REPOSITIONING OF CHOCOLATE MILK AS A CONTENDER FOR MARKET FOR ENERGY DRINKS/SPORTS DRINKS

A Thesis

by

YANG HU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Senarath Dharmasena
Committee Members,	Oral Capps Jr.
	Ramkumar Janakiraman
Head of Department,	C. Parr Rosson III

May 2016

Major Subject: Agricultural Economics

ABSTRACT

Consumption of chocolate milk in the United States is growing as an alternative beverage to sports and energy drinks. Recent literature suggests that consumption of chocolate milk vis-à-vis sports and energy drinks is an effective recovery aid after prolonged workouts. In this light, knowledge of price sensitivity, substitutes/ complements and demographic profiling with respect to consumption of chocolate milk is important for manufacturers, retailers and advertisers of chocolate milk.

Using household-level purchase data for chocolate milk, energy drinks, and sports drinks and related demopraphic characteristics from the 2011 Nielsen Homescan data, we estimated three beverage demand models to show that chocolate milk is a substitute for energy drinks in consumption. Sports drinks are complementary in consumption for energy drinks. Chocoate milk and energy drinks are complements for sports drinks in consumption.

According to own-price elasticity of demand for three beverages, we find that all of them are between -1 and 0 which means that they are price inelastic. The household size, age, education, race, region, the presence of children, gender of household heads are significant determinants of demand for chocolate milk. Household size, age, employment status, education, race, region, the presence of children in a household, gender of household heads significantly affect the demand of energy drinks. Significant demographic variables affecting the demand of sports drinks include household size, age, education, race, region, the presence of children, gender of household heads.

It is important to note that data used in this work only capture at home purchase/consumption of chocolate milk, energy drinks and sports drinks. As a result, household's behavior with respect to away-from home consumption of these beverages is not captured in this thesis.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Dharmasena, and my committee members, Dr. Capps and Dr. Janakiraman, for their guidance and support throughout the course of this research.

Thanks also go to my friends Wen zheng, Xiaoli Liu and the department faculty and staff for making my time at Texas A&M University a great experience. I would also like to thank Alicia Copeland who assisted me in data cleaning for this thesis.

Finally, thanks to my mother and father for their endless love.

TABLE OF CONTENTS

LIST OF TABLES

1. INTRODUCTION AND LITERATURE REVIEW

According to the NPD Group (2010) and Nielsen (2010), consumption of chocolate milk in the United States has been growing. Servings of chocolate milk grew from 1.2 billion in 2009 to 1.4 billion in 2010. The Tetra Pak Dairy Index (2013) reported that, in the U.S., the demand of flavored milk grew by 1% (compound annual growth rate) between 2009 and 2012 and the volume of consumption of flavored milk was growing and expected to increase to nearly ten percent by 2015. However, after the United States Department of Agriculture (USDA) updated official standards of school meal, flavored milk was removed from school cafeterias in some states in response to the concerns about childhood obesity. This policy change might have had some impact on the demand for chocolate milk among children. Another problem faced by the milk market is the rising competition for milk from other beverages, such as protein shakes and plant-based beverages like soymilk and almond milk (Dharmasena and Capps, 2014). As a consequence, chocolate milk industry would face a challenge in terms of moving to a new target market, or position differently.

The market for energy drinks in the United States has become a multibillion dollar business. Also, it is a fast growing segment in the beverage industry, second to bottled water (Agriculture and Agri-Food Canada, 2008). According to the Beverage Marketing Corporation (2015), the consumption of energy drinks grew by 6.4% in volume from 2013 to 2014, sports drinks increased 3% in volume, and sports beverage segment exceeded 1 billion gallons for the first time in 2011. As of 2014, the consumption of sports drinks was reported to be 1.4 billion gallons. Sales of energy drink mixes have grown by 434% between 2011 and 2013. Energy drinks reached 10.5 billion dollars in 2014, which was an increase of 4.9% compared to 2013. The number of gallons of energy drinks consumed grew to be 4 billion in 2014, which was a 5.4% growth compared to previous year (IRI, 2014).

More importantly, many studies have documented over the past couple of years, showing chocolate milk as a substitute for energy or sports drinks. Compared with energy drinks, researchers find that chocolate milk is better in reducing debilitating muscle breakdown and increasing endurance for those who are physically active (Lunn, 2011). When runners drank fat free chocolate milk after a strenuous run, on average, they ran 23% longer and had a 38% increase in markers of muscle building compared to when they drank a carbohydrate-only sports beverage with the same amount of calories (Lunn, 2011). Karp (2006) emphasized that chocolate milk contained high carbohydrate and protein content which were effective for people to recover from strenuous exercise.

In contrast, one of the most pressing issues of energy drinks is the ingredients containing many stimulants, such as the caffeine and guarana. Excessive consumption of energy drinks may increase the risk for caffeine overdose and result in greater potential for acute caffeine toxicity (Reissig, 2009). Initially, the primary consumers of energy drinks were athletes. However, as the energy drinks market expanded into various niche markets, the majority of energy drinks were targeted at teenagers and young adults 18 to 34 year old (Heckman, 2010). According to Kaminer (2010), 30% of youths between ages 12 and 17 regularly consumed energy drinks. However, excessive caffeine is not recommended for people under the age of 18. Although many brands of energy drinks try to dispel consumer's concerns about caffeine, this fact has triggered increased negative media coverage, and consumers' look for healthier alternative beverages.

Due to the ingredient advantage of chocolate milk and weakened outlook of the fliud milk market in the United States, it is a unique opportunity for chocolate milk processors to enter the fastest growing beverage market as an alternative recovery drink. This could potentially provide an additional occasion for consumers to buy chocolate milk and drive sales up (Markets and Markets report, 2016).

In fact, the dairy industry has been repositioning chocolate milk as a contender in the fast-growing market for protein bars, shakes and energy beverages. Since 2012, Milk Processor Education Program (MilkPEP), the group responsible for the "*Got Milk?*" campaign, has invested $15 million a year into chocolate-milk campaign to strengthen the role of chocolate milk as a new-age sports/energy drink (Yang, 2014). Also, MilkPEP set their next 20-year campaign as "*propel milk back into a position of power*," (Berry, 2014). In 2012, MilkPEP launched "*My after*" campaign to strengthen the consciousness that consuming low-fat chocolate milk was better for athletes.

Additionally, chocolate milk, like sports or energy drinks, is aligning with professional athletes and celebrities, incorporating sports games and music to advertise their products. Recently, NBA stars, professional football players, swimmers and running groups have been gradually taking chocolate milk as their recovery drinks. Chocolate milk has become the official refuel beverage of many prominent sports organizations and teams, like IRONMAN® triathlon series, Rock' n Roll Marathon series, and Challenged Athletes Foundation (Built with Chocolate Milk, 2014).

Below we discuss few past studies that are available in the extant literature dealing with estimating demand for choclate milk, and energy drinks. Dharmasena and Capps (2009) used a Heckman sample selection procedure to estimate the demand for chocolate milk in the United States, for calendar year 2008 using a Nielsen Homescan panel data. They found that the own-price elasticity of demand for chocolate milk was estimated to be -0.04. Factors affecting the probability of purchase of chocolate milk were price of chocolate milk, household income, age of household head, and education status of household head. Maynard (1999) estimated the own price elasticity of demand for flavored milk which was in a range from -1.4 to -1.47. They used weekly scanner data for the period 1996 through 1998. Capps and Hanselman (2012) employed the Barten synthetic demand system to estimate own price, cross-price, and expenditure elasticities for major energy drink brands by using weekly survey data from October 2007 to October 2010.According their study, the own-price elasticity of demand for energy drinks was estimated to be from -0.99 to -1.69.

While the media has linked chocolate milk benefits with the emphasis on healthy ingredients and performance, according to our knowledge, systematic scientific analysis documenting consumer demand for chocolate milk and sports/ energy drinks in the United States is not available.

2. OBJECTIVES

A thorough and complete analysis of demand for chocolate milk, energy drinks, and sports drinks are important to uncover demand interrelationships. Additionally, the price sensitivity, substitutes or complements and demographic profiling with respect to consumption of chocolate milk, energy drinks, and sports drinks are important for manufacturers, retailers and advertisers of these beverage products. Specific objectives of this study are to: (1) estimate the own-price elasticities, cross-price elasticities and income elasticity of chocolate milk and energy and sports drinks, (2) determine the socio-economic-demographic factors affecting the purchase of chocolate milk, energy drinks and sports drinks in the United States.

3. THEORETICAL MODEL

Out of total number of households included in the sample, some households did not buy chocolate milk, energy and/or sports drinks during the sampling period. In this case, the dollar amount households' spent on these beverages was recorded as zero. If the fraction of the observations of the dependent variables take this limit value (lower limit being zero), the dependent variable is said to be censored. This kind of consuming behavior would lead to corner solutions for some nontrivial fraction of the sampled households. Application of ordinary least squares (OLS) to estimate this kind of regression gives rise to biased estimates even asymptotically (Kennedy, 2003). As a result, Tobit model is suggested as a method to explicitly model the corner solution dependent variables (or censored dependent variables). Tobit model is applied to outcome variables that are roughly continuous over positive values but have a positive probability of equaling zero (Tobin, 1958 and Heckman, 1979).

The Tobit model is defined as a latent variable model as follows:

$$(1) \qquad Y_i = \beta \boldsymbol{X}_i + \mu_i, \quad \beta \boldsymbol{X}_i + \mu_i > 0 \qquad \mu_i \sim Normal(0, \sigma^2)$$

$$Y_i = 0 \qquad \beta \boldsymbol{X}_i + \mu_i \leq 0$$

where $i = 1,2,3,\ldots\ldots n$ is the number of observations, Y_i is the censored dependent variable, $\boldsymbol{X}_i$ is a vector of explanatory variables, β is the vector of unknown parameters to be estimated. μ_i is the normally distributed error. For Tobit model, there are two expectations of Y dependent variable, conditional expectation, $E(Y|Y > 0, \boldsymbol{X})$ and unconditional expectation $E(Y)$. Conditional expected value of Y is expressed in equation (2), while unconditional expected value is shown in equation (3).

(2) Conditional expectation: $E(Y|Y > 0, \boldsymbol{X}) = \boldsymbol{X}\beta + \sigma(\frac{f(z)}{F(z)})$

(3) Unconditional expectation: $E(Y) = E(Y|Y > 0) * P(Y > 0|\boldsymbol{X})$

$$= E(Y|Y > 0) * F(z)$$

$$= \boldsymbol{X}\beta F(z) + \sigma(f(z))$$

Where $z = \frac{\boldsymbol{X}\beta}{\sigma}$, is the standardized value and $\lambda = \frac{f(z)}{F(z)}$ which is called inverse mills ratio, is the ratio between the standard normal probability density function, pdf $(f(z))$ and standard normal cumulative density function, cdf $(F(z))$, each evaluated at z. In Tobit model, the coefficients represent the effect of an independent variable on the latent dependent variable. Therefore, the coefficients with each explanatory variable must be transformed into meaningful marginal effects. There are two types of marginal effects. The first one is *conditional marginal effects* which reflect the marginal effects on consumption that contains the households actually bought the beverage. The other is the *unconditional marginal effects* for consumption of beverage which include all the households whether or not buy the beverage.

If X_i is a continuous variable, the conditional marginal effect of X_i on $E(Y|Y > 0, \boldsymbol{X})$is represented by:

(4) $$\frac{\partial E(Y|Y>0)}{\partial \boldsymbol{X}} = \beta(1 - z\frac{f(z)}{F(z)} - \frac{f(z)^2}{F(z)^2})$$

The unconditional marginal effect of X_i on $E(Y)$ is shown by:

(5) $$\frac{\partial E(Y)}{\partial \boldsymbol{X}} = \beta F(z)$$

From equation 3, we know that $E(Y) = E(Y|Y > 0) * F(z)$, therefore:

$$\frac{\partial E(Y)}{\partial \boldsymbol{X}} = F(z)\frac{\partial E(Y|Y>0)}{\partial \boldsymbol{X}} + E(Y|Y > 0)\frac{\partial F(z)}{\partial \boldsymbol{X}} \tag{6}$$

The total change in the unconditional expected value of dependent variable Y is represented by the sum of (i), the change in the expected value of Y being above the limit weighted by the probability of being above the limit $[F(z)\frac{\partial E(Y|Y>0)}{\partial \boldsymbol{X}}]$and (ii) the change in the probability of being above the limit weighted by the expected value of Y being above the limit $[E(Y|Y > 0)\frac{\partial F(z)}{\partial \boldsymbol{X}}]$ (McDonald & Moffitt's, 1980).

The elasticity of Y with respect to x_1, conditional on $Y > 0$, is

$$\frac{\partial E(Y|Y>0)}{\partial x_1} \times \frac{x_1}{E(Y|Y>0)} \tag{7}$$

This can be computed when x_1 appears in various functional forms, including level, logarithmic, and quadratic forms.

The significance level considered in this thesis is set at p-value 0.05 (or 95% significance level).

4. EMPIRICAL MODEL

For the fraction of households with censored data (censored quantity or expenditure), price of the product was not observed. Therefore we used an auxiliary regression to forecast the price for these households where price was not observed. To accomplish this, observed price for each beverage was regressed on household income, household size, and the region.

(8) $P_{i,observed} = \alpha_1 + \alpha_2 \times HH_{i,income} + \alpha_3 \times HH_{i,size} + \alpha_4 \times HH_{i,region} + \mu_i$

where $i=1,2,3,\ldots.n$, number of households.

These variables are extensively used in the literature in imputing missing prices (Kyureghian, *et al*., 2011, Alviola and Capps, 2010, and Dharmasena and Capps, 2014).The household income relates to the different levels of product quality as it is reflected by the price of the product. The household region reflects the spatial differences in price. Household size not only reflects the composition of the households but also relates to the amount dollars that households spend on the product,assuming that large households tend to buy less expensive products. The parameters estimated from the auxiliary regression were then used to impute prices for these households with zero-expenditure observations. The forecast price through the auxiliary regression would act as an instrument (or proxy) for the observed price, which is commonly used to address the issue of endogeneity of prices (Capps *et al*., 1994, Alviola and Capps, 2010, Kyureghian, Capps, and Nayga, 2011, and Dharmasena and Capps, 2012). Summary statistics of observed and imputed price for each beverage are shown in Table 1. The mean prices and standard deviations were consistent for both observed prices as well as imputed prices. Specifically, the mean of the observed price of chocolate milk is $0.049 /ounce which is consistent with the mean of imputed price which is $0.051 /ounce. The mean of observed price for energy drinks is $0.129 /ounce which is again consistent with the imputed price for energy drinks, $0.131 /ounce. The observed mean price for sports drinks is $0.052/ounce and imputed price is $0.053/ounce,which is consistent with the mean of the oberseved price.

