Consumers Apple Variety Choices Based On National Household-Level Data

Qiujie Zheng[1], Vicki A. McCracken[2], Mykel R. Taylor[3]

[1]School of Economic Sciences, Washington State University, Pullman, WA 99164

(email: qiujiezheng@gmail.com, phone: 415-827-4854)

[2]School of Economic Sciences, Washington State University, Pullman, WA 99164

(email: mccracke@wsu.edu, phone: 509-335-4728)

[3]School of Economic Sciences, Washington State University, Pullman, WA 99164

(email: m_taylor@wsu.edu, phone: 509-335-8503)

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association 2010 AAEA,CAES, & WAEA Joint Annual Meeting, Denver, Colorado, July 25-27, 2010

Consumers Apple Variety Choices Based On National Household-level Data

Qiujie Zheng, Vicki A. McCracken, Mykel R. Taylor

School of Economic Sciences, Washington State University

Introduction

Consumers face choices in purchasing fruits and, even within a particular fruit category, they face a number of options including variety decisions. This is particularly relevant for fresh apples, as there are a large number of different varieties available for consumers to choose from. We consider different varieties of apples as differentiated products and examine the factors that affect consumer fresh apple purchase decisions.

Demographics such as income, age, education, race, and the presence of children are all contributing factors to the revealed purchase decision, either through their impact on preferences and/or budget constraints. National or state level aggregate consumption data on apples, even if available by variety, does not allow for research on the relative impacts of these personal characteristics. Use of household-level ACN Homescan data will allow researchers to overcome the restrictions of aggregate data and investigate the impact of household characteristics, as well as seasonality and geographic location effects.

Consumer apple variety choices can be expressed as the valuation of apple variety-specific internal attributes such as firmness, crispness, juiciness, sweetness and tartness. Preferences for internal apple attributes constitute part of the reason why consumers decide to purchase a specific variety. Apple breeders are especially interested in consumers' valuation of the variety-specific internal attributes. It is critical that their breeding programs align so that they are developing varieties that meet market needs and consumer preferences.

This paper combines prices, household demographics, and state dependence variables, as well as regional and seasonal effects to analyze household apple variety choices and purchases. The objectives of our overall project are to a) identify the important factors that affect consumers' variety choices; b) estimate demand elasticities for different varieties of apples; and c) share implications of the study with apple breeders for their breeding prioritization of apple traits that are valued by consumers. This paper focuses on the first objective.

Literature

It is advantageous to use household data to analyze apple variety choice problems since it provides a large sample with rich demographic information. However, there are estimation issues to address when using household level data because households who do not consume apples are censored in the dataset.

Tobin (1958) first showed that direct use of the ordinary least squares (OLS) method on a censored response variable will cause bias and inconsistency. The use of a Tobit model restricts the factors to have the same impact on the purchase decision and the purchase quantity decision. Allowing a separate estimation of the purchase and quantity decisions, Heckman (1979) developed a two-step single equation model in which an inverse Mills ratio is calculated from the first step based on the probability of purchase and used as an instrument in the second step regression on the uncensored observations.

The problem of limited dependent variables is particularly complicated in a multivariate system model. Amemiya (1974) developed a computationally simple but consistent estimator for multivariate and simultaneous equation models with truncated dependent variables. There are a number of censored demand system estimators in the literature, such as Wales and Woodland

(1983), and Lee and Pitt (1986, 1987). But direct maximum likelihood estimation of these systems is computationally demanding when censoring occurs in multiple equations.

Lee (1978) generalized the single equation two-step Amemiya estimator to a multiple equation model. Lee showed that the two-stage estimators resulting from this procedure are asymptotically more efficient than other two-stage estimators. Heien and Wessells (1990) (henceforth HW) employed a two-step method by using the whole sample (including the censored observations) in both steps to estimate a system of demand equations for a group of food commodities. In the first step, a probit maximum likelihood is used on each commodity to estimate the purchase or not decision. An inverse Mills ratio is calculated as the specification of omitted variable which can be used on the censored sample in the second step to alleviate the sample selection bias. In the second step, a complete demand system is estimated using all the observations by seemingly unrelated regression (SUR) with inverse mills ratio as an instrument variable. This procedure has been used extensively in the empirical literature, such as Heien and Durham (1991), Gao *et al* (1995), Nayga (1996), Wang *et al* (1996) and Byrne *et al* (1996).

Vermeulen (2001) also pointed out that the HW approach yields inconsistent estimates of the regression coefficients. Because of the incorrect unconditional mean expressions for the censored dependent variables, the results are not different from the OLS estimation on all the observations.

Tauchmann (2008) showed that the multivariate generalizations to the classical Heckman two-step approach that account for cross-equation correlation, and use of the inverse Mills ratio as a correction term are consistent only if certain restrictions apply to the true error-covariance structure. He considered three variants of this estimator and concluded the debate on which estimator is the best choice for estimating the multivariate sample selection model. If efficiency is the major concern and numerical computing time does not matter, the full information maximum likelihood (FIML) instead of two-step method is the better choice. If computational simplicity and consistency is the major concern, the equation-by-equation Heckman appears to be the best choice. If a small mean square error and computational simplicity are the major concerns, one might even argue in favor of the inconsistent SUR estimator that conditions equation-by-equation on the outcome of the upstream choice problem. If both consistency and a small mean square error are desired and the computational burden of FIML is to be avoided, the GLS estimator that conditions on the entire selection pattern is the best choice. Following Tauchmann, we use equation-by-equation Heckman estimator, which is computationally simple and consistent.

