics* 34(1): 51-65.

Ros, E. 2010. Health benefits of nut consumption. *Nutrients* 2(7): 652-682.

Russo, C., R. Green and R.E. Howitt. 2008. Estimation of supply and demand elasticities of California commodities. *SSRN.* http://dx.doi.org/10.2139/ssrn.1151936

Sabaté, J. 2003. Nut consumption and body weight. *American Journal of Clinical Nutrition* 78(3): 647S-650S.

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

Spence, A.M. 1973. *Price, quality, and quantity interdependencies*. Memorandum No. 161, Center for Research on Economic Growth, Stanford University, Stanford, CA, USA.

Trajtenberg, M. 1989. The welfare analysis of product innovations, with an application to computed tomography scanners. *Journal of Political Economy* 97(2): 444-479.

U.S. Census Bureau. 2010. *Census of population and housing 2012*. *Population and housing unit counts, CPH-2-1, United States summary.* U.S. Government Printing Office, Washington, DC, USA.

U.S. Census Bureau. 2020. *Current population reports.* U.S. Census Bureau, Washington, DC, USA. Available at: https://www.census.gov/main/www/cprs.html

Veall, M.R. and K.F. Zimmermann. 1996. Pseudo-R2 measures for some common limited dependent variable models. *Journal of Economic Surveys* 10(3): 241-259.

Wells, G.J., S.E. Miller and C.S. Thompson. 1986. Farm-level demand for pecans reconsidered. *Journal of Agricultural and Applied Economics* 18(1): 157-160.

Wooldridge, J.M. 1995. Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics* 68: 115-132.

Wooldridge, J.M. 2010. *Econometric analysis of cross section and panel data.* MIT Press, Cambridge, MA, USA.