Table 1 Summary Statistics of Observed Prices and Imputed Prices for Each Beverage

	Observed Price (U.S. dollars per ounce)		Imputed Price (U.S. dollars per ounce)	
	Mean	Standard Deviation	Mean	Standard Deviation
Chocolate milk	0.049	0.02	0.051	0.01
Energy drinks	0.129	0.06	0.131	0.01
Sports drinks	0.052	0.15	0.053	0.003

Source: calculated by authors

Pearson correlation coefficient among these three beverage prices are shown in Table 2. The correlation between chocolate milk price and energy drinks price is significant and it is equal to 0.13 which means there is a very weak positive linear relationship between chocolate milk price and energy drinks price. The correlation coefficient between chocolate milk price and sports drinks price, the price of energy drinks and price of sports drinks are not statistically significant under p-value of 0.05.

Therefore, we fail to reject the null hypothesis that there exists a perfect linear correlation, which means that there is no significant linear relationship between chocolate milk price and sports drinks price, as well as between energy drinks price and sports drinks price.

Table 2 Correlation Test for Beverage Prices

	Chocolate Milk Price	Energy Drinks Price	Sports Drinks Price
Chocolate Milk Price	1	**0.13** (<.0001)	0.007 (0.0510)
Energy Drinks Price		1	-0.00033 (0.9347)
Sports Drinks Price			1

Source: Computed by author
Notes: Pearson correlation coefficients, N = 62029, Prob > |r| under H0: Rho=0

Once the imputed price for each type of beverage is obtained, the independent variables to estimate the Tobit model included imputed prices for those observations with previously zero prices, observed prices, household income, presence of children in the household, region, race, employment status, level of education, and gender of household head.

We test several hypotheses regarding purchases of chocolate milk, energy drinks, and sports drinks. They are as follows: [1] Chocolate milk is a substitute in consumption for energy drinks and sports drinks and therefore have positive cross-price elasticities; energy drinks are substitutes in consumption for sports drinks so that the cross price elasticity is positive. [2] Households with young household heads consume more chocolate milk, energy drinks and sports drinks; [3] Members of full-time-empoyed households consume a greater share of chocolate milk, energy drinks and sports drinks away from home; [4] Households with more educated household heads consume less chocolate milk,energy drinks and sports drinks; [5] households with white household heads consume more chocolate milk than other racial groups and households with Hispanic household heads consume less chocolate milk; Households with Oriental household heads consume less energy drinks and sports drinks than other races. [6] Households with children consume more chocolate milk; households with children who are teenagers consume more energy drinks and sports drinks.

Several functional forms were investigated in this thesis such as linear, quadratic, and semi-log. We found that semi-log functional form outperformed other functional forms in chocolate milk demand model and sports drinks demand model, based on model fit, significance of the variables, and the results of loss metrics like Akaike Information Criterion (AIC). However, for the energy drinks Tobit model, price of chocolate milk in linear form outperformed price represented in natural logarithm. Therefore, we used the semi-log functional form to calculate the conditional and unconditional marginal effects associated with each explanatory variable, except for linear functional form for price of chocolate milk in energy drink Tobit model. Therefore, the demand functions for three beverages are as follows:

(9) Chocolate milk demand model:

$$Y_{quantity} = \delta_1 + \delta_2 \log price_{chocolatemilk} + \delta_3 \log price_{energydrinks} + \delta_4 \log price_{sportsdrinks} + \delta_5 \log HHincome + \boldsymbol{\delta_6 X}_{demographic\ variables} + \mu$$

(10) Energy drinks demand model:

$$Y_{quantity} = \rho_1 + \rho_2 price_{chocolatemilk} + \rho_3 \log price_{energydrinks} + \rho_4 \log price_{sportsdrinks} + \rho_5 \log HHincome + \boldsymbol{\rho_6 X}_{demographic\ variables} + \mu$$

(11) Sports drinks demand model:

$$Y_{quantity} = \varphi_1 + \varphi_2 price_{chocolatemilk} + \varphi_3 \log price_{energydrinks} + \varphi_4 \log price_{sportsdrinks} + \varphi_5 \log HHincome + \boldsymbol{\varphi_6 X}_{demographic\ variables} + \mu$$

Where $\boldsymbol{X}_{demographic\ variables}$= $X_{age\ of\ household\ head\ 25-29}$, $X_{age\ of\ household\ head\ 30-34}$, $X_{age\ of\ household\ head\ 35-44}$, $X_{age\ of\ household\ head\ 45-54}$, $X_{age\ of\ household\ head\ 55-64}$, $X_{age\ of\ household\ head\ 65\ or\ older}$, $X_{employment\ status\ part-time}$, $X_{employment\ status\ full-time}$, $X_{education\ high\ school}$, $X_{education\ undergraduate}$, $X_{education\ post-college}$, X_{Black}, $X_{Oriental}$, X_{other}, $X_{Hispanic}$, $X_{New\ England}$, $X_{Middle\ Atlantic}$, $X_{East\ North\ central}$, $X_{West\ North\ central}$, $X_{South\ Atlantic}$, $X_{East\ South\ central}$, $X_{West\ South\ central}$, $X_{Moutain}$, $X_{children\ less\ than\ 6\ years}$, $X_{children\ 6-12\ years}$, $X_{children\ 13-17\ years}$, $X_{children\ under\ 6\ and\ 6-12\ years}$,

$X_{children\ 6-12\ and\ 13-17\ years}$, $X_{children\ under\ 6\ 6-12\ 13-17}$, $X_{female\ head\ only}$, $X_{male\ head\ only}$.

Conditional marginal effect for semi-log price variable:

$$(12) \qquad \frac{\partial E(Y|Y>0)}{\partial p} = \frac{\beta}{p^c}\left(1 - z\frac{f(z)}{F(z)} - \frac{f(z)^2}{F(z)^2}\right)$$

Unconditional marginal effect for semi-log price variable:

$$(13) \qquad \frac{\partial E(Y)}{\partial p} = \frac{\beta}{p^u} F(z)$$

Where p^c is the average price in the censored sample, p^u is the average of the unconditional price. Conditional and unconditional own-price, cross-price and income elasticity are represented as follows.

Conditional elasticities:

$$(14) \quad \text{Own-Price:} \quad \varepsilon_{ii}^C = \frac{\beta}{p_i^C}\left(1 - z\frac{f(z)}{F(z)} - \frac{f(z)^2}{F(z)^2}\right)\frac{p_i^C}{Q_i^C}$$

$$(15) \quad \text{Cross-Price:} \quad \varepsilon_{ij}^C = \frac{\beta}{p_j^C}\left(1 - z\frac{f(z)}{F(z)} - \frac{f(z)^2}{F(z)^2}\right)\frac{p_j^C}{Q_i^C}$$

$$(16) \quad \text{Income:} \quad \varepsilon_I^C = \frac{\beta}{I_i^C}\left(1 - z\frac{f(z)}{F(z)} - \frac{f(z)^2}{F(z)^2}\right)\frac{I_i^C}{Q_i^C}$$

Where ε_{ii}^C is the conditional own-price elasticity for *i*; ε_{ij}^C is the conditional cross-price elasticity measures the quantity demanded for *i* to a change in the price of *j*; ε_I^C is the conditional income elasticity for *i*th beverage.

For the linear price (chocolate milk price in energy drinks Tobit model), conditional cross-price elasticity is

$$(17) \quad \varepsilon_{ij}^C = \beta\left(1 - z\frac{f(z)}{F(z)} - \frac{f(z)^2}{F(z)^2}\right)\frac{p_j^C}{Q_i^C}$$

Unconditional demand elasticities:

(18) Own-Price: $$\varepsilon_{ii}^{u} = \frac{\beta}{p_i^u} F(z) \frac{p_i^u}{Q_i^u}$$

(19) Cross-Price: $$\varepsilon_{ij}^{u} = \frac{\beta}{p_j^u} F(z) \frac{p_j^u}{Q_i^u}$$

(20) Income: $$\varepsilon_{I}^{u} = \frac{\beta}{I_i^u} F(z) \frac{I_i^u}{Q_i^u}$$

Where ε_{ii}^{u} is the unconditional own-price elasticity for *i*; ε_{ij}^{u} is the unconditional cross-price elasticity measures the quantity demanded for *i* to a change in the price of *j*; ε_{I}^{u} is the unconditional income elasticity for *i*th beverage.

For the linear price (chocolate milk price in energy drinks Tobit model), unconditional cross-price elasticity is

(21) $$\varepsilon_{ij}^{u} = \beta \mathrm{F}(z) \frac{p_j^u}{Q_i^u}$$

Where I^c is conditional mean income and I^u is unconditional mean income, Q_i^C is the conditional mean quantity, Q_i^u is the unconditional mean of quantity. From equation (6), we could obtain the changes in the probability of being above the limit for consumption of each beverage category in response to a change in an explanatory variable.

(22) $$\frac{\partial F(z)}{\partial X} = \frac{1}{E(Y|Y>0)} \left(\frac{\partial E(Y)}{\partial X} - F(z) \frac{\partial E(Y|Y>0)}{\partial X} \right)$$

5. DATA

The data used in this study is based on 2011 Nielsen Homescan panel, which provides detailed beverage-purchase information from 62,029 households from across the United States. The Nielsen Homescan data are a nationwide panel of households who scan their food purchases for at-home use from all retail outlets (grocery stores, department stores, convenience stores, drug stores, and club stores. Data includes detailed product characteristics, quantities, and expenditures for each food item purchased by each household and socioeconomic demographic characteristic (for a detailed account of Nielsen Homescan panel data, see Dharmasena, 2010).

The energy drinks include brands such as Red Bull, Monster, Tampico, Talking Rain, RockStar, etc. The sports drinks include three brands: Gatorade, Powerade, and All Sport R.

Table 3 shows the summary statistics for all variables included in the model. We standardized the quantity as liquid ounces for all beverages and the expenditures are expressed in dollars. A *unit value,* which is taken as a proxy for price is generated by dividing total expenditure by quantity for each beverage. This unit value variable is considered as price paid for each beverage category and is expressed as dollars per ounce. The mean price for chocolate milk, energy drinks and sports drinks are $0.049 /ounce, $0.129 /ounce, $0.052/ounce.

Table 3 Summary Statistics of the Variables Used in the Model

Variable	Mean	Standard Deviation
Price of chocolate milk	0.049	0.024
Price of energy drinks	0.129	0.056
Price of sports drinks	0.052	0.149
Household size	2.360	1.290
Household income	58.32	31.93
Age of household head 25-29	0.018	0.042
Age of household head 30-34	0.038	0.191
Age of household head 35-44	0.147	0.354
Age of household head 45-54	0.276	0.447
Age of household head 55-64	0.297	0.457
Age of household head 65 or older	0.222	0.415
Employment status part-time	0.178	0.383
Employment status full-time	0.390	0.488
Education high school	0.237	0.425
Education undergraduate	0.618	0.485
Education post-college	0.120	0.325
Black	0.094	0.292
Oriental	0.029	0.166
Other	0.040	0.196
Hispanic	0.051	0.220
New England	0.045	0.208
Middle Atlantic	0.131	0.337
East North central	0.181	0.385
West North central	0.086	0.281
South Atlantic	0.198	0.398
East South central	0.060	0.237
West South central	0.102	0.303
Mountain	0.073	0.260
Children less than 6 hears	0.028	0.164
Children 6-12 years	0.052	0.223
Children 13-17 years	0.067	0.249
Children under 6 and 6-12 years	0.024	0.154
Children under 6 and 13-17 years	0.004	0.064
Children 6-12 and 13-17 years	0.033	0.179
Children under 6, 6-12, and 13-17	0.005	0.070
Female head only	0.250	0.433
Male head only	0.096	0.295

Source: Nielsen Homescan data 2011, calculated by authors.

Household size is separated into 9 groups, according to Nielsen Homescan 2011 demographic specifications. The code value of household size we assigned to is equal to the number of household members. If household members are more than nine, the value of household size is nine. The mean value for household size is 2.36 (Table3) which means most households have 2 to 3 members.

There are 16 household income categories (the table that appears on page 48) and the mean value of household income is 58.32(Table 3) which means the income of most households fall into $50,000-$59,999.

The base case for the age of household head is considered as the household head that is less than 25 years. From Table 3, we know that households with 25-29 years and the households with 30-34 years are small proportions of the sample; only 1.8% of the households fall into the category of 25-29 years and 3.8% of the households are 30-34 years. The household heads who are 35-44 years constitute 14.7% of the sample. 27.6% of the household heads are 45 to 54 years old. Household heads who are 55-64 years make up 29.7% of the sample. Households over 65 years account for more than 20 percent of the sample.

Employment status is an indicator variable representing whether the household head is full time employed, part-time employed or employed neither full time nor part time. We treat the household heads with neither full-time nor part-time as the base case in this thesis. 17.8% of the household heads have part-time jobs and 39% household heads work full-time. Also we consider the education status of households. The base case of household heads' education status considered is less than high school educated.

23.7% of the household heads have high school degree. While there are 12% of household heads that earned post-college education, more than 60% of the household heads had undergraduate degrees. Race is grouped as White, Black, Oriental(Asian) and other. Race white is used as the base case for this analysis. 9.4% of the sample is Black. Oriental household heads account for 2.9% of the sample. Four percent of the household heads belong to other race catogery. 5.1% of the household heads are Hispanic. Regions are labeled as New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific (detailed classification information is shown in Table 4).