Data

The data used for this paper are from the ACN Homescan panel of U.S. households. The data consist of retail purchases of at-home foods. Panel participants scan in their purchase information at home after they finish shopping. The purchase data are then uploaded to the ACN computer system. This information is paired with the household's demographic information provided when they signed up with the program. The dataset is drawn from a sample of households that are demographically balanced within 19 markets and 4 Census regions in the United States, which are claimed to be fully representative of U.S. households. Household level purchase data and demographic information were included for 7,195 household panelists who were in the sample during at least 10 out of 12 months beginning January 3, 1999 through January 1, 2000.

The dataset contains 728,777 household-level observations on fresh apple purchases (in pounds) from 1998 to 2006. After removing two observations with missing demographic information and all the duplicates, we cut off the lowest 1% volume (i.e. volume less than 0.24

pounds and the highest 1% price (i.e. price larger than 4.44 dollars/pound). After this data cleaning procedure, 709,403 observations remained for analysis.

Because of our interest in varietal purchase decisions, we break the apple purchases down to the five most commonly purchased varieties of apples. These are Red Delicious (28.94% of all observations), Gala (14.04%), Granny Smith (11.41%), Golden Delicious (10.35%) and McIntosh (5.07%). The remaining apple purchases fall into the other category (30.19%).

There is rich demographic information for each observation in this dataset. We used purchase date, household size, income, age, presence of kids, education, race and region variables in our analysis. The Nielsen data divides income into 19 categories. We define the household income variable (*inc*) by the midpoints of each category and setting the over $200,000 category as $225,000. A similar method is used for the age variable which has nine categories. We define the age variable (*hhage*) as the midpoint of each category and set the under 25 category as 20 and the over 65 category as 70. The three education categories include individuals with a high school education (*hhed_hs*), some college or a college education (*hhed_col)* and post college education (*hhed_pc)*. For the age and education variables, the female head of household information is used if there is female head in the household; otherwise the male head of household information is used. Individuals are divided into race/ethnic groups defined as white, black, Asian, Hispanic and other. The regions are divided into east, central, south and west. Urban is a binary variable to distinguish urban or rural areas. Year and month dummies are created from purchase dates to account for seasonal purchase effects.

On the basis of censoring rate for consumer apple purchase behavior, we aggregate the purchase information to a monthly level. The expenditures and purchased quantities are aggregated to a monthly level to calculate a monthly average price.

On a monthly level, the censoring rate for all apple purchase is 65.81% with that for each variety very high (Red Delicious 87.39%, Gala 93. 73%, Granny Smith 94.67%, Golden Delicious 95.28%, McIntosh 97.57%, and Other 87.89%). Checking the monthly apple purchase frequency, we found that the mean times of monthly apple purchase are 4.10. Due to the high censoring rate, we decided to cut off the households if they purchase apples less than 4 months per year. This removed 42.35% of the observations, but resulted in a lower censoring rate (49.15% for all apples, 81.11% for Red Delicious, 90.3% for Gala, 92.04% for Granny Smith, 92.68% for Golden Delicious, 96.41% for McIntosh, and 81.27% for Other).

Method

In this paper, we use a two-step estimation method using a probit model for each variety to estimate a consumer's decision to purchase apples or not. The inverse Mills ratio is calculated from this probit model and used in the second step as an instrument. Based on Tauchmann (2008), we choose to use the equation-by-equation OLS on uncensored observations, which is computationally simple and consistent as mentioned previously.

We model the apple variety purchase decision and quantity decision as a system of equations with limited dependent variables as follows:

$$
\begin{aligned}
&y^{*}_{ijt} = x'_{ijt}\beta_i + \varepsilon_{ijt} \\
&d^{*}_{ijt} = z'_{ijt}\alpha_i + v_{ijt} \\
&d_{ijt} = \begin{cases} 1 & \text{if} \quad d^{*}_{ijt} > 0 \\ 0 & \text{if} \quad d^{*}_{ijt} \le 0 \end{cases} \\
&y_{ijt} = d_{ijt} y^{*}_{ijt}
\end{aligned}
\qquad (1)
$$

where for the ith equation, the jth household and the tth observation, y_{ijt} and d_{ijt} are the observed dependent variables of apple variety purchase quantity and purchase decision, respectively; y^*_{ijt} and d^*_{ijt} are the corresponding latent variables; x_{it} and z_{it} are vectors of exogenous variables, such as prices, demographics, seasons and regions; β_i and α_i are conformable vectors of parameters; ε_{it} and v_{it} are random errors. We assume $\varepsilon_t = [\varepsilon_{1jt}, \ldots, \varepsilon_{mjt}]'$ and $v_t = [v_{1jt}, \ldots, v_{mjt}]'$ are normally distributed random vectors with covariance matrix $\mathrm{var}(\varepsilon_t, v_t) = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma'_{\varepsilon v} \\ \Sigma_{\varepsilon v} & \Sigma_{vv} \end{bmatrix}$.

The data is an unbalanced panel of households over 9 years because all households are not observed over the entire time period (households rotate out of the panel on an annual basis). Due to the heterogeneous nature of the population and the importance of unobservable variables, a random effects probit model is used to model the unobserved effects on purchase decisions.

Summary statistics for all the explanatory variables are presented in Table 1. A random effects probit model as in equation (1) is used for each of the six apple varieties to estimate the impact of various factors on consumers apple varietal purchase decisions. Prices (imputed for censored observations), household size, income and its squared term, household age, education level, race, urban/rural, regions, year and month dummies and state dependent variables are used to determine the probability of making a specific apple variety purchase.

Due to censoring, retail prices are not observed if the household didn't purchase apples in a specific month. Therefore, a market price must be imputed for the censored observations. We assume the prices vary by season, region, and urban versus rural locations. Since household characteristics also play a role in on the price that a household would actually pay, household income, household size and their square terms are also considered as explanatory variables. We regressed observed prices for each apple variety on the year, month, region dummies, urban/rural binary variable, household income, household income squared, household size, and household size squared, and then used the estimated coefficients to impute a price for the censored observations.

When households repeatedly purchase specific apple varieties, their past choice can affect the probability of choosing the specific varieties again. Following Moeltner and Englin (2004), we consider this state dependence effect by constructing four state dependence variables. We denote *sqtot* as total apple variety monthly purchase count for a household during his/her participation to the current time point. We denote *sq_run* as number of consecutive apple variety monthly purchase count to the current time point. We also define two analogous variables for monthly non-purchase of specific variety of apples as *sq_np* and *sq_npr*. There will be subscripts following the variable names to denote different varieties.

In our apple variety model, the conditional mean of purchase can be expressed as:

$$E\left[y_{ijt}\,|x_{ijt}, z_{ijt}; d_{ijt} = 1\right] = x'_{ijt}\beta_i + \delta_i \frac{\phi(z'_{ijt}\alpha_i)}{\Phi(z'_{ijt}\alpha_i)} \tag{2}$$

Hence the second step estimation will be:

$$y_{ijt} = x'_{ijt}\beta_i + \delta_i \frac{\phi(z'_{ijt}\hat{\alpha}_i)}{\Phi(z'_{ijt}\hat{\alpha}_i)} + \xi_{ijt} \tag{3}$$

Due to the panel characteristics of our data, we also use random effects OLS model to estimate the factors effect on per capita apple variety purchase quantity. The variables in x_{ijt} are purchase prices, household income and its squared term, age, presence of kids, education, race, urban/rural, region and seasonal dummies. The inverse Mills ratio $\phi(z'_{ijt}\hat{\alpha}_i)/\Phi(z'_{ijt}\hat{\alpha}_i)$ calculated from the first step probit is included in the model as an instrumental variable. We utilize the random effects OLS regression[1] on the uncensored observations separately for each varietal purchase in the second step.

The purpose of this paper is to analyze the factors which impact on consumers apple variety choice and to address consumers preference for the internal attributes of specific varieties. Table 2 summarizes the apple variety characteristics including internal attributes, color, function and availability for the varieties analyzed in this paper. In our results, we will analyze and compare the impact of different factors by variety.

Results

First Step Probit Model

The results of the random effects probit model are reported in Table 3 for each apple variety. The coefficients for price of granny smith (*pp_gs*), golden delicious (*pp_gd*), gala (*pp_ga*) and other (*pp_ot*) varieties are negative and significant at the 5% level. Coefficients of household size (*hhsize*) are positive and significant at the 5% level for all varieties except for the other varieties model. This indicates that larger sized households tend to have a higher probability of purchasing these five apple varieties given all else constant. Coefficients for the income and income squared variables are significant at the 5% level for gala, granny smith and other, and are significant at the 10% level for red delicious. The income factors do not have a significant effect on purchase decisions of golden delicious and McIntosh.

Household age coefficients are positive and significant at the 5% level for red delicious, gala, McIntosh and other. This indicates that older households, all other variables in our model held constant, are more likely to purchase the varieties with sweet attributes such as the red delicious and gala varieities and also the varieties with tender attributes such as McIntosh. The impact of age on the probability of purchase is negative and significant at the 5% level for

[1] We tested H_0: Pool OLS model vs. H_a: Random Effects model. The null hypothesis is rejected for all varieties.

granny smith apples. This is a variety with a very tart flavor suggesting that older consumers do not prefer such tart apples. The coefficients for presence of kids are significantly positive at the 5% level for red delicious, granny smith and golden delicious, hence suggesting that these are popular varieties for households with kids. Education was significant for all varieties except McIntosh. In particular, the higher the household educational level, the more likely the household will purchase gala, granny smith and other varieties. In contrast, higher educated households are less likely to purchase red delicious and golden delicious. Regarding race/ethnicity effects, black households have significantly different purchase probabilities compared to white households for all varieties. Black households have higher purchase probability for red delicious and golden delicious than white households and a lower purchase probability for gala, granny smith and golden delicious. Asian households have significantly lower purchase probability for red delicious, granny smith and McIntosh compared to white households. Hispanic households have higher purchase probabilities for red delicious and lower probabilities for McIntosh compared to otherwise equivalent white households.

Households located in urban areas tend to have a higher probability of purchasing gala and McIntosh apples than rural households. All varieties are affected by the regional location of households. Households in the central region are more likely to purchase red delicious than households in any other region. And relative to equivalent households in the central region, households in the western region are more likely to purchase gala and granny smith. Households in the south are more likely to purchase golden delicious, and those in the east are more likely to purchase McIntosh, all in comparison to their counterparts in the central region.

Seasonal effects measured by year and month dummies, are statistically significant at the 5% level. From year 1998 to year 2006, consumers have a statistically significantly lower probability of purchasing red delicious, granny smith and golden delicious in each subsequent year.. The gala purchase probability is higher each year from 1998 to 2002, obtains a peak at 2002, and then decreases each year after 2002 but it is still higher than the 1998 level. Due to differences in harvest schedules and relative storability of each variety, most of the monthly effects are significant but there are different patterns for each variety. Consumers are more likely to purchase red delicious apples in March, galas in September, and granny smiths, golden delicious and McIntosh in October.

All the state dependent variables for all varieties have statistically significant impacts on the probability of purchase decision at the 5% level. Except McIntosh, all the signs of the state dependent variables are as expected. The significantly positive total purchase counts accumulation and negative total non-purchase counts accumulation indicate that a consumer's historically varietal purchase preference has positively affected their current point purchase decisions. The significantly positive consecutive purchase counts and negative consecutive purchase counts represent that a consumer's current purchase decision depends on their consecutive purchase habits and the current purchase decisions depend on the past consecutive purchase behavior.