Table 4 Census Bureau Regions and States

New England	**Middle Atlantic**	**East North Central**
Connecticut, Maine, Massachusetts, New Rhode Island, Vermont Hampshire	New Jersey, New York, Pennsylvania	Indiana, Illinois, Ohio, Wisconsin， Michigan
West North Central	**South Atlantic**	**East South Central**
Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota	Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia	Alabama, Kentucky, Tennessee， Mississippi
West South Central	**Mountain**	**Pacific**
Arkansas, Louisiana, Oklahoma, Texas	Arizona, Colorado, Idaho, New Mexico, Montana, Utah, Nevada, Wyoming	Alaska, California, Oregon, Washington， Hawaii

Source: U.S. Department of Commerce Economics and Statistics Administration U.S. Census Bureau.

Pacific region is treated as the base case for this analysis. 4.5% of the household heads are from New England, 13.1% from Middle Atlantic, 18.1% from East North Central, 8.6% from West North central, 19.8% from South Atlantic, 6% from East South, 10.2% from West South, and 7.3% from Mountain.

Variable with respect to the presence of the children in the households is classified into 8 categories based on the children's age. The base case considered in this study is the households with no children. The other 7 categories are households with children less than 6 years, children 6 to 12 years, children 13-17 years, children under 6 and in the range of 6-12 years, children under 6 and 13 to 17 years, children 6-12 and 13-17 years, and the households with children under 6, 6 to 12 and 13-17 years. Households with children who are from 13 to 17 years account for 6.7% of the total sample. The base case of household's gender is defined as a household with both female and male. If the household is headed by both female and male, then we considered the female's demographic characteristics. Households headed by female only made up 25% of the sample. Male only household heads composed 9.6% of the dataset.

6. EMPIRICAL RESULTS AND DISCUSSION

Table 5 shows the summary statistics for price, quantity, expenditure and market penetration (number of households purchased the beverage under consideration, out of total number of households sampled in this study) for three beverage categories studied in this thesis.

Table 5 Summary Statistics for Price, Quantity and Market Penetration in 2011 in U.S.

	Chocolate Milk	Energy Drinks	Sports Drinks
Market penetration	26.09%	7.23%	35.78%
Unconditional average price($)	0.049	0.13	0.053
Conditional average price($)	0.05	0.13	0.052
Average conditional quantity(oz)	423	441.12	756.55
Average unonditional quantity(oz)	110.38	31.87	270.73

Source: calculated by authors.

It is shown that 26.1% households purchased chocolate milk, 7.23% households purchased energy drinks, 35.7% households chose sports drinks. Conditional average price (*unit value*) for chocolate milk, energy drinks, sports drinks are $0.049/ounce, $0.13/ounce, and $0.052/ounce, respectively. Unconditional average price for chocolate milk, energy drinks, and sports drinks are respectively $0.05/ounce, $0.13/ounce, and $0.053/ounce. The average conditional quantity consumed per household per year for chocolate milk is -423 ounces, which is higher than the unconditional quantity, around 110 ounces, because average conditional quantity only takes into account the households who actually consumed the chocolate milk. Average conditional quantity of energy drinks consumed per household per year is 441 ounces. The unconditional average quantity of energy drinks consumed is only 31.87 ounces. For sports drinks, the average conditional quantity is around 757 ounces per year and the unconditional quantity is less than 300 ounces.

Table 6 shows the Chi-square-test results to test the joit effects of categorical demographic variables and Table 7 presents the Tobit regression results.

The significant (at p-value 0.05 level) economic determinants for chocolate milk are price of chocolate milk, price of energy drink, price of sports drinks. The household income did not have a significant effect on the demand of chocolate milk. Significant demographic variables affecting the demand for chocolate milk includes household size, age, education, race, Hispanic origin, region, the presence of children in a household and gender of the household head.

Statistically significant determinants of demand for energy drinks are the price of energy drinks, price of chocolate milk, and price of sports drinks; household size, age, employment status, education, race, Hispanic origin, region, the presence of children in a household, and household heads' gender. Household income did not have a significant effect on the demand for energy drinks.

Statistically significant factors affecting the demand of sports drinks are price of chocolate milk, energy drinks, and sports drinks. Significant demographic determinants were household income household size, household income, age, education, race, region, the presence of children in a household, and gender of the household head.

Table 6 Chi-square test for joint significance of demographic variables considered in chocolate milk, energy drinks and sports drinks Tobit models

Chocolate milk		Energy drinks		Sports drinks	
P-value associated with Chi-sq	Lable	P-value associated with Chi-sq	Lable	P-value associated with Chi-sq	Lable
<.0001	agehh2529 = 0 agehh3034 = 0 agehh3544 = 0 agehh4554 = 0 agehh5564 = 0 agehhgt64 = 0	<.0001	agehh2529 = 0 agehh3034 = 0 agehh3544 = 0 agehh4554 = 0 agehh5564 = 0 agehhgt64 = 0	<.0001	agehh2529 = 0 agehh3034 = 0 agehh3544 = 0 agehh4554 = 0 agehh5564 = 0 agehhgt64 = 0
0.8042	emphhpt = 0 emphhft = 0	0.0015	emphhpt = 0 emphhft = 0	0.0428	emphhpt = 0 emphhft = 0
<.0001	eduhhhs = 0 eduhhu = 0 eduhhpc = 0	<.0001	eduhhhs = 0 eduhhu = 0 eduhhpc = 0	<.0001	eduhhhs = 0 eduhhu = 0 eduhhpc = 0
<.0001	black = 0 , oriental = 0 other = 0	0.0003	black = 0 oriental = 0 other = 0	0.0009	black = 0 oriental = 0 other - 0 = 0
<.0001	newengland = 0 middleatlantic = 0 eastnorthcentral = 0 westnorthcentral = 0 southatlantic = 0 eastsouthcentral = 0 westsouthcentral = 0 mountain = 0	<.0001	newengland = 0 middleatlantic = 0 eastnorthcentral = 0 westnorthcentral = 0 southatlantic = 0 eastsouthcentral = 0 westsouthcentral = 0 mountain = 0	<.0001	newengland = 0 middleatlantic = 0 eastnorthcentral = 0 westnorthcentral = 0 southatlantic = 0 eastsouthcentral = 0 westsouthcentral = 0 mountain = 0
<.0001	aclt6_only = 0 ac6_12only = 0 ac13_17only = 0 aclt6_6_12only = 0 aclt6_13_17only = 0 ac6_12and13_17only = 0 aclt6_6_12and13_17 = 0	<.0001	aclt6_only = 0 ac6_12only = 0 ac13_17only = 0 aclt6_6_12only = 0 aclt6_13_17only = 0 ac6_12and13_17only = 0 aclt6_6_12and13_17 = 0	<.0001	aclt6_only = 0 ac6_12only = 0 ac13_17only = 0 aclt6_6_12only = 0 aclt6_13_17only = 0 ac6_12and13_17only = 0 aclt6_6_12and13_17 = 0
<.0001	fhonly = 0 mhonly = 0	<.0001	fhonly = 0 mhonly = 0	<.0001	fhonly = 0 mhonly = 0

Source: calculated by authors.

The results of median conditional marginal effects are shown in Table 8. In order to reduce the influences by outliers and skewed data, we used the median values in deriving marginal effects as well as elasticities in this thesis. Table 9 presents the results of median changes in probability of consumption for the change in each explanatory variable.

Table 7 Tobit Regression Results for Chocolate Milk, Energy Drinks, and Sports Drinks

	Chocolate milk			**Energy drinks**			**Sports drinks**		
Variable	Estimate	Std Error	p-Value	Estimate	Std Error	p-Value	Estimate	Std Error	p-Value
Intercept	**-4813.94**	203.73	<.0001	**-5204.96**	288.98	<.0001	**−7004.03**	251.97	<.0001
Price of chocolate milk	**-1008.42**	20.75	<.0001	**2460.87**	54.5	<.0001	**−78.12**	32.92	0.0118
Price of energy drink	**-146.67**	49.50	0.003	**-1506.67**	902.13	0.0064	**-319.24**	61.04	<.0001
Price of sports drink	**-161.25**	16.11	<.0001	**-179.82**	30.47	<.0001	**−1639.65**	17.36	<.0001
Household size	**75.45**	8.07	<.0001	**145.63**	14.79	<.0001	**163.54**	10.14	<.0001
Household income	-18.49	10.99	0.0926	8.99	21.14	0.6707	**38.04**	14.11	0.007
Age of household head 25-29	-37.25	149.88	0.8037	-136.99	221.9	0.537	−83.88	177.83	0.6371
Age of household head 30-34	43.27	146.06	0.7670	-205.93	216.38	0.3412	−121.92	173.66	0.4827
Age of household head 35-44	78.05	143.86	0.5875	**-493.84**	212.59	0.0202	−163.61	170.89	0.3384
Age of household head 45-54	100.96	143.49	0.4818	**-614.89**	211.88	0.0037	−238.37	170.45	0.1620
Age of household head 55-64	16.20	143.47	0.9101	**-962.17**	212.32	<.0001	**−509.21**	170.49	0.0028
Age of household head 65 or older	-187.13	143.85	0.1933	**-1369.22**	214.45	<.0001	**−720.42**	171.07	<.0001
Employment status part-time	-11.39	17.65	0.5185	-48.4	35.59	0.1739	−15.14	22.68	0.5045
Employment status full-time	-2.43	15.68	0.8770	**73.05**	30.72	0.0174	37.61	20.04	0.0605
Education high school	-27.46	40.73	0.5002	**-182.9**	77.11	0.0177	68.05	53.38	0.2024
Education undergraduate	**-111.11**	40.11	0.0056	**-328.26**	75.58	<.0001	−6.1	52.47	0.9075
Education post-college	**-237.18**	44.36	<.0001	**-576.51**	85.35	<.0001	**−145.52**	57.49	0.0114
Black	**-336.19**	23.84	<.0001	-29.92	42.83	0.4847	−17.81	27.49	0.5169
Oriental	**-243.6**	40.3	<.0001	**-261**	73.46	0.0004	**−161.99**	47.66	0.0007
Other	**-77.73**	34.58	0.0246	**126.41**	58.73	0.0314	**82.72**	42.07	0.0493
Hispanic	**-74.54**	30.84	0.0156	**105.14**	52.44	0.0449	19.67	37.19	0.5968

Table 7 Continued

Variable	Chocolate milk			Energy drinks			Sports drinks		
	Estimate	Std Error	p-Value	Estimate	Std Error	p-Value	Estimate	Std Error	p-Value
New England	-40.97	36.89	0.2667	**-451.73**	71.06	<.0001	76.86	45.03	0.0878
Middle Atlantic	**169.63**	26.78	<.0001	**-374.22**	49.02	<.0001	36.2	33.27	0.2765
East North central	**116.35**	24.81	<.0001	**-401.72**	45.99	<.0001	5.09	32.59	0.8758
West North central	**161.00**	28.19	<.0001	**-349.84**	54.27	<.0001	-24.36	37.49	0.5158
South Atlantic	**63.86**	24.48	0.0091	**-337.75**	42.8	<.0001	**165.20**	29.72	<.0001
East South central	**201.58**	31.26	<.0001	**-298.1**	59.74	<.0001	**296.76**	39.71	<.0001
West South central	**120.92**	27.67	<.0001	**-128.1**	47.62	0.0071	**252.84**	33.86	<.0001
Mountain	**-65.70**	30.59	0.0317	-78.48	52.27	0.1333	**116**	37.41	0.0019
Children less than 6 years	65.63	40.23	0.1028	**-198.69**	73.5	0.0069	**-164.76**	50.72	0.0012
Children 6-12 years	**155.27**	29.99	<.0001	-104.01	56.54	0.0659	**95.67**	37.48	0.0107
Children13-17 years	**173.56**	26.66	<.0001	**265.72**	46.87	<.0001	**480.68**	32.9	<.0001
Children under6 and 6-12 years	**103.17**	44.47	0.0203	**-460.37**	85.45	<.0001	**-320.62**	56.39	<.0001
Children under 6 and 13-17 years	150.50	88.59	0.0893	-173.20	159.94	0.2788	21.36	110.97	0.8474
Children 6-12 and 13-17 years	**112.65**	38.87	0.0038	**-152.66**	71.41	0.0325	**293.44**	47.92	<.0001
Children under 6, 6-12, 13-17	103.38	84.23	0.2196	**-321.23**	152.28	0.0349	-13.48	106.02	0.8988
Female head only	**-69.18**	18.47	0.0002	49.42	36.7	0.1780	**-233.51**	23.89	<.0001
Male head only	**-101.56**	25.34	<.0001	**291.32**	46.48	<.0001	-54.66	31.68	0.0844
Sigma	**1141.31**	6.79	<.0001	**1531.57**	17.98	<.0001	**1541.74**	7.70	<.0001

Source: calculated by authors, the significance of the estimated coefficients is based on a p-value of 0.05.

Table 8 Median Conditional Marginal Effect

Variable	Chocolate milk	Energy drinks	Sports drinks
Household size	19.04	24.64	56.48
Age of household head 25-29	-9.4	-23.17	-28.95
Age of household head 30-34	10.9	-34.83	-42.07
Age of household head 35-44	19.69	-83.53	-56.46
Age of household head 45-54	25.47	-104.01	-82.26
Age of household head 55-64	4.08	-162.76	-175.72
Age of household head 65 or older	-47.22	-231.61	-248.61
Employment status part-time	-2.87	-8.19	-5.22
Employment status full-time	-0.61	12.36	12.98
Education high school	-6.93	-30.94	23.48
Education undergraduate	-28.04	-55.53	-2.1
Education post-college	-59.85	-97.52	-50.22
Black	-84.43	-5.06	-6.15
Oriental	-61.47	-44.15	-55.9
Other	-19.61	21.38	28.55
Hispanic	-18.81	17.78	6.79
New England	-10.34	-76.41	26.52
Middle Atlantic	42.81	-63.30	12.49
East North central	29.36	-67.95	1.76
West North central	40.63	-59.18	-8.41
South Atlantic	16.16	-57.13	57.01
East South central	50.87	-50.42	102.41
West South central	0.23	-21.67	87.25
Mountain	-16.58	-13.27	40.03
Children less than 6 years	16.56	-33.61	-56.85
Children 6-12 years	39.18	-17.59	33.01
Children 13-17 years	43.79	44.95	165.88
Children under6 and 6-12 years	26.04	-77.88	-110.64
Children under6 and 13-17 hears	37.98	-29.30	7.37
Children 6-12 and 13-17 years	28.43	-25.82	101.26
Children under6,6-12,and 13-17	26.08	-54.34	-4.65
Female head only	-17.46	8.36	-80.58
Male head only	-25.63	49.28	-18.86

Source: calculated by authors.