Second Step OLS Model

Using the Heckman two-step method, we allow the set of factors that affects consumers varietal purchase decision to differ from those affecting consumers quantity of purchase. Hence in the second step, we analyze the impact that factors have effect on per capita varietal purchase quantity, including the inverse mills ratio calculated from the first step coefficients as an instrument variable.

Table 4 reports the coefficients of factors affecting per capita quantity of purchase in the second step random effects OLS model estimated separately for each variety. The coefficients of prices for all varieties are negative and significant at the 5% level. The coefficients of income and income squared terms are both significant for all varieties. Since we have squared terms of income, we calculate marginal income effect to compare the income effects among varieties. The marginal income effects are -0.0912 for red delicious, -0.0604 for gala, -0.0546 for granny smith, -0.0937 for golden delicious, -0.1003 for McIntosh, and -0.0859 for other. The fact that higher income causes less apple purchase quantity per capita may due to the substitution of other fruits.

The coefficients for household age are positive and significant at the 5% level for red delicious and gala which means that older household tends to purchase more red delicious and gala per capita, varieties which are sweeter. The coefficients of kid presence are negative and significant at the 5% level for all varieties indicating the presence of kids decrease the quantity per capita purchase of apples. In probit model, the coefficients of kids presence are positive and significant at the 5% level for red delicious, granny smith and golden delicious. The signs are switched from the probit model to the OLS model which indicates that households with kids are more likely buy red delicious, granny smith and golden delicious apples but the kids eat fewer apples than the adults. The educational level coefficients are all positive and significant indicating more highly educated households purchase higher quantities per capita of all varieties than their otherwise equivalent less educated counterparts. For red delicious, granny smith, McIntosh and other, only Hispanic households purchase lower quantities than their white counterpart households. There are no race/ethnicity differences for the other race/ethnicity groups for these same varieties. Hispanic and other race/ethnic groups have very different purchase behaviors for the gala variety. For golden delicious variety, black, Asian and other race/ethnic households all differ significantly from white households in their per capita purchase behavior.

Considering regional effects, the urban/rural differences exist only for gala apples at the 5% level, in which case urban households tend to buy more gala per capita than their rural counterparts. For red delicious and McIntosh, there are obvious regional effects with the quantity per capita purchase being significantly different for households from the east and west relative to households from the central region of the country. For gala apples, households from the south and west consumed significantly lower per capita quantities than their central region counterparts.

Consistent with the finding for the probit models, there are significant time impacts, both annual and seasonal, on the per capita purchase of specific varieties. Both gala and golden delicious varieties experienced a decreasing and then gradual increasing pattern of purchase quantity per capita from 1998 to 2006. Granny Smith has an increasing trend during the 9 years. The monthly effects are obvious for all apple varieties. The red delicious variety has the highest quantity per capita purchase in January and lowest in September. Per capita purchases of gala are highest in January and lowest in August. Granny Smith purchases are highest in May and lowest in September. Golden delicious purchases are highest in January and lowest in September. And McIntosh apple purchases are highest in October and lowest in June. These different seasonal patterns across varieties likely correlate with the storage quality of the variety. The inverse mills ratio for all the varieties are all negative and significant at the 5% level indicating that the correction factor (for censoring) is necessary to obtain consistent estimates for the coefficients in the per capita purchase quantity apple variety models.

Conclusions

The models estimated in this research provide empirical evidence to help us understand the relative importance of different product-specific and household characteristics on apple varietal purchase decisions. Price plays an important role in both the purchase and quantity decisions of a household. Purchases of red delicious apples appear to be particularly price sensitive. The effect of household income on purchase quantity is negative, possibly due to substitution effects from other apple varieties or other fruits. No other apple variety considered in the model is negatively affected by increases in income, suggesting that Red Delicious apples may be considered an inferior good by households that purchase apples. The finding that age of household head and presence of kids in a household had different impacts on purchases by varieties can be explained and related to the internal attributes and uses/functions of specific apple varieties. Other demographics such as educational level and race/ethnicity had differential impacts on household apple varietal purchases. Some regional affects were found, for some varieties. The time impacts, represented by year and month dummies, are significant but differed by variety because their storability varies which will affect their quality late in the season or availability early in the season.

Future Research

Because of the correlation between different apple variety choices and purchases, a system approach to estimation still needs to be considered to gain efficiency in the parameter estimates. The aggregation to monthly level needs to be adjusted to better apply to the storability and purchase frequency of the apples. Intrinsic characteristics can be incorporated into a hedonic pricing model for additional information concerning what influences the price paid for an apple. All of these will be considered in our future research efforts.

References

Agüero J. M., and B. W. Gould. 2003. "Household Composition and Brazilian Food Purchases: An Expenditure System Approach." *Canadian Journal of Agricultural Economics* 51: 323-345.

Byrne, P. J., O. Capps, Jr., and A. Saha. 1996. "Analysis of Food-Away-from-Home Expenditure Patterns for U.S. Households." *American Journal of Agricultural Economics* 78: 614-27.

Gao, X. M., E. J. Wails, and G. L. Cramer. 1995. "A microeonometric model analysis of US consumer demand for alcoholic beverages." *Applied Economics* 27: 59-69.

Harris, J. M., and N. Blisard. "Characteristics of the Nielsen Homescan Data."

Heckman, J. 1979. "Sample selection bias as a specification error." *Econometrica* 47: 153-61.

Heien, D., and C. Durham. 1991. "A Test of the Habit Formation Hypothesis Using Household Data." *The Review of Economics and Statistics* 73: 189-199.

Heien, D., and C. R. Wessells. 1990. "Demand Systems Estimation with microdata: A Censored Regression Approach." *Journal of Business and Economic Statistics* 54: 165-71.

Hutasuhut, M., Chang, H. S., Griffith, G., C. O'Donnell, and H. Doran. 2001. "The demand for beef in Indonesia: implications for Australian agribusiness." *University of New England, Working Paper Series in Agricultural and Resource Economics*.

Lee, L.-F., and M. M. Pitt. 1986. "Microeconometric demand systems with binding nonnegativity constraints: the dual approach." *Econometrica* 54: 1237-42.

Lee, L.-F., and M. M. Pitt. 1987. "Microeconometric models of rationing, imperfect markets, and non-negativity constraint." *Journal of Econometrics* 36: 89-110.

Lee, L.-F. 1978. "Simultaneous Equations Models With Discrete and Censored Dependent Variables." In *Structural analysis of Discrete Data with Econometric Applications*, eds. P. Manski and D. McFadden, Cambridge, MA: MIT Press, pp. 346-364.

Moeltner K., and J. Englin. 2004. "Choice Behavior Under Time-Variant Quality: State Dependence Versus 'Play-It-by-Ear' in Selecting Ski Resorts." *Journal of Business & Economic Statistics* 22 (2): 214-224.

Nayga Jr., R. M. 1996. "Sample selectivity models for away from home expenditures on wind and beer." *Applied Economics* 28: 1421-5.

Shonkwiler, J. S., and S. T. Yen. 1999. "Two-step estimation of a censored system of equations." *American Journal of Agricultural Economics* 81(4): 972-82.

Su, S.-J. B., and S. T. Yen. 1999. “Two-step estimation of a censored system of equations.” *American Journal of Agricultural Economics* 81(4): 972-82.

Tauchmann, H. 2005. “Efficiency of two-step estimators for censored systems of equations: Shonkwiler and Yen reconsidered.” *Applied Economics* 37: 367-374.

Tauchmann, H. 2008. “Consistency of Heckman-type two-step estimators for the multi-variate sample-selection model.” *Applied Economics* 1-8, iFirst.

Tobin, J. 1958. “Estimation of Relationships for Limited Dependent Variables.” *Econometrica* 26: 24-36.

Vermeulen, F. 2001. “A note on Heckman-type corrections in models for zero expenditures.” *Applied Economics* 33: 1089-1092.

Wales, T. J. and A. D. Woodland. 1983. “Estimation of consumer demand systems with binding non-negativity constraints.” *Journal of Econometrics* 21: 263-85.

Wang, J., X. M. Gao, E. J. Wailes, and G. L. Cramer. 1996. “US consumer demand for alcoholic beverages: cross-section estimation of demographic and economic effects.” *Review of Agricultural Economics* 18: 477-88.

Yen, S. T., K. Kah, and S.-J. Su. 2002. “Household demand for fat and oils: two-step estimation of a censored demand system.” *Applied Economics* 34(14): 1799-806.

Table 1. Summary Statistics for the Whole Sample of Monthly Apple Purchase.

Variable	**Definition**	**Mean**	**Std Dev**	**Minimum**	**Maximum**
pp_rd	Price of red delicious. Equal to observed price for observed purchase; equal to imputed price for non-purchase.	0.899767	0.224812	0	4.436893
pp_ga	Price of Gala	1.044078	0.188097	0	4.407407
pp_gs	Price of Granny Smith	1.151487	0.213181	0	4.435484
pp_gd	Price of Golden Delicious	0.998207	0.159906	0	4.422857
pp_mc	Price of McIntosh	0.947532	0.145615	0	4.419355
pp_ot	Price of Other	1.07982	0.261252	0	4.44
hhsize	Number of household members	2.640895	1.393715	1	9
inc	Household income in thousand dollars	5.485657	3.076592	0.25	22.5
hhage	Age in years	52.01642	12.87186	20	70
Dummy variables (yes=1, no=0)					
kids	There are children under 18	0.311814	0.463235	0	1
hhed_hs	Household head has grade school, some high school or graduated high school education	0.27616	0.447097	0	1
hhed_col	Household head has some college or graduated college education	0.60679	0.488463	0	1
hhed_pc	Household head has post college graduate education	0.117051	0.321481	0	1
white	Household head is white	0.797387	0.401947	0	1
black	Household head is black	0.093534	0.291179	0	1
asian	Household head is asian	0.021434	0.144825	0	1
hispanic	Household head is hispanic	0.071842	0.258226	0	1
other	Household head is not white, black, asian or hispanic	0.015804	0.124717	0	1
urban	Resides in urban area	0.85475	0.352353	0	1
east	Resides in East	0.187859	0.3906	0	1
central	Resides in Central	0.244259	0.429647	0	1
south	Resides in South	0.36384	0.481104	0	1
west	Resides in West	0.204042	0.403	0	1
y1	Purchase in year 1998	0.088972	0.284703	0	1
y2	Purchase in year 1999	0.07859	0.269099	0	1
y3	Purchase in year 2000	0.078625	0.269153	0	1
y4	Purchase in year 2001	0.080622	0.272254	0	1
y5	Purchase in year 2002	0.082619	0.275306	0	1
y6	Purchase in year 2003	0.084444	0.278052	0	1
y7	Purchase in year 2004	0.169507	0.3752	0	1
y8	Purchase in year 2005	0.175498	0.380393	0	1

y9	Purchase in year 2006	0.161123	0.367645	0	1
m1	Purchase in January	0.083333	0.276386	0	1
m2	Purchase in February	0.083333	0.276386	0	1
m3	Purchase in March	0.083333	0.276386	0	1
m4	Purchase in April	0.083333	0.276386	0	1
m5	Purchase in May	0.083333	0.276386	0	1
m6	Purchase in June	0.083333	0.276386	0	1
m7	Purchase in July	0.083333	0.276386	0	1
m8	Purchase in August	0.083333	0.276386	0	1
m9	Purchase in September	0.083333	0.276386	0	1
m10	Purchase in October	0.083333	0.276386	0	1
m11	Purchase in November	0.083333	0.276386	0	1
m12	Purchase in December	0.083333	0.276386	0	1

Table 2. Apple Variety Characteristics.