Table 9 Median Change in Probability of Consumption for being above the Limit for Change in each Demographic Variable for Chocolate Milk, Energy Drinks, and Sports Drinks

Variable	Chocolate milk	Energy drinks	Sports drinks
Household size	0.022	0.017	0.04
Age of household head 25-29	-0.01	-0.016	-0.021
Age of household head 30-34	0.013	-0.024	-0.03
Age of household head 35-44	0.023	-0.057	-0.04
Age of household head 45-54	0.029	-0.071	-0.059
Age of household head 55-64	0.004	-0.11	-0.127
Age of household head 65 or older	-0.054	-0.157	-0.179
Employment status part-time	-0.003	-0.006	-0.004
Employment status full-time	-0.001	0.008	0.009
Education high school	-0.008	-0.021	0.017
Education undergraduate	-0.032	-0.038	-0.002
Education post-college	-0.069	-0.066	-0.037
Black	-0.098	-0.003	-0.004
Oriental	-0.071	-0.030	-0.041
Other	-0.023	0.015	0.020
Hispanic	-0.022	0.012	0.005
New England	-0.012	-0.052	0.019
Middle Atlantic	0.049	-0.043	0.009
East North central	0.034	-0.046	0.002
West North central	0.047	-0.040	-0.005
South Atlantic	0.019	-0.039	0.042
East South central	0.059	-0.034	0.075
West South central	0.047	-0.015	0.063
Mountain	-0.019	-0.009	0.030
Children less than 6 years	0.019	-0.023	-0.041
Children 6-12 years	0.045	-0.012	0.025
Children 13-17 years	0.050	0.030	0.121
Children under6 and 6-12 years	0.03	-0.053	-0.078
Children under6 and 13-17 hears	0.044	-0.020	0.006
Children 6-12 and 13-17 years	0.033	-0.017	0.074
Children under6,6-12,and 13-17	0.030	-0.037	-0.005
Female head only	-0.020	0.006	-0.058
Male head only	-0.029	0.033	-0.014

Source: calculated by authors.

For chocolate milk, the average change in probability of consumption for change in household size at the median is 0.022, which means an increase in one household family member would increase the chocolate milk consumption by 2.2%. Median conditional marginal effect shown in Table 8 shows that adding one more household member would increase the chocolate milk consumption by about 19 ounces per year. A household head who had post college education is 7% less likely to consume chocolate milk compared with the base case of household head with less than high school education. Household head with the post college education would consume approximately 60 ounces less chocolate milk per year compared to the base category.

Compared with the base case, white household heads, other race including black, oriental, and other are 2.3~9.8% less likely to consume chocolate milk. Black household heads consume 84 ounces chocolate milk less than the household heads classified as White. Household head classified as Oriental consume 61.47 ounces less than the base case, while the household heads who belong to other race category consume 19.6 ounces less than white household heads. Hispanic household heads consume 18.8 ounces less chocolate milk with 2.2% less chance of purchase than non-Hispanic household head. Region has been classified into nine parts which Pacific is treated as the base case. Except New England, others significantly affect consumption of chocolate milk. Households living in Middle Atlantic, East North central, West North central, South Atlantic, East South central, and West South central are 1.9% - 5.9% more likely to buy chocolate milk than the households living in Pacific and consume 0.23-50.87 ounces more chocolate milk. However, the household living in Mountain is 1.9% less likely to consume chocolate milk with 16.58 ounces less chocolate milk consumption than the base case.

The presence of children in a household increases the probability of chocolate milk consumption relative to household with no children. Households who have 6-12 years children consume 39 ounces more chocolate milk with 4.5% more likely than the base case. It is 5% more likely to consume chocolate milk for the households whose children are 13-17 years old and these households consume 44 ounces more chocolate milk. Households whose children are between 6-12 years old and less than 6 years purchase 26 ounces more chocolate milk, which amounts to 3 percent smaller probability than the households without children. Households who have children at the age from 6 to 12 and 13 to 17 increase the probability of chocolate milk consumption by 3.3% and one additional child who is in these age range causes 28.4 ounces more consumption of chocolate milk compared with the households who have no children. Compared with the household heads by both male and female, only male or female household heads purchase about 27 ounces or 26 ounces less chocolate milk per year with 2.9% or 2% less likely to consume chocolate milk, respectively.

For energy drinks, an increase in the household size results in 25 ounces more purchase of energy drinks. The change in probability of consumption for change in household size is 1.7%. Household heads that are more than 35 years old are less likely to purchase energy drinks than the base case that household heads are less than 25 years old. With increasing age comes increased quantity of reduced purchase of energy drinks. Specifically, household heads who are from 35 to 44 years old are 5.7% less likely to purchase energy drinks which are 84 ounces less energy drinks than the base case. The median change in the probability for household heads from 45 to 54 years old is -0.07. Therefore, a household heads whose ages are in this range is 7% less likely to consume energy drinks. The household heads in 45-54 years old consume 104 ounces less energy drinks than the household headed by a person younger than 25. When household heads are 55-64 years old, the consumption of energy drinks is reduced by 163 ounces on the basis of the household heads who are less than 25 years old with 11% probability. Households headed by more than 65 years old are 15.7% less likely to purchase energy drinks and they consume 232 ounces less energy drinks than the base case.

The household heads that have full time jobs prefer to consume 12 ounces more energy drinks with 0.8% greater probability than the household headed by a person with no job. As long as the education level of a household head is higher than high school (household heads who take less than high school education are treated as the base case), they will reduce their consumption of energy drinks. Further, the household head who has an undergraduate degree consume 56 ounces less energy drinks than the base case with 3.8% less probability. Households with high school education are associated with a 31 ounces decline relative to the base case of less than a high school education. The marginal change in the probability of change in households with high school education for energy drinks is -0.021. Households with post-college education are 6.6% less than the households who have less than high school education and they would consume around 98 ounces less energy drinks.

Considering the race of household heads, oriental household heads are 3% less likely to purchase energy drinks and consume 44 ounces less than the base case, White people. Hispanic household heads are 12% more likely to purchase energy drinks and consume 18 ounces more than Whites.

Regionally, except the Mountain region, household heads living in other regions consume less energy drinks and are less likely to purchase energy drinks than the base case – the Pacific region. Households living in the Mountain region are not statistically significant with respect to purchase of energy drinks. The households who are living in New England are 5.2% less likely to consume energy drinks compared with Pacific region with around 76 ounces less consumption. The probability of purchasing energy drinks for the household heads living in East North central decreases by 4.6 percentage points and the quantity that they prefer to purchase decreases by around 68 ounces per year. The household heads living in Middle Atlantic consume 63.3 ounces less energy drinks per year with 4.3% less likelihood to consume energy drinks compared with Pacific region. Households in West North central are 4% less likely to purchase energy drinks and the consumption decreased by around 59 ounces per year. Households living in South Atlantic consume 57 ounces less energy drinks per year with 3.9 less likelihood to purchase energy drinks. Households living in East South central have a 3.4% less probability of consuming energy drinks and they consume around 51 ounces less than the base case of Pacific region. Household heads living in West South central consume around 22 ounces less energy drinks per year with 1.5% less probability to purchase energy drinks.

The presence of children in a household whose ages are between 13-17 years old increased by 44.95 ounces energy drinks with 3% more likely to purchase energy drinks. By contrast, household heads that have children less than 6 years purchase 33 ounces less energy drinks than households with no children with 2.3% less likelihood. Households with children under 6 and 6-12 years consume about 78 ounces less energy drinks with 5.3% less likelihood. If the households have 6-12 year-old and 12-17 year-old children, these households are 1.7% less likely to purchase energy drinks and consume around 26 ounces less than the base case. Households with children under 6, 6-12, and 13-17 years, are 3.7% less likely to consume energy drinks with 54 ounces less consumption.

For the gender of household heads, households headed by male consume 49 ounces more energy drinks with 3.3% more likely than the base case of household heads are both male and female.

From the perspective of the demand for sports drinks, household size also has significant effect on sports drinks demand. Presense of one extra member in the household wold increase the consumption of sports drinks by 56 ounces with 4% more probability to purchase sports drinks. Households head of 55-65 years old are 12.7% less likely to consume sports drinks and they consume 176 ounces less sports drinks than the base case (which is the households with household heads less than 25 years). Similarly, household heads who are more than 65 years old consume 249 ounces less sports drinks than people younger than 25 years. Household heads in this age range are 17.9% less likely to consume sports drinks than the base case.

Higher levels of education again reduce the amount of sports drinks consumed. Household heads received post college degree consumed 50 ounces less sports drinks per year and the likelihood of purchase by 3.7 percent less relative to the base case of person owned less than high school education.

Oriental household heads consume 56 ounces less sports drinks and are 4 percent less likely to purchase sports drinks compared with White households. Other races are 2% more likely to purchase sports drinks with 28.5 ounces more consumption than those who are classified as White. Households in the South Atlantic, East South central, West South central, and Mountain purchase 57, 102, 87, 40 ounces more sports drinks per year, respectively, than households living in the Pacific States with 3%-7.5% greater probability of consumption.

Overall, the presence of children whose age is less than 6 years old reduce the consumption of sports drinks. When households have children less than 6 years of age, it is 4.1% less likely for these households to purchase sports drinks compared with the base case of households without children and they consume 57 ounces less sports drinks. Households with children who are under 6 and 6-12 years old are 7.8% less likely to purchase sports drinks and they consume 110 ounces less than the base case. However, when households have children whose ages are between 6-12 or 13 -17 years, these households consume 33 or 166 ounces more sports drinks per year than the households without children. There is 12% more likely for households with 6 -12 years old children to purchase sports drinks than the base case and 2.5% more probability for households with 13-17 years old children. Households headed by female consume 80 ounces less sports drinks per year than the base case of households headed by a female and a male. It is 5.8% less likely for female household heads to purchase sports drinks than the base case.

Table 10 reports the median unconditional marginal effects. The difference between conditional and unconditional marginal effects is that, unconditional marginal effects are smaller than conditional marginal effects, in absolute value.

From Table 10, we know that additional member in a given household will increase the purchase of chocolate milk by 14.92 ounces. Household heads with undergraduate degree consume around 22 ounces less chocolate milk than the households who have less than high school degree. Households headed by the people who have post-college degrees decrease their consumptions of chocolate milk by 47 ounces per year. Household heads who are black consume around 67ounces less chocolate milk, oriental household heads consume 48 ounces less, and household heads who are other race consume 16 ounces less chocolate milk per year than the white household heads. Hispanic households consume 15 ounces less chocolate milk than the households who are no Hispanic.

Regionally, compared with the households living in the Pacific region, the households living in Middle Atlantic consume 34 ounces more, in East North central consume 23 ounces more, in West North central consume 32 ounces more, and in South Atlantic consume 13 ounces more, in East South central consume around 40 ounces more, living in West South central consume 23.9 ounces more, while household in Mountain consume 13 ounces less chocolate milk.

Table 10 Median Unconditional Marginal Effects

Variable	Chocolate milk	Energy drinks	Sports drinks
Household size	14.92	7.71	40.5
Age of household head 25-29	-7.37	-7.25	-20.77
Age of household head 30-34	8.56	-10.9	-30.19
Age of household head 35-44	15.44	-26.14	-40.52
Age of household head 45-54	19.97	-32.55	-59.03
Age of household head 55-64	3.20	-50.94	-126.1
Age of household head 65 or older	-37.03	-72.49	-178.41
Employment status part-time	-2.25	-2.56	-3.75
Employment status full-time	-0.48	3.87	9.31
Education high school	-5.43	-9.68	16.85
Education undergraduate	-21.98	-17.38	-1.51
Education post-college	-46.93	-30.52	-36.04
Black	-66.53	-1.58	-4.41
Oriental	-48.21	-13.82	-40.12
Other	-15.38	6.69	20.49
Hispanic	-14.75	5.57	4.87
New England	-8.10	-23.91	19.03
Middle Atlantic	33.57	-19.81	8.97
East North central	23.02	-21.27	1.26
West North central	31.86	-18.52	-6.03
South Atlantic	12.63	-17.88	40.91
East South central	39.89	-15.78	73.49
West South central	23.93	-6.78	62.62
Mountain	-13	-4.15	28.73
Children less than 6 years	12.99	-10.52	-40.80
Children 6-12 years	30.72	-5.51	23.69
Children 13-17 years	34.35	14.07	119.04
Children under6 and 6-12 years	20.41	-24.37	-79.4
Children under6 and 13-17 hears	29.78	-9.17	5.29
Children 6-12 and 13-17 years	22.29	-8.08	72.67
Children under6,6-12,and 13-17	20.46	-17.01	-3.34
Female head only	-13.69	2.61	-57.83
Male head only	-20.10	15.422	-14.47

Source: calculated by authors.

Households with the children 6-12 years consume around 31 ounces more chocolate milk per year compared with the households with no children. Households have children under 6 years and 6-12 years increase the consumption of chocolate milk by 20.41 ounces per year. If the households have children who are between 6-12 years and between 13-17 years, they would purchase 22.3 ounces more chocolate milk per year. The households headed by only female or only male consume around 14 ounces or 20 ounces less chocolate milk than the households headed by both female and male.

The unconditional marginal effect of household size for energy drinks demand is 7.71 which mean increasing one person in a household will lead to 7.71 ounces increased in consumption of energy drinks. Compared with the household heads who are less than 25 years old, the household heads who are 35-44 years would decrease their consumption of energy drinks by around 26.14 ounces; household heads with 45-54 years consume 33 ounces less; household heads who are 55-64 years consume relatively 51 ounces less; households whose head is 65 years or more older consume around 73 ounces less. Household heads who are in full time jobs consume only around 4 ounces more energy drinks than the household heads who are not in full-time and part-time jobs. If the household heads have more than high school degrees, the consumption of energy drinks decrease by 10-31 ounces per year. Oriental households purchase around 14 ounces less energy drinks than the white household heads and the households who are other races consume 6.7 ounces more energy drinks. Hispanic households consume around 5.6 ounces more energy drinks than non-Hispanic households. Households living in New England consume 24 ounces less energy drinks compared with the people living in the Pacific region. Household heads living in the Middle Atlantic consume around 20 ounces less energy drinks and households living in the East North central consume 21 ounces less energy drinks per year than the households who are living in the Pacific region. Household heads living in the West North central, South Atlantic, and East South central consume 18.5-15.78 ounces less energy drinks than the base case.