Variety	Internal Attributes	Color	Function	Availability
Red Delicious	sweet crispy juicy	from striped red to solid midnight red	best eaten fresh or in salads	year-round starting in September
Gala	very sweet crispy juicy	vary in color, from cream to red- and yellow-striped	ideal for snacking	year-round U.S.-grown be harvested beginning in mid-July
Granny Smith	very tart	distinctive green flesh	all-purpose apple work equally well as a snack or in pies and sauce	year-round harvested beginning in August
Golden Delicious	sweet crisp mellow	a pale yellow skin and flesh, sometimes with a red blush	all-purpose great for eating out of hand, baking and salads	year-round appear on the market in September
McIntosh	tangy tart tender juicy	white flesh	best used for snacking and applesauce some people enjoy its tart flavor in pies as well	from September through May

Note: US Apple Association. http://www.usapple.org/consumers/appleguide/guide.cfm.

Table 3. Random Effects Probit Model for Six Apple Varieties.

Variable	Coefficient (Z-value)					
	Red Delicious	Gala	Granny Smith	Golden Delicious	McIntosh	Other
pp_i	-0.0135*	-0.1032***	-0.0575***	-0.0793***	0.0246	-0.0773***
	(-1.84)	(-8.93)	(-5.55)	(-6.47)	(1.25)	(-11.36)
hhsize	0.0206***	0.0144***	0.0142***	0.0144***	0.0194***	-0.0050
	(6.1)	(3.35)	(3.15)	(3.26)	(3.31)	(-1.51)
inc	-0.0127***	0.0252***	0.0264***	0.0038	0.0067	0.0205***
	(-4.16)	(7.15)	(6.69)	(0.92)	(1.26)	(7.3)
inc2	0.0004*	-0.0008***	-0.0007***	-0.0005*	-0.0006	-0.0007***
	(1.94)	(-3.59)	(-2.94)	(-1.74)	(-1.64)	(-4.03)
hhage	0.0017***	0.0050***	-0.0073***	-0.0001	0.0031***	0.0038***
	(4.98)	(11.64)	(-15.51)	(-0.12)	(5.47)	(11.74)
kids	0.0297***	0.0138	0.0849***	0.0382***	0.0034	-0.0046
	(2.89)	(1.05)	(6.23)	(2.83)	(0.19)	(-0.46)
hhed_col	-0.0375***	0.0729***	0.0737***	-0.0520***	-0.0025	0.0639***
	(-4.54)	(7.03)	(6.25)	(-4.77)	(-0.18)	(7.89)
hhed_pc	-0.0636***	0.1226***	0.0857***	-0.0639***	0.0225	0.1195***
	(-4.76)	(7.57)	(4.68)	(-3.67)	(1.03)	(9.48)
black	0.0971***	-0.0836***	-0.0629***	0.1898***	-0.2384***	-0.1310***
	(7.6)	(-5.2)	(-3.44)	(11.74)	(-10.22)	(-10.21)
asian	-0.0547**	0.0013	-0.1932***	0.0165	-0.2675***	0.1754***
	(-2.14)	(0.04)	(-5.36)	(0.49)	(-5.23)	(7.75)
hispanic	0.0981***	0.0143	0.0279	0.0183	-0.0979***	0.0163
	(7.06)	(0.82)	(1.5)	(0.99)	(-3.93)	(1.22)
other	0.0189	-0.0249	-0.0134	0.0313	-0.1088***	-0.0230
	(0.8)	(-0.84)	(-0.43)	(1)	(-2.56)	(-1.01)
urban	-0.0186*	0.0499***	0.0155	0.0191	0.0426**	0.0247**
	(-1.7)	(3.66)	(1.01)	(1.32)	(2.29)	(2.33)
east	-0.1495***	-0.1156***	0.0357**	-0.0404***	0.5455***	0.0808***
	(-12.3)	(-7.48)	(2.06)	(-2.56)	(30.59)	(7.06)
south	-0.0026	0.0868***	0.1144***	0.0307**	-0.2036***	-0.0560***
	(-0.26)	(6.98)	(7.92)	(2.35)	(-11.94)	(-5.71)
west	-0.1853***	0.1384***	0.2279***	-0.1973***	-0.4673***	0.1871***
	(-15.42)	(9.56)	(13.78)	(-12.38)	(-20.91)	(16.66)
y2	-0.1209***	0.2224***	-0.0925***	-0.0531***	0.0346**	-0.0552***
	(-12.58)	(9.09)	(-6.8)	(-4.31)	(2)	(-5.56)
y3	-0.1955***	1.2210***	-0.2011***	-0.1203***	0.0270	-0.2211***
	(-19.14)	(59.07)	(-14.14)	(-9.14)	(1.49)	(-21.09)
y4	-0.2601***	1.2139***	-0.2498***	-0.2033***	-0.0619***	-0.2206***
	(-24.31)	(58.44)	(-16.84)	(-14.68)	(-3.21)	(-20.47)
y5	-0.2764***	1.2235***	-0.2093***	-0.2465***	-0.1617***	-0.2681***
	(-24.78)	(58.09)	(-13.78)	(-16.73)	(-7.77)	(-24.17)