Households with children who are less than 6 years consume 10.5 ounces less energy drinks, while households whose children are 13-17 years consume 14 more ounces energy drinks compared with the household with no children. If the households have children who are 13-17 years, these households consume around 14 ounces more energy drinks per year. Households with children under 6, 6-12 years, and 13-17 years consume 17 ounces less energy drinks and households whose children are 6-12 years and 13-17 years purchase 8 ounces less energy drinks compared with the base case. The households headed by only male consume 15 ounces more than the households headed by both male and female.

For sports drinks, one member increase in one household will purchase 40.5 ounces more sports drinks. Households who are more than 55 years consume around 150 ounces less sports drinks than the households less than 25 years. Household heads who have post-college education consume about 36 ounces less sports drinks. Oriental households consume around 40 ounces less sports drinks while other race consume around 21 ounces more sports drinks than the base case.

Household heads living in South Atlantic, East South central, West South central, and Moutain consume more than 29 ounces more sports drinks than the Pacific region.

Households with children less than 6 years consume 40 ounces less sports drinks. Households with chidren between 6 and 12 years consume 24 ounces more sports drinks. Households with children are 13-17 years consume 119 ounces more sports drinks. Households with children who are under 6 and 6-12 years consume 79 ounces less sports drinks. Household heads whoes children are 6-12 years and 13-17 years consume around 73 ounces more sports drinks. Female household heads consume 58 ounces less sports drinks.

Based on the coefficient estimates, we calculated the conditional and unconditional own-price, cross-price and income elasticities for all beverages. Table 11 represents the mean value of conditional and unconditional elasticities. The unconditional elasticities estimates are consistently smaller than the conditional elasticities. Because unconditional values take into account the household who have yet to purchase the beverage that probably are not loyal to these beverages, should the price of this beverage increase a little, these consumers sway away from this beverage.

For chocolate milk, the conditional own-price elasticity of demand is -0.62, which means that consumers are relatively insensitive to own price changes. The conditional cross-price elasticities of demand of chocolate milk with energy drinks and sports drinks are -0.09, -0.1, which implies energy drinks and sports drinks are complementary beverages for chocolate milk in consumption. The conditional income elasticity of demand for chocolate milk is -0.01, however this was not statistically significant at p-value 0.05.

Table 11 Unconditional and Conditional Own-price, Cross-price, and Income Elasticities of Demand for Chocolate Milk, Energy Drinks, and Sports Drinks

	Chocolate milk	Energy drinks	Sports drinks
Unconditional			
Chocolate milk	**-2.049**	**-0.298**	**-0.328**
Energy drinks	**0.253**	**-3.079**	**-0.368**
Sports drinks	**-0.093**	**-0.364**	**-1.778**
Income	-0.038	0.018	0.03
Conditional			
Chocolate milk	**-0.624**	**-0.0911**	**-0.099**
Energy drinks	**0.047**	**-0.599**	**-0.071**
Sports drinks	**-0.038**	**-0.146**	**-0.748**
Income	-0.011	0.004	0.017

Notes: The numbers printed in bold are statistically significant at a p-value of 0.05.

For energy drinks, the conditional own-price elasticity of demand is -0.6, indicating that energy drinks are price inelastic in demand. The cross-price elasticity of demand of energy drinks with chocolate milk and sports drinks are 0.05, and -0.07. Therefore, chocolate milk is a substitute for energy drinks in consumption, but sports drinks are complementary in consumption for energy drinks. The income elasticity is 0.004 which is statistically not significant.

For sports drinks, the conditional own-price elasticity of demand for sports drinks is -0.75, indicating that sports drinks are less elastic than energy drinks. The cross-price elasticity of demand for sports drinks with chocolate milk and energy drinks are -0.036, and -0.146. Therefore, chocolate milk and energy drinks are complementary in consumption for sports drinks. The income elasticity of demand is 0.017 which is statistically not significant.

7. CONCLUSIONS

Using household-level purchase data for chocolate milk, energy drinks, and sports drinks and related demographic characteristics from the 2011 Nielsen Homescan data, we estimated three beverage demand models to show that chocolate milk is a substitute for energy drinks in consumption. Sports drinks are complementary in consumption for energy drinks. Chocolate milk and energy drinks are complements for sports drinks in consumption.

According to the own-price elasticity of demand for three beverages, we find that all of them are between -1 and 0 which means that the percentage change in quantity is smaller than that of change in price. Therefore, if the price is raised, the total revenue for retailer increases and raised price will not result in the loss of potential buyers. In addition, we find that demographic characteristics of households have an impact on the demand for chocolate milk, energy drinks, and sports drinks. The household size, age, education, race, region, the presence of children, gender of household head are significant determinants of demand for chocolate milk. Household size, age, employment status, education, race, region, the presence of children in a household, gender of household head significantly affect the demand of energy drinks. Significant demographic variables affecting the demand of sports drinks include household size, age, education, race, region, the presence of children, gender of household head.

It is important to note that data used in this work only capture at home purchase/ consumption of chocolate milk, energy drinks and sports drinks. As a result, household's behavior with respect to away-from home consumption of theses beverages is not captured in this thesis.

REFERENCES

Alviola, 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.

Berry D. Mar 2014. "Feed Them Milk: A New and Needed Approach to Nourishing Americans." Retrieved from: http://berryondairy.blogspot.com/2014/03/feed-them-milk-new-and-needed-approach.html (accessed on January 28, 2016)

Beverage Marketing Corporation, 2015.(n.d.). Retrieved from: http://www.beveragemarketing.com/ (accessed on January 28, 2016)

Built with Chocolate Milk, 2014. (n.d.). Retrieved from: http://builtwithchocolatemilk.com/ (accessed on January 28, 2016)

Capps, O. Jr., Hanselman, R. D. 2012. "A Pilot Study of the Market for Energy Drinks." *Journal of Food Distribution Research* 43(3): 15-29.

Capps, O.Jr., R. Kirby, and G.Williams. 1994. "A Comparison of Demand for Meat Products in the Pacific Rim Region." *Journal of Agricultural and Applied Economics* 19(1):210-224

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.

_____.2014. "U.S. Demand for Wellness and Functional Beverages and Implications on Nutritional Intake: An Application of EASI Demand System Capturing Diverse

Preference Heterogeniety." Paper presented at AAEA Annual Meeting, Minneapolis, MN, 27-29 July.

_____.2012. "Intended and Unintended Consequences of a Proposed National Tax on Sugar-Sweetened Beverages to Combat the U.S. Obesity Problem." Health Economics 21(6)：669-694[DOI: 10.1002/HEC.1738].

_____.2011. "Is Chocolate Milk the New-Age Energy\Sports Drink in the United States?" Paper presented at SAEA Annual Meeting, Corpus Christi, TX, 5-8 February.

_____.2010. "The Non-alcoholic Beverage Market in the United States: Demand Interrelationships, Dynamics, Nutrition Issues and Probability Forecast Evaluaion" Unpublished Phd dissertation, Texas A&M University.

Dharmasena, S., O. Capps, Jr., Kosub B. 2015. "Demand and Market Competitiveness of Almond Milk as a Dairy Alternative Beverage in the United States." Paper presented at SAEA Annual Meeting, Atlanta, GA, 31 January - 3 February.

Godfrey, V. Sept 2012. "2011-2012 Annual School Channel Survey ." Retrieved from: http://www.foodbev.com/news/vivien-godfrey-and-the-challenges-facing/#.VmH_JLerSM8 (accessed on January 28, 2016)

Heckman M.A.,Sherry K., and Gonzalez E. de M. May 2010. "Energy Drinks: An Assessment of Their Market Size, Consumer Demographics, Ingredient Profile, Functionality, and Regulations in the United States." *Comprehensive Reviews in Food Science and Food Safety* 9(3): 303-317.

Karp J. R., Johnston J. D., Tecklenburg S., Mickleborough T. D.,et al. Feb 2006.

"Chocolate Milk as a Post-Exercise Recovery Aid." *International Journal of Sport Nutrition* & Exercise *Metabolism* 16(1):78-91.

Kaminer Y. Jul 2010. "Problematic Use of Energy Drinks by Adolescents." *Child and Adolescent Psychiatric Clinics of North America* 19(3):643-50[DOI: 10.1016/j.chc.2010.03.015].

Kyureghian, G., O.Capps, Jr., and R.Nayga. 2011. "A Missing Variable Imputation Methodology with an Empirical Application." *Advances in Econometrics* 27A:313-337.

Linshi, J. Feb12, 2015. "These 2 Charts Show the Biggest Change in America's Breakfast." Retrieved from: http://time.com/3705987/skipping-breakfast-cereal-kellogg/ (accessed on January 28, 2016)

Lunn, WR., Pasiakos, SM., Colletto, MR., Karfonta, KE., et al. 2011. "Chocolate Milk & Endurance Exercise Recovery: Protein Balance, Glycogen and Performance." *Medicine & Science in Sports & Exercise.*

Markets and Markets. (n.d.) "Global Sports and Energy Drinks Market is Expected to Reach $52 Billion by 2016." Retrieved from: http://www.marketsandmarkets.com/PressReleases/energy-drinks.asp (accessed on January 28, 2016)

Maynard L.J., Liu D.Y. 1999. "Fragility in Dairy Product Demand Analysis." Paper presented at AAEA Annual Meeting, Nashville, TN, 8-11 August.

McDonald, J.F., and R.A.Moffitt. 1980. "The Uses of Tobit Analysis." *Review of*

Economics and Statistics 62[2]:318-321.

Reissig C.J., Strain E.C., Griffiths R. R. Jan 2009. “Caffeinated Energy Drinks—A Growing Problem.” *Drug and Alcohol Dependence* 99: 1-10.

Tetra Pak Dairy Index, An Annual News and Information Source about the Dairy Industry, Issue 6-June 2013, Flavored Milk. Retrieved from: http://www.april.org.ro/images/fisiere/75a7942c855e8257d233082694ecb233.pdf (accessed on January 28, 2016)

Tobin, J. 1958. “Estimation of Relationships for Limited Dependent Variables.” *Econometrica* 26[1]:24-36.

Wooldridge, J. M. 2012. *Introductory Econometrics: A Modern Approach*, 5th. ed. South-Western College Pub.

Yang, J. L. 2014. “Beverage of Champions: Chocolate Milk Gets an Olympic-Style Makeover.” Retrieved from: https://www.washingtonpost.com/business/economy/beverage-of-champions-chocolate-milk-gets-an-olympic-style-makeover/2014/01/31/a13261b6-89c8-11e3-916e-e01534b1e132_story.html (accessed on January 28, 2016)

APPENDIX A

1. Derivation of Equation (22):

$$\frac{\partial E(Y)}{\partial \boldsymbol{X}} = F(z)\frac{\partial E(Y|Y>0)}{\partial \boldsymbol{X}} + E(Y|Y>0)\frac{\partial F(z)}{\partial \boldsymbol{X}} \text{ (equation 6)}$$

$$\frac{\partial E(Y)}{\partial \boldsymbol{X}} - F(z)\frac{\partial E(Y|Y>0)}{\partial \boldsymbol{X}} = E(Y|Y>0)\frac{\partial F(z)}{\partial \boldsymbol{X}}$$

$$\frac{\partial F(z)}{\partial \boldsymbol{X}} = [\frac{\partial E(Y)}{\partial \boldsymbol{X}} - F(z)\frac{\partial E(Y|Y>0)}{\partial \boldsymbol{X}}]\frac{1}{E(Y|Y>0)}$$

APPENDIX B

Table 12 Demographic Specifications and the Code Value

Household Income	Code Value
Under $5000	5
$5000-$7999	6.5
$8000-$9999	9
$10,000-$11,999	11
$12,000-$14,999	13.5
$15,000-$19,999	17.5
$20,000-$24,999	22.5
$25,000-$29,999	27.5
$30,000-$34,999	32.5
$35,000-$39,999	37.5
$40,000-$44,999	42.5
$45,000-$49,999	47.5
$50,000-$59,999	55
$60,000-$69,999	65
$70,000-$99,999	85
$100,000 & Over	112.5
Household Size	**Code Value**
Single Member	1
Two Members	2
Three Members	3
Four Members	4
Five Members	5
Six Members	6
Seven Members	7
Eight Members	8
Nine+ Members	9

Source: Nielsen Homescan 2011 demographic specifications

APPENDIX C

SAS CODES

1. SAS Codes for processing raw data from 2011 Nielsen Homescan panel:

1.1 Chocolate Milk:

```
libname Yang 'Z:\Yang\Yang';
PROC IMPORT out=Yang.Products
        DATAFILE= "z:\Ali\Original TSV files\products.tsv"
        dbms=dlm replace;
                delimiter='09'x;
   GETNAMES=YES;
RUN;

data yang.test; set yang.products;
if department_code=0 then delete;
if department_code=1 then delete;
if department_code=2 then delete;
if department_code=4 then delete;
if department_code=5 then delete;
if department_code=6 then delete;
if department_code=7 then delete;
if department_code=8 then delete;
if department_code=9 then delete;
run;

data yang.test1; set yang.test;
if product_module_code ne 3592 then delete;
if index(upc_descr, " ST ")>0 then delete;
if index(upc_descr, " VAN ")>0 then delete;
run;

PROC SORT DATA=Yang.Test1 OUT=Yang.Test1 ;
 BY upc upc_ver_uc ;
RUN ;

PROC IMPORT out=Yang.Purchases2011
        DATAFILE= "z:\Ali\Original TSV files\purchases_2011.tsv"
        dbms=dlm replace;
                delimiter='09'x;
   GETNAMES=YES;
RUN;

PROC SORT DATA=yang.Purchases2011 OUT=yang.Purchases2011 ;
```

```
  BY upc upc_ver_uc ;
RUN ;

data Yang.TestPurchases2011;
        merge Yang.Test1 Yang.Purchases2011;
        by upc upc_ver_uc;
run;

data yang.TestPurchases2011; set yang.TestPurchases2011;
  if product_module_code eq '.' then delete;
run;

data Yang.TestPurchases2011; set yang.TestPurchases2011;
  if trip_code_uc eq '.' then delete;
run;

data Yang.TestPurchases2011; set Yang.TestPurchases2011;
  Q = quantity * multi * size1_amount;
run;

data Yang.TestPurchases2011; set Yang.TestPurchases2011;
  P = total_price_paid / Q;
run;

PROC IMPORT out=Yang.Trips2011
        DATAFILE= "z:\Ali\Original TSV files\trips_2011.tsv"
        dbms=dlm replace;
                        delimiter='09'x;
   GETNAMES=YES;
RUN;

PROC SORT DATA=Yang.Trips2011 OUT=Yang.Trips2011 ;
  BY  trip_code_uc ;
RUN ;

PROC SORT DATA=Yang.TestPurchases2011 OUT=Yang.TestPurchases2011 ;
  BY  trip_code_uc ;
RUN ;

data Yang.TestPurchasesTrips2011;
        merge Yang.TestPurchases2011 Yang.Trips2011;
        by trip_code_uc;
run;

data Yang.TestPurchasesTrips2011; set Yang.TestPurchasesTrips2011;
  if upc eq '.' then delete;
run;
```

```
PROC SORT DATA=Yang.TestPurchasesTrips2011 OUT=Yang.TestPurchasesTrips2011;
  BY  household_code ;
RUN ;

DATA yang.ChocolateMilk2011; SET yang.TestPurchasesTrips2011;
        BY household_code;
        RETAIN total_CH_exp total_CH_oz total_CH_coupon;
         IF first.household_code THEN
                DO;
                total_CH_exp = 0;
                total_CH_oz = 0;
                Total_CH_coupon = 0;
        END;
        total_CH_exp = total_CH_exp + total_price_paid;
        total_CH_oz = total_CH_oz + Q;
        total_CH_coupon = total_CH_coupon + coupon_value;
        If last.household_code then output;
        KEEP household_code total_CH_exp total_CH_oz total_CH_coupon;
        run;

        PROC IMPORT out=YAng.Panelists2011
       DATAFILE= "z:\Ali\Original TSV files\panelists_2011.tsv"
       dbms=dlm replace;
                        delimiter='09'x;
   GETNAMES=YES;
RUN;

PROC SORT DATA=yang.panelists2011 OUT=yang.panelists2011 ;
  BY  household_code ;
RUN ;

data Yang.CH;
        merge YAng.ChocolateMilk2011 Yang.Panelists2011;
        by household_code;
run;
```

1.2 Energy Drinks:

```
libname Yang 'Z:\Yang\Yang';

PROC IMPORT out=Yang.Products
       DATAFILE= "Z:\Ali\Original TSV files\products.tsv"
       dbms=dlm replace;
                        delimiter='09'x;
   GETNAMES=YES;
RUN;

data yang.test; set yang.products;
```

```
if department_code=0 then delete;
if department_code=2 then delete;
if department_code=3 then delete;
if department_code=4 then delete;
if department_code=5 then delete;
if department_code=6 then delete;
if department_code=7 then delete;
if department_code=8 then delete;
if department_code=9 then delete;
run;

data yang.test1; set yang.test;
if product_module_code=622 then delete;
if product_module_code=819 then delete;
if product_module_code=358 then delete;
if product_module_code=186 then delete;
if product_module_code=820 then delete;
if product_module_code=813 then delete;
if product_module_code=224 then delete;
if product_module_code=204 then delete;
if product_module_code=230 then delete;
if product_module_code=222 then delete;
if product_module_code=142 then delete;
if product_module_code=140 then delete;
if product_module_code=132 then delete;
if product_module_code=200 then delete;
if product_module_code=170 then delete;
if product_module_code=182 then delete;
if product_module_code=174 then delete;
if product_module_code=186 then delete;
if product_module_code=1487 then delete; *bottled water;
if product_module_code=1052 then delete; *ice pops - unfrozen;
if product_group_code=1501 then delete;
if product_group_code=1505 then delete;
if product_group_code=1506 then delete;
if product_group_code=1507 then delete;
if product_group_code=504 then delete;
if product_group_code=1001 then delete;
if product_group_code=1020 then delete;
if product_group_code=1010 then delete;
if product_group_code=511 then delete;
if product_group_code=506 then delete;
if product_group_code=1004 then delete;
if product_group_code=1005 then delete;
if product_group_code=1017 then delete;
if product_group_code=1012 then delete;
if product_group_code=1013 then delete;
if product_group_code=1007 then delete;
```

```
if product_group_code=508 then delete;
if product_group_code=1002 then delete;
if product_group_code=512 then delete;
if product_group_code=1015 then delete;
if product_group_code=503 then delete;
if product_group_code=1014 then delete;
if product_group_code=1016 then delete;
if product_group_code=510 then delete;
if product_group_code=513 then delete;
if product_group_code=1019 then delete;
if product_group_code=514 then delete;
if product_group_code=1009 then delete;
if product_group_code=1021 then delete;
if product_group_code=1008 then delete;
if product_group_code=1006 then delete; *coffee;
if product_group_code=505 then delete;
if product_group_code=1011 then delete;
if product_group_code=501 then delete;
if product_group_code=507 then delete;
if product_group_code=1018 then delete;
if product_group_code=. then delete;
if product_module_code=1046 then delete;
if product_module_code=1482 then delete;
if index(upc_descr, " SOY ")>0 then delete;
if index(brand_descr, "PEPSI WILD CH")>0 then delete;
if index(brand_descr, "PEPSI VANILLA")>0 then delete;
if index(upc_descr, " E-J ")=0 then delete;
run;
data yang.test2; set yang.test;
if product_module_code=622 then delete;
if product_module_code=819 then delete;
if product_module_code=358 then delete;
if product_module_code=186 then delete;
if product_module_code=820 then delete;
if product_module_code=813 then delete;
if product_module_code=224 then delete;
if product_module_code=204 then delete;
if product_module_code=230 then delete;
if product_module_code=222 then delete;
if product_module_code=142 then delete;
if product_module_code=140 then delete;
if product_module_code=132 then delete;
if product_module_code=200 then delete;
if product_module_code=170 then delete;
if product_module_code=182 then delete;
if product_module_code=174 then delete;
if product_module_code=186 then delete;
if product_module_code=1487 then delete; *bottled water;
```

```
if product_module_code=1052 then delete; *ice pops - unfrozen;
if product_group_code=1501 then delete;
if product_group_code=1505 then delete;
if product_group_code=1506 then delete;
if product_group_code=1507 then delete;
if product_group_code=504 then delete;
if product_group_code=1001 then delete;
if product_group_code=1020 then delete;
if product_group_code=1010 then delete;
if product_group_code=511 then delete;
if product_group_code=506 then delete;
if product_group_code=1004 then delete;
if product_group_code=1005 then delete;
if product_group_code=1017 then delete;
if product_group_code=1012 then delete;
if product_group_code=1013 then delete;
if product_group_code=1007 then delete;
if product_group_code=508 then delete;
if product_group_code=1002 then delete;
if product_group_code=512 then delete;
if product_group_code=1015 then delete;
if product_group_code=503 then delete;
if product_group_code=1014 then delete;
if product_group_code=1016 then delete;
if product_group_code=510 then delete;
if product_group_code=513 then delete;
if product_group_code=1019 then delete;
if product_group_code=514 then delete;
if product_group_code=1009 then delete;
if product_group_code=1021 then delete;
if product_group_code=1008 then delete;
if product_group_code=1006 then delete; *coffee;
if product_group_code=505 then delete;
if product_group_code=1011 then delete;
if product_group_code=501 then delete;
if product_group_code=507 then delete;
if product_group_code=1018 then delete;
if product_group_code=. then delete;
if product_module_code=1046 then delete;
if product_module_code=1482 then delete;
if index(upc_descr, " SOY ")>0 then delete;
if index(brand_descr, "PEPSI WILD CH")>0 then delete;
if index(brand_descr, "PEPSI VANILLA")>0 then delete;
if index(upc_descr, " E-D ")=0 then delete;
run;
data Yang.ed_ed_data;
set Yang.test2 Yang.test1;
run;
```

```
PROC SORT DATA=Yang.ed_ed_data OUT=Yang.ed_ed_data;
  BY upc upc_ver_uc ;
RUN ;

PROC IMPORT out=Yang.Purchases2011
            DATAFILE= "Z:\Ali\Original TSV files\purchases_2011.tsv"
            dbms=dlm replace;
                              delimiter='09'x;
     GETNAMES=YES;
RUN;

PROC SORT DATA=yang.Purchases2011 OUT=yang.Purchases2011 ;
  BY upc upc_ver_uc ;
RUN ;

data Yang.TestPurchases2011;
             merge Yang.ed_ed_data Yang.Purchases2011;
             by upc upc_ver_uc;
run;

data yang.TestPurchases2011; set yang.TestPurchases2011;
   if product_module_code eq '.' then delete;
run;

data Yang.TestPurchases2011; set yang.TestPurchases2011;
   if trip_code_uc eq '.' then delete;
run;

data Yang.TestPurchases2011; set Yang.TestPurchases2011;
   Q = quantity * multi * size1_amount;
run;

data Yang.TestPurchases2011; set Yang.TestPurchases2011;
   P = total_price_paid / Q;
run;

PROC IMPORT out=Yang.Trips2011
            DATAFILE= "Z:\Ali\Original TSV files\trips_2011.tsv"
            dbms=dlm replace;
                              delimiter='09'x;
     GETNAMES=YES;
RUN;

PROC SORT DATA=Yang.Trips2011 OUT=Yang.Trips2011 ;
  BY  trip_code_uc ;
RUN ;

PROC SORT DATA=Yang.TestPurchases2011 OUT=Yang.TestPurchases2011 ;
```

```
  BY  trip_code_uc ;
RUN ;

data Yang.TestPurchasesTrips2011;
        merge Yang.TestPurchases2011 Yang.Trips2011;
        by trip_code_uc;
run;

data Yang.TestPurchasesTrips2011; set Yang.TestPurchasesTrips2011;
   if upc eq '.' then delete;
run;

PROC SORT DATA=Yang.TestPurchasesTrips2011 OUT=Yang.TestPurchasesTrips2011;
  BY  household_code ;
RUN ;

DATA yang.EnergyDrinks2011; SET yang.TestPurchasesTrips2011;
        BY household_code;
        RETAIN total_ED_exp total_ED_oz total_ED_coupon;
         IF first.household_code THEN
                DO;
                total_ED_exp = 0;
                total_ED_oz = 0;
                Total_ED_coupon = 0;
        END;
        total_ED_exp = total_ED_exp + total_price_paid;
        total_ED_oz = total_ED_oz + Q;
        total_ED_coupon = total_ED_coupon + coupon_value;
        If last.household_code then output;
        KEEP household_code total_ED_exp total_ED_oz total_ED_coupon;
        run;

        PROC IMPORT out=YAng.Panelists2011
       DATAFILE= "Z:\Ali\Original TSV files\panelists_2011.tsv"
       dbms=dlm replace;
                        delimiter='09'x;
    GETNAMES=YES;
RUN;

PROC SORT DATA=yang.panelists2011 OUT=yang.panelists2011 ;
  BY  household_code ;
RUN ;

data Yang.ED;
        merge YAng.EnergyDrinks2011 Yang.Panelists2011;
        by household_code;
run;
```