	-0.2948***	1.1573***	-0.2720***	-0.2425***	-0.2025***	-0.3215***
y6	(-25.17)	(53.93)	(-17.06)	(-15.83)	(-9.1)	(-27.83)
	-0.3754***	1.1304***	-0.3813***	-0.4648***	-0.0836***	-0.3509***
y7	(-34.81)	(54.22)	(-24.95)	(-32.14)	(-4.42)	(-32.66)
	-0.3416***	1.0429***	-0.4113***	-0.4314***	-0.1767***	-0.3729***
y8	(-29.86)	(48.9)	(-25.14)	(-28.34)	(-8.79)	(-32.83)
	-0.3871***	1.0686***	-0.4215***	-0.4364***	-0.2718***	-0.3729***
y9	(-31.11)	(48.19)	(-23.96)	(-26.21)	(-12.18)	(-30.6)
	-0.0272***	-0.0295**	-0.0547***	-0.0684***	-0.0712***	0.0103
m2	(-2.93)	(-2.5)	(-4.36)	(-5.53)	(-4.38)	(1.12)
	0.0325***	-0.0080	0.0460***	-0.0278**	-0.0816***	0.0884***
m3	(3.53)	(-0.68)	(3.74)	(-2.26)	(-5)	(9.72)
	-0.0079	-0.0186	-0.0329***	-0.0600***	-0.1615***	0.0227**
m4	(-0.85)	(-1.58)	(-2.62)	(-4.84)	(-9.62)	(2.46)
	-0.0316***	-0.0872***	-0.0355***	-0.0952***	-0.2567***	-0.0765***
m5	(-3.37)	(-7.23)	(-2.82)	(-7.59)	(-14.72)	(-8.16)
	-0.1707***	-0.3719***	-0.1527***	-0.2088***	-0.4701***	-0.3116***
m6	(-17.76)	(-28.23)	(-11.78)	(-16.03)	(-24.66)	(-31.63)
	-0.2330***	-0.4709***	-0.1920***	-0.2633***	-0.6398***	-0.4139***
m7	(-23.81)	(-34.05)	(-14.56)	(-19.75)	(-30.74)	(-40.73)
	-0.1755***	-0.1789***	-0.2040***	-0.2744***	-0.8006***	-0.2947***
m8	(-18.06)	(-14.32)	(-15.35)	(-20.43)	(-35.07)	(-29.67)
	-0.0530***	0.2931***	-0.0842***	-0.0606***	0.0816***	-0.0585***
m9	(-5.56)	(26.04)	(-6.54)	(-4.83)	(5.15)	(-6.17)
	0.0191**	0.2225***	0.0530***	0.0170	0.2374***	0.1753***
m10	(2.02)	(19.55)	(4.25)	(1.37)	(15.55)	(19.06)
	-0.0769***	0.0046	0.0694***	-0.0886***	0.0572***	0.0651***
m11	(-7.99)	(0.39)	(5.58)	(-6.97)	(3.58)	(6.97)
	-0.1043***	-0.1224***	-0.0783***	-0.1644***	-0.1204***	-0.1320***
m12	(-10.75)	(-10)	(-6.08)	(-12.62)	(-7.12)	(-13.69)
	0.0313***	0.0444***	0.0400***	0.0460***	0.0628***	0.0290***
sqtot_i	(72.75)	(61.71)	(55.03)	(60.87)	(51.29)	(73.28)
	-0.0128***	-0.0038***	-0.0046***	-0.0063***	-0.0101***	-0.0063***
sq_np_i	(-51.97)	(-15.88)	(-16.15)	(-21.84)	(-25.62)	(-27.92)
	0.0926***	0.1278***	0.1162***	0.1415***	0.1786***	0.1125***
sq_run_i	(65.43)	(51.68)	(45.26)	(50.57)	(35.34)	(73.26)
	-0.0070***	-0.0070***	-0.0016***	-0.0010**	0.0050***	-0.0097***
sq_npr_i	(-15.88)	(-20.72)	(-4.12)	(-2.52)	(9.76)	(-22.41)
	-0.6952***	-3.0496***	-1.3874***	-1.3541***	-2.2441***	-1.0963***
_cons	(-24.77)	(-76.53)	(-34.59)	(-36.62)	(-46.19)	(-39.47)

Table 4. Random Effects OLS Model for Six Apple Varieties.