1.3 Sports Drinks:

```
libname Yang 'Z:\Yang\Yang';

data yang.test; set yang.products;
if department_code=0 then delete;
if department_code=2 then delete;
if department_code=3 then delete;
if department_code=4 then delete;
if department_code=5 then delete;
if department_code=7 then delete;
if department_code=6 then delete;
if department_code=8 then delete;
if department_code=9 then delete;
run;

data yang.test1; set yang.test;
if product_module_code=622 then delete;
if product_module_code=819 then delete;
if product_module_code=358 then delete;
if product_module_code=186 then delete;
if product_module_code=820 then delete;
if product_module_code=813 then delete;
if product_module_code=224 then delete;
if product_module_code=204 then delete;
if product_module_code=230 then delete;
if product_module_code=222 then delete;
if product_module_code=142 then delete;
if product_module_code=140 then delete;
if product_module_code=132 then delete;
if product_module_code=200 then delete;
if product_module_code=170 then delete;
if product_module_code=182 then delete;
if product_module_code=174 then delete;
if product_module_code=186 then delete;
if product_module_code=1487 then delete; *bottled water;
if product_module_code=1052 then delete; *ice pops - unfrozen;
if product_group_code=1501 then delete;
if product_group_code=1505 then delete;
if product_group_code=1506 then delete;
if product_group_code=1507 then delete;
if product_group_code=504 then delete;
if product_group_code=1001 then delete;
if product_group_code=1020 then delete;
if product_group_code=1010 then delete;
if product_group_code=511 then delete;
if product_group_code=506 then delete;
if product_group_code=1004 then delete;
```

```
if product_group_code=1005 then delete;
if product_group_code=1017 then delete;
if product_group_code=1012 then delete;
if product_group_code=1013 then delete;
if product_group_code=1007 then delete;
if product_group_code=508 then delete;
if product_group_code=1002 then delete;
if product_group_code=512 then delete;
if product_group_code=1015 then delete;
if product_group_code=503 then delete;
if product_group_code=1014 then delete;
if product_group_code=1016 then delete;
if product_group_code=510 then delete;
if product_group_code=513 then delete;
if product_group_code=1019 then delete;
if product_group_code=514 then delete;
if product_group_code=1009 then delete;
if product_group_code=1021 then delete;
if product_group_code=1008 then delete;
if product_group_code=1006 then delete; *coffee;
if product_group_code=505 then delete;
if product_group_code=1011 then delete;
if product_group_code=501 then delete;
if product_group_code=1018 then delete;
if product_group_code=. then delete;
if product_module_code=1046 then delete;
if product_module_code=1482 then delete;
if index(upc_descr, " SOY ")>0 then delete;
if index(brand_descr, "PEPSI WILD CH")>0 then delete;
if index(brand_descr, "PEPSI VANILLA")>0 then delete;
if index(upc_descr, " E-D ")>0 then delete;
if index(brand_descr, " E-J ")>0 then delete;
if index(brand_descr, "MOUNTAIN DEW")>0 then delete;
if index(brand_descr, "SIERRA MIST R")>0 then delete;
if index(brand_descr, "PEPSI DT")>0 then delete;
if index(brand_descr, "PEPSI R")>0 then delete;
if index (brand_descr, "PEPSI CAFFEIN")>0 then delete;
if index (brand_descr, "DIET RITE")>0 then delete;
if index (brand_descr, "SLICE")>0 then delete;
run;

data yang.test4; set yang.test;
if index (brand_descr, "GATORADE")=0 then delete;
run;
data Yang.test5; set yang.test;
if index (brand_descr, "POWERADE")=0 then delete;
run;
data yang.test6; set yang.test;
```

```
if index(brand_descr, "ALL SPORT R")=0 then delete;
run;
data yang.test7;set yang.test4 yang.test5 yang.test6;
run;
data yang.test8; set yang.test7;
if index(product_module_code,1050)=0 then delete ;
run;

PROC SORT DATA=Yang.Test7 OUT=Yang.Test7 ;
  BY upc upc_ver_uc ;
RUN ;

PROC IMPORT out=Yang.Purchases2011
        DATAFILE= "z:\Ali\Original TSV files\purchases_2011.tsv"
        dbms=dlm replace;
                        delimiter='09'x;
    GETNAMES=YES;
RUN;

PROC SORT DATA=yang.Purchases2011 OUT=yang.Purchases2011 ;
  BY upc upc_ver_uc ;
RUN ;

data Yang.TestPurchases2011;
         merge Yang.Test7 Yang.Purchases2011;
         by upc upc_ver_uc;
run;

data yang.TestPurchases2011; set yang.TestPurchases2011;
   if product_module_code eq '.' then delete;
run;

data Yang.TestPurchases2011; set yang.TestPurchases2011;
   if trip_code_uc eq '.' then delete;
run;

data Yang.TestPurchases2011; set Yang.TestPurchases2011;
   Q = quantity * multi * size1_amount;
run;

data Yang.TestPurchases2011; set Yang.TestPurchases2011;
   P = total_price_paid / Q;
run;

PROC IMPORT out=Yang.Trips2011
        DATAFILE= "z:\Ali\Original TSV files\trips_2011.tsv"
        dbms=dlm replace;
                        delimiter='09'x;
```

```
    GETNAMES=YES;
RUN;

PROC SORT DATA=Yang.Trips2011 OUT=Yang.Trips2011 ;
  BY  trip_code_uc ;
RUN ;

PROC SORT DATA=Yang.TestPurchases2011 OUT=Yang.TestPurchases2011 ;
  BY  trip_code_uc ;
RUN ;

data Yang.TestPurchasesTrips2011;
        merge Yang.TestPurchases2011 Yang.Trips2011;
        by trip_code_uc;
run;

data Yang.TestPurchasesTrips2011; set Yang.TestPurchasesTrips2011;
   if upc eq '.' then delete;
run;

PROC SORT DATA=Yang.TestPurchasesTrips2011 OUT=Yang.TestPurchasesTrips2011;
  BY  household_code ;
RUN ;

DATA yang.sportdrink2011; SET yang.TestPurchasesTrips2011;
        BY household_code;
        RETAIN total_sp_exp total_sp_oz total_sp_coupon;
         IF first.household_code THEN
                DO;
                total_sp_exp = 0;
                total_sp_oz = 0;
                Total_sp_coupon = 0;
        END;
        total_sp_exp = total_sp_exp + total_price_paid;
        total_sp_oz = total_sp_oz + Q;
        total_sp_coupon = total_sp_coupon + coupon_value;
        If last.household_code then output;
        KEEP household_code total_sp_exp total_sp_oz total_sp_coupon;
        run;

        PROC IMPORT out=YAng.Panelists2011
       DATAFILE= "z:\Ali\Original TSV files\panelists_2011.tsv"
       dbms=dlm replace;
                        delimiter='09'x;
    GETNAMES=YES;
RUN;

PROC SORT DATA=yang.panelists2011 OUT=yang.panelists2011 ;
```

```
  BY  household_code ;
RUN ;

data Yang.sports_drink;
        merge YAng.sportdrink2011 Yang.Panelists2011;
        by household_code;
run;
```

2. SAS codes for Tobit model:

2.1 Chocolate Milk:

```
libname Yang 'Z:\Yang\Yang';

proc means data= Yang.ch;
run;
data Yang.ch_dot;set Yang.ch;
price_chocolate = total_CH_exp/total_CH_oz;
if household_income=03 then  household_income=5;
if household_income=04 then  household_income=6.5;
if household_income=06 then  household_income=9;
if household_income=08 then  household_income=11;
if household_income=10 then  household_income=13.5;
if household_income=11 then  household_income=17.5;
if household_income=13 then  household_income=22.5;
if household_income=15 then  household_income=27.5;
if household_income=16 then  household_income=32.5;
if household_income=17 then  household_income=37.5;
if household_income=18 then  household_income=42.5;
if household_income=19 then  household_income=47.5;
if household_income=21 then  household_income=55;
if household_income=23 then  household_income=65;
if household_income=26 then  household_income=85;
if household_income=27 then  household_income=112.5;
run;

data Yang.eee;set Yang.ch_dot;
if price_chocolate ne 0 then delete;
run;
data Yang.ch_dot1;set Yang. ch_dot;
if index(household_code,"2099025")>0 then delete;
if index(household_code,"8073408")>0 then delete;
if index(household_code,"8210882")>0 then delete;
if index(household_code,"8236525")>0 then delete;
if index(household_code,"8315819")>0 then delete;
if index(household_code,"8515976")>0 then delete;
if index(household_code,"8657627")>0 then delete;
```

```
if index(household_code,"8674944")>0 then delete;
if index(household_code,"30034887")>0 then delete;
if index(household_code,"30059156")>0 then delete;
if index(household_code,"30562133")>0 then delete;
if index(household_code,"30732159")>0 then delete;
if index(household_code,"30745170")>0 then delete;
if index(household_code,"8644217")>0 then delete;
if index(household_code,"30643618")>0 then delete;
if index(household_code,"8593529")>0 then delete;
run;
proc means data=Yang.ch_dot1;
var price_chocolate  total_CH_oz  ;title "conditional mean to get the max price";
run;
data yang.ch_conditional_mean;set yang.ch_dot;
if index(household_code,"2099025")>0 then delete;
if index(household_code,"8073408")>0 then delete;
if index(household_code,"8210882")>0 then delete;
if index(household_code,"8236525")>0 then delete;
if index(household_code,"8315819")>0 then delete;
if index(household_code,"8515976")>0 then delete;
if index(household_code,"8657627")>0 then delete;
if index(household_code,"8674944")>0 then delete;
if index(household_code,"30034887")>0 then delete;
if index(household_code,"30059156")>0 then delete;
if index(household_code,"30562133")>0 then delete;
if index(household_code,"30732159")>0 then delete;
if index(household_code,"30745170")>0 then delete;
if index(household_code,"8644217")>0 then delete;
if index(household_code,"30643618")>0 then delete;
if index(household_code,"8593529")>0 then delete;
if index(household_code,"2001406")>0 then delete;
if index(household_code,"2001670")>0 then delete;
if index(household_code,"2001777")>0 then delete;
if index(household_code,"2018988")>0 then delete;
if index(household_code,"2029790")>0 then delete;
if index(household_code,"2049206")>0 then delete;
if index(household_code,"2050892")>0 then delete;
if index(household_code,"2057150")>0 then delete;
if index(household_code,"2066676")>0 then delete;
if index(household_code,"2083426")>0 then delete;
if index(household_code,"2097247")>0 then delete;
if index(household_code,"2098235")>0 then delete;
if index(household_code,"8005801")>0 then delete;
if index(household_code,"8030882")>0 then delete;
if index(household_code,"8045762")>0 then delete;
if index(household_code,"8046245")>0 then delete;
if index(household_code,"8074841")>0 then delete;
if index(household_code,"8096931")>0 then delete;
```

```
if index(household_code,"8123724")>0 then delete;
if index(household_code,"8162112")>0 then delete;
if index(household_code,"8176181")>0 then delete;
if index(household_code,"8191323")>0 then delete;
if index(household_code,"8193635")>0 then delete;
if index(household_code,"8203954")>0 then delete;
if index(household_code,"8231965")>0 then delete;
if index(household_code,"8245003")>0 then delete;
if index(household_code,"8251995")>0 then delete;
if index(household_code,"8253652")>0 then delete;
if index(household_code,"8270506")>0 then delete;
if index(household_code,"8277284")>0 then delete;
if index(household_code,"8314713")>0 then delete;

if index(household_code,"8327678")>0 then delete;
if index(household_code,"8331605")>0 then delete;
if index(household_code,"8343180")>0 then delete;
if index(household_code,"8350288")>0 then delete;
if index(household_code,"8370169")>0 then delete;

if index(household_code,"8420574")>0 then delete;
if index(household_code,"8529905")>0 then delete;
if index(household_code,"8579702")>0 then delete;
if index(household_code,"8599531")>0 then delete;
if index(household_code,"8607537")>0 then delete;
if index(household_code,"8633094")>0 then delete;
if index(household_code,"8634622")>0 then delete;
if index(household_code,"8644988")>0 then delete;
if index(household_code,"9005129")>0 then delete;
if index(household_code,"9150684")>0 then delete;
if index(household_code,"9170619")>0 then delete;
run;
proc means data= yang.ch_conditional_mean;
var price_chocolate  total_CH_oz  ;title "conditional mean ";
run;

data Yang.ch_delete;set Yang. ch;
if index(household_code,"2099025")>0 then delete;
if index(household_code,"8073408")>0 then delete;
if index(household_code,"8210882")>0 then delete;
if index(household_code,"8236525")>0 then delete;
if index(household_code,"8315819")>0 then delete;
if index(household_code,"8515976")>0 then delete;
if index(household_code,"8657627")>0 then delete;
if index(household_code,"8674944")>0 then delete;
if index(household_code,"30034887")>0 then delete;
if index(household_code,"30059156")>0 then delete;
if index(household_code,"30562133")>0 then delete;
```

```
if index(household_code,"30732159")>0 then delete;
if index(household_code,"30745170")>0 then delete;
if index(household_code,"8644217")>0 then delete;
if index(household_code,"30643618")>0 then delete;
if index(household_code,"8593529")>0 then delete;
if index(household_code,"2001406")>0 then delete;
if index(household_code,"2001670")>0 then delete;
if index(household_code,"2001777")>0 then delete;
if index(household_code,"2018988")>0 then delete;
if index(household_code,"2029790")>0 then delete;
if index(household_code,"2049206")>0 then delete;
if index(household_code,"2050892")>0 then delete;
if index(household_code,"2057150")>0 then delete;
if index(household_code,"2066676")>0 then delete;
if index(household_code,"2083426")>0 then delete;
if index(household_code,"2097247")>0 then delete;
if index(household_code,"2098235")>0 then delete;
if index(household_code,"8005801")>0 then delete;
if index(household_code,"8030882")>0 then delete;
if index(household_code,"8045762")>0 then delete;
if index(household_code,"8046245")>0 then delete;
if index(household_code,"8074841")>0 then delete;
if index(household_code,"8096931")>0 then delete;
if index(household_code,"8123724")>0 then delete;
if index(household_code,"8162112")>0 then delete;
if index(household_code,"8176181")>0 then delete;
if index(household_code,"8191323")>0 then delete;
if index(household_code,"8193635")>0 then delete;
if index(household_code,"8203954")>0 then delete;
if index(household_code,"8231965")>0 then delete;
if index(household_code,"8245003")>0 then delete;
if index(household_code,"8251995")>0 then delete;
if index(household_code,"8253652")>0 then delete;
if index(household_code,"8270506")>0 then delete;
if index(household_code,"8277284")>0 then delete;
if index(household_code,"8314713")>0 then delete;

if index(household_code,"8327678")>0 then delete;
if index(household_code,"8331605")>0 then delete;
if index(household_code,"8343180")>0 then delete;
if index(household_code,"8350288")>0 then delete;
if index(household_code,"8370169")>0 then delete;

if index(household_code,"8420574")>0 then delete;
if index(household_code,"8529905")>0 then delete;
if index(household_code,"8579702")>0 then delete;
if index(household_code,"8599531")>0 then delete;
if index(household_code,"8607537")>0 then delete;
```

```
if index(household_code,"8633094")>0 then delete;
if index(household_code,"8634622")>0 then delete;
if index(household_code,"8644988")>0 then delete;
if index(household_code,"9005129")>0 then delete;
if index(household_code,"9150684")>0 then delete;
if index(household_code,"9170619")>0 then delete;
run;
data Yang.ch_first;set Yang.ch_delete;
if household_size=1 then hsize=1;
if household_size=2 then hsize=2;
if household_size=3 then hsize=3;
if household_size=4 then hsize=4;
if household_size=5 then hsize=5;
if household_size=6 then hsize=6;
if household_size=7 then hsize=7;
if household_size=8 then hsize=8;
if household_size=9 then hsize=9;

if household_income=03 then hinc=5;
if household_income=04 then hinc=6.5;
if household_income=06 then hinc=9;
if household_income=08 then hinc=11;
if household_income=10 then hinc=13.5;
if household_income=11 then hinc=17.5;
if household_income=13 then hinc=22.5;
if household_income=15 then hinc=27.5;
if household_income=16 then hinc=32.5;
if household_income=17 then hinc=37.5;
if household_income=18 then hinc=42.5;
if household_income=19 then hinc=47.5;
if household_income=21 then hinc=55;
if household_income=23 then hinc=65;
if household_income=26 then hinc=85;
if household_income=27 then hinc=112.5;
*age of household head "agehh",agef is agefemale;
agehh=female_head_age;
if female_head_age=0 then agehh=male_head_age;
if agehh=1 then agehhlt25=1; else agehhlt25=0;
if agehh=2 then agehh2529=1; else agehh2529=0;
if agehh=3 then agehh3034=1; else agehh3034=0;
if agehh=4 or agehh=5 then agehh3544=1; else agehh3544=0;
if agehh=6 or agehh=7 then agehh4554=1; else agehh4554=0;
if agehh=8 then agehh5564=1; else agehh5564=0;
if agehh=9 then agehhgt64=1; else agehhgt64=0;
*employment status of the household head "emphh";
emphh=female_head_employment;
if female_head_employment=0 then emphh=male_head_employment;
if emphh=1 or emphh=2 then emphhpt=1; else emphhpt=0;
```

```
if emphh=3 then emphhft=1; else emphhft=0;
if emphh=9 then emphhnfp=1; else emphhnfp=0;
*education of the household head "eduhh";
eduhh=female_head_education;
if female_head_education=0 then eduhh=male_head_education;
if eduhh=1 or eduhh=2 then eduhhlths=1; else eduhhlths=0;
if eduhh=3 then eduhhhs=1; else eduhhhs=0;
if eduhh=4 or eduhh=5 then eduhhu=1; else eduhhu=0;
if eduhh=6 then eduhhpc=1; else eduhhpc=0;
*race of the household;
if race=1 then White=1; else White=0;
if race=2 then Black=1; else Black=0;
if race=3 then Oriental=1; else Oriental=0;
if race=4 then Other=1; else Other=0;

*hispanic origin;
if hispanic_origin=1 then hisp_yes=1; else hisp_yes=0;
if hispanic_origin=2 then hisp_no=1; else hisp_no=0;

*region;
if region_code=1 then NewEngland=1; else NewEngland=0;
if region_code=2 then MiddleAtlantic=1; else  MiddleAtlantic=0;
if region_code=3 then EastNorthCentral=1; else EastNorthCentral=0;
if region_code=4 then WestNorthCentral=1; else WestNorthCentral=0;
if region_code=5 then SouthAtlantic=1; else SouthAtlantic=0;
if region_code=6 then EastSouthCentral=1; else EastSouthCentral=0;
if region_code=7 then WestSouthCentral=1; else WestSouthCentral=0;
if region_code=8 then Mountain=1; else Mountain=0;
if region_code=9 then Pacific=1; else Pacific=0;

*age and presence of children "ac";
if age_and_presence_of_children=1 then aclt6_only=1; else aclt6_only=0;
if age_and_presence_of_children=2 then ac6_12only=1; else ac6_12only=0;
if age_and_presence_of_children=3 then ac13_17only=1; else ac13_17only=0;
if age_and_presence_of_children=4 then aclt6_6_12only=1; else aclt6_6_12only=0;
if age_and_presence_of_children=5 then aclt6_13_17only=1; else aclt6_13_17only=0;
if age_and_presence_of_children=6 then ac6_12and13_17only=1; else ac6_12and13_17only=0;
if age_and_presence_of_children=7 then aclt6_6_12and13_17=1; else aclt6_6_12and13_17=0;
if age_and_presence_of_children=9 then no_child=1; else no_child=0;
*houdehold head's gender "fhonly=femal household head only" and "mhonly=male household
head only";
if male_head_age=0 then fhonly=1; else fhonly=0;
if female_head_age=0 then mhonly=1; else mhonly=0;
if female_head_age ne 0 and male_head_age ne 0 then fhmh=1; else fhmh=0;

price_chocolate = total_CH_exp/total_CH_oz;
loghinc=log(hinc);
*create a dummy variable for the quantity of soymilk;
```

```
if total_CH_oz>0 then Dummy_CH=1; else Dummy_CH=0;
*replace dots (.s) with a zero;

if total_CH_oz=.  then total_CH_oz=0;
if total_CH_exp=. then total_CH_exp=0;
if total_CH_coupon=. then total_CH_coupon=0;
*renaming quanitity purchased;
Q_CH=total_CH_oz;

keep
household_code
hinc
hsize
agehhlt25
agehh2529
agehh3034
agehh3544
agehh4554
agehh5564
agehhgt64
emphhpt
emphhft
emphhnfp
eduhhlths
eduhhhs
eduhhu
eduhhpc
white
black
oriental
other

hisp_yes
hisp_no

NewEngland
MiddleAtlantic
EastNorthCentral
WestNorthCentral
SouthAtlantic
EastSouthCentral
WestSouthCentral
Mountain
Pacific

aclt6_only
ac6_12only
ac13_17only
```

```
aclt6_6_12only
aclt6_13_17only
ac6_12and13_17only
aclt6_6_12and13_17
no_child

fhonly
mhonly
fhmh

price_chocolate
loghinc

Q_CH
Dummy_CH
total_CH_oz
total_CH_exp
total_CH_coupon

run;
proc means data=Yang.ch_first; title " mean";
run;
data yang.income_chocolate;set yang.ch_first;
if index(total_ch_oz or total_ch_exp,"0")>0 then delete;
run;

proc means data=yang.income_chocolate;
run;

* to get the auxiliary regression;
Proc reg data= Yang.ch_first;
model price_chocolate=loghinc hsize NewEngland MiddleAtlantic EastNorthCentral
WestNorthCentral SouthAtlantic EastSouthCentral WestSouthCentral Mountain;
RUN;

data Yang.ch_second; set Yang.ch_first;
if (total_CH_oz ne 0 or total_CH_exp ne 0) then price_chocolate=total_CH_exp/total_CH_oz;
if (total_CH_oz = 0 or total_CH_exp = 0)
then price_chocolate=0.05787 + 0.00033378*loghinc -0.00115*hsize-0.00207*NewEngland
-0.00141*MiddleAtlantic -0.01724*EastNorthCentral-0.01524*westNorthCentral
-0.00145*SouthAtlantic-0.00688*EastSouthCentral-0.00167*WestSouthCentral
-0.00716*Mountain;
RUN;
proc means data=Yang.ch_second;
var price_chocolate   Q_CH   Dummy_CH  hinc;title "unconditional mean of consumption";
run;

data Yang.imputed_price_test; set Yang.ch_first;
```

```
if (total_CH_oz ne 0 or total_CH_exp ne 0) then price_chocolate=. ;
if (total_CH_oz = 0 or total_CH_exp = 0)
then price_chocolate=0.05787 + 0.00033378*loghinc -0.00115*hsize-0.00207*NewEngland
-0.00141*MiddleAtlantic -0.01724*EastNorthCentral-0.01524*westNorthCentral
-0.00145*SouthAtlantic-0.00688*EastSouthCentral-0.00167*WestSouthCentral
-0.00716*Mountain;
RUN;
proc means data= Yang.imputed_price_test;
var price_chocolate ; title "mean of imputed price";
run;

libname Yang 'Z:\Yang\Yang';
data Yang.tobit; merge Yang.ch_second Yang.ed_second Yang.sp_second;
by household_code;
run;

data Yang.tobit1 ; set Yang.tobit;
logprice_chocolate=log(price_chocolate);
logprice_energy=log(price_energy);
logprice_sportsdrink=log(price_sportsdrink);
loghinc=log(hinc);
run;

Proc QLIM data=Yang.tobit1 ;
        model Q_CH= logprice_chocolate logprice_energy logprice_sportsdrink hsize loghinc
agehh2529 agehh3034 agehh3544 agehh4554 agehh5564 agehhgt64 emphhpt emphhft eduhhhs
eduhhu eduhhpc
black oriental other hisp_yes newengland middleatlantic eastnorthcentral westnorthcentral
southatlantic eastsouthcentral westsouthcentral mountain aclt6_only ac6_12only ac13_17only
aclt6_6_12only aclt6_13_17only ac6_12and13_17only aclt6_6_12and13_17
fhonly mhonly;
        endogenous Q_CH ~ censored(lowerbound=0);
        nloptions maxiter=500; /*maximum number of iterations set at 300*/
        *hetero Q_CH ~ P_chocolate P_energy hinc agehh2529 agehh3034 agehh3544
agehh4554 agehh5564 agehhgt64 emphhpt emphhft eduhhhs eduhhu eduhhpc
                east midwest south black oriental other hisp_yes aclt6_only ac6_12only
ac13_17only aclt6_6_12only aclt6_13_17only
                ac6_12and13_17only aclt6_6_12and13_17 fhonly mhonly;
        output out=Yang.Tobit_out conditional expected marginal xbeta;*conditional expected
value of Q_CH,
        unconditional expected value of Q_CH,
        unconditional marginal effect of all variables, xbeta;
run;

Proc means data=Yang.Tobit_out;
var  Meff_logPrice_chocolate  Meff_logPrice_energy
Meff_logPrice_sportsdrink;*unconditional marginal effect of logprice_chocolate and
logprice_energy;
```