Variable	Coefficient (t-value)					
	Red Delicious	Gala	Granny Smith	Golden Delicious	McIntosh	Other
pp_i	-0.8278*** (-87.34)	-0.3993*** (-31.91)	-0.3739*** (-41.88)	-0.4002*** (-34.09)	-0.5188*** (-26.56)	-0.4394*** (-49.22)
inc	-0.1138*** (-18.13)	-0.1106*** (-18.08)	-0.0955*** (-17.45)	-0.1191*** (-19.1)	-0.1266*** (-13.96)	-0.1058*** (-20.78)
inc2	0.0042*** (10.24)	0.0043*** (11.72)	0.0034*** (10.18)	0.0046*** (10.79)	0.0048*** (7.71)	0.0035*** (10.89)
hhage	0.0035*** (4.96)	0.0025*** (3.37)	0.0002 (0.34)	0.0007 (1.17)	0.0007 (0.76)	0.0000 (-0.05)
kids	-1.0179*** (-60.2)	-0.8275*** (-44.87)	-0.6920*** (-46.17)	-0.8089*** (-53.46)	-0.8897*** (-38.07)	-0.8726*** (-59.46)
hhed_col	0.1833*** (10.84)	0.1220*** (6.91)	0.1084*** (6.65)	0.1305*** (8.99)	0.1573*** (7.51)	0.1156*** (8.05)
hhed_pc	0.4388*** (15.84)	0.2975*** (11.05)	0.2453*** (10.03)	0.3196*** (13.94)	0.3195*** (9.91)	0.3442*** (16.14)
black	-0.0027 (-0.1)	0.0329 (1.2)	-0.0135 (-0.55)	-0.1010*** (-5.19)	-0.0401 (-1.01)	-0.0291 (-1.26)
asian	0.0541 (0.99)	-0.0335 (-0.7)	-0.0816 (-1.58)	0.0681 (1.42)	-0.0455 (-0.45)	-0.0249 (-0.69)
hispanic	-0.2523*** (-8.9)	-0.1727*** (-5.94)	-0.1742*** (-7.32)	-0.2166*** (-8.77)	-0.1653*** (-3.95)	-0.1989*** (-8.62)
other	-0.0199 (-0.4)	0.1029** (1.97)	0.0371 (0.86)	0.1502*** (3.13)	-0.1190* ()-1.66	0.0572 (1.33)
urban	0.0298 (1.34)	0.0524** (2.25)	-0.0015 (-0.07)	0.0124 (0.66)	0.0020 (0.07)	-0.0162 (-0.9)
east	-0.1452*** (-5.72)	-0.0432 (-1.63)	-0.0400* (-1.74)	-0.0222 (-1.12)	-0.1766*** (-6.52)	-0.1118*** (-5.86)
south	-0.0324 (-1.6)	-0.0573*** (-2.75)	-0.0153 (-0.79)	0.0080 (0.49)	-0.0277 (-0.99)	-0.0604*** (-3.55)
west	-0.1390*** (-5.52)	-0.0869*** (-3.66)	-0.0428** (-1.99)	-0.1198*** (-5.76)	-0.2503*** (-6.14)	-0.1077*** (-5.82)
y2	-0.3642*** (-19.65)	-0.0878 (-1.28)	0.0996*** (5.3)	-0.0949*** (-4.55)	-0.0279 (-0.81)	0.2280*** (11.71)
y3	-0.2932*** (-14.81)	-0.3755*** (-6.41)	0.1463*** (7.35)	-0.0901*** (-4.16)	0.0442 (1.26)	0.2254*** (10.86)
y4	-0.2925*** (-14.3)	-0.3856*** (-6.55)	0.1914*** (9.42)	-0.0673*** (-3.04)	-0.0090 (-0.25)	0.2352*** (11.52)
y5	-0.2702*** (-13.09)	-0.3897*** (-6.58)	0.1972*** (9.96)	-0.0393* (-1.75)	0.0101 (0.27)	0.2544*** (12.4)
y6	-0.3475*** (-16.61)	-0.3930*** (-6.63)	0.2301*** (11.45)	-0.0246 (-1.11)	0.0104 (0.27)	0.2802*** (13.63)

y7	-0.0844***	-0.2352***	0.3534***	0.2144***	0.2118***	0.5107***
	(-4.34)	(-4.01)	(18.58)	(10.02)	(6.62)	(27.27)
y8	-0.0862***	-0.2112***	0.3932***	0.1535***	0.1994***	0.5789***
	(-4.47)	(-3.59)	(20.48)	(7.36)	(6.16)	(31.02)
y9	-0.0226	-0.2476***	0.3995***	0.1870***	0.2769***	0.5609***
	(-1.12)	(-4.16)	(20.16)	(8.68)	(8.11)	(29.49)
m2	-0.0891***	-0.0762***	-0.0386**	-0.0305	-0.0203	-0.0700***
	(-4.72)	(-3.47)	(-1.99)	(-1.37)	(-0.62)	(-3.57)
m3	-0.0590***	-0.0361*	0.0007	-0.0043	0.0322	-0.0633***
	(-3.17)	(-1.66)	(0.04)	(-0.2)	(0.98)	(-3.29)
m4	-0.0785***	-0.0706***	-0.0181	-0.0065	-0.0246	-0.1068***
	(-4.18)	(-3.24)	(-0.94)	(-0.29)	(-0.72)	(-5.48)
m5	-0.0784***	-0.0985***	0.0426**	-0.0205	-0.0232	-0.0899***
	(-4.15)	(-4.42)	(2.22)	(-0.91)	(-0.66)	(-4.49)
m6	-0.1593***	-0.0874***	-0.0208	-0.0190	-0.0807**	-0.0527**
	(-8.04)	(-3.45)	(-1.03)	(-0.8)	(-2)	(-2.4)
m7	-0.1287***	-0.0836***	-0.0142	-0.0194	-0.0641	-0.0096
	(-6.3)	(-3.07)	(-0.69)	(-0.79)	(-1.39)	(-0.41)
m8	-0.1578***	-0.1520***	-0.0633***	-0.0445*	-0.0589	-0.1247***
	(-7.85)	(-6.43)	(-3.05)	(-1.8)	(-1.1)	(-5.63)
m9	-0.2474***	-0.0933***	-0.0701***	-0.0598***	0.0352	-0.1840***
	(-12.81)	(-4.52)	(-3.55)	(-2.65)	(1.11)	(-9.08)
m10	-0.1081***	-0.0390*	-0.0048	-0.0153	0.1263***	0.0339*
	(-5.72)	(-1.88)	(-0.26)	(-0.7)	(4.19)	(1.78)
m11	-0.1187***	-0.0894***	-0.0016	-0.0654***	0.0730**	-0.0182
	(-6.12)	(-4.15)	(-0.08)	(-2.9)	(2.32)	(-0.94)
m12	-0.0868***	-0.1132***	-0.0220	-0.0735***	0.0329	-0.0614***
	(-4.43)	(-5.06)	(-1.12)	(-3.16)	(0.98)	(-3.01)
invmills_i	-0.4015***	-0.3122***	-0.1873***	-0.3828***	-0.2865***	-0.6595***
	(-22.97)	(-18.58)	(-11.83)	(-25.55)	(-13.32)	(-40.89)
_cons	3.9399***	3.3838***	2.4027***	3.1008***	3.1661***	3.4651***
	(66.17)	(34.07)	(40.3)	(53.71)	(34.01)	(61.47)