run; *meff represents the unconditional marginal effect;

data Yang.CH_marginal;set Yang.Tobit_out;
* z is the normalized index,sas does not provide z directly,z=xbeta_y/sigma;
z=xbeta_Q_ch/**1141.306862**;
*capfz is the cdf standard normal,sas does not provide capfz directly;
capfz=probnorm(z);*F(z);
*fz is the standard normal density function,sas does not provide capfz directly, * 2.5066272 is the result of (1/sqrt(2*pi));;
fz=exp(-z****2**/**2**)/**2.5066272**;
*expected_Q_CH is the unconditional expected value of the dependent variable;
*expected_Q_ch also serve as the predicted value of the dependent variable;
*sas captures the unconditional expected values;
expected_Q_CH=xbeta_Q_ch*capfz+**1141.306862***fz;
*cexpected_Q_CH is the conditional expected value of the dependent variable;
*sas captures the condtional expected values;
cexpected_Q_CH=xbeta_Q_ch+**1141.306862***fz/capfz;

*unconditional marginal effects of price and household income;*sas captures the unconditional marginal effects of variables(logprice,loghinc);
meff_logprice_chocolate=(**-1008.419964**)*capfz;*this is captured by sas automatically;
meff_price_chocolate=(**-1008.419964**)*capfz/**0.0503423**;*0.05034123 is the unconditional mean of chocolate price;
meff_logprice_energy=(**-146.672036**)*capfz;
meff_price_energy=(**-146.672036**)*capfz/**0.1309351**;*0.1309351 is the unconditional mean of energy price;
meff_logprice_sportsdrink=(**-161.245539**)*capfz;*this is captured by sas automatically;
meff_price_sportsdrink=(**-161.245539**)*capfz/**0.0527959**;*0.05247959 is the unconditional mean of chocolate price;

meff_loghinc=(**-18.491725**)*capfz;
meff_hinc=(**-18.491725**)*capfz/**58.3209305**;*58.3209305 is the unconditional mean of income;
*unconditional Elasticities for chocolate milk;
Un_own_price_chocholate=((**-1008.419964**)*capfz)/**110.3789024**;* 110.3789024 is the uncondtional mean of Q_CH;
Un_cross_price_energy=((**-146.672036**)*capfz)/**110.3789024**;* -125.059086 is the beta of logprice_energy;
Un_cross_price_sportsdrink=((**-161.245539**)*capfz)/**110.3789024**;* -162.157154 is the beta of logprice_sportsdrink;
Un_income_Ch=((**-18.491725**)*capfz)/**110.3789024**;*-18.491725 is the beta of hinc;

* conditional marginal effects for price;
c_meff_logprice_chocolate= (**-1008.419964**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz));
c_meff_price_chocolate=(**-1008.419964**)*(**1**-z*fz/capfz-
(fz/capfz)*(fz/capfz))/**0.0485519**;*0.0485468 is the conditional mean of chocolate price;
c_meff_logprice_energy=(**-146.672036**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz));
c_meff_price_energy=(**-146.672036**)*(**1**-z*fz/capfz-
(fz/capfz)*(fz/capfz))/**0.1294626**;*0.1294626 is the conditional mean of energy price;

c_meff_logprice_sportsdrink=(**-161.245539**)*(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz));
c_meff_price_sportsdrink=(**-161.245539**)*(**1-**z*fz/capfz-
(fz/capfz)*(fz/capfz))/**0.0523869**;*0.0523869 is the conditional mean of sportsdrink price;

c_meff_loghinc=(**-18.491725**)*(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz));
c_meff_hinc=(**-18.491725**)*(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))/**58.6873842**; *58.6873842 is the conditional mean of hinc;
*conditional marginal effects;
c_Meff_hsize=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))***75.453887**;
c_meff_agehh2529=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-37.253199**);
c_meff_agehh3034=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**43.273577**);

c_meff_agehh3544=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**78.045181**);
c_meff_agehh4554=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**100.925704**);

c_meff_agehh5564=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**16.195370**);
c_meff_agehhgt64=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-187.128383**);
c_meff_emphhpt=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-11.394092**);

c_meff_emphhft=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-2.427708**);
c_meff_eduhhhs=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-27.461497**);
c_meff_eduhhu=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-111.105331**);
c_meff_eduhhpc=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-237.184717**);

c_meff_black=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-336.189108**);
c_meff_Oriental=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-243.596435**);
c_meff_other=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-77.726278**);

c_meff_hisp_yes=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-74.540999**);
c_meff_newengland=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-40.967165**);
c_meff_middleatlantic=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))* **169.628833**;
c_meff_eastnorthcentral=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**116.345409**);

c_meff_westnorthcentral=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**161.001819**);
c_meff_southatlantic=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**63.862200**);
c_meff_eastsouthcentral=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**201.578866**);
c_meff_westsouthcentral=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))* **0.915146**;
c_meff_mountain=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-65.702931**);

c_meff_aclt6_only=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**65.630548**);
c_meff_ac6_12only=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**155.272847**);

c_meff_ac13_17only=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))***173.558739**;

c_meff_aclt6_6_12only=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**103.170134**);
c_meff_aclt6_13_17only=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**150.509949**);
c_meff_ac6_12and13_17only=(**1-**z*fz/capfz-(fz/capfz)*(fz/capfz))*(**112.645866**);

c_meff_aclt6_6_12and13_17= (**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**103.383587**);
c_meff_fhonly=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-69.177739**);
c_meff_mhonly=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-101.556465**);

**now to calculate the marginal change in probability, we use the McDonald and Moffitt Decomposition
Reference:
McDonald, J.F., and R. A. Moffitt, "The Uses of Tobit Analysis" The Review of Economics and Statistics, May 1980, 62(2):318-321;

*probability of being above the limit for chocolate,formula 17;
pr_meff_logprice_chocolate=(meff_logprice_chocolate-capfz*(c_meff_logprice_chocolate))/cexpected_Q_CH;
pr_meff_logprice_energy=(meff_logprice_energy-capfz*(c_meff_logprice_energy))/cexpected_Q_CH;
pr_meff_logprice_sportsdrink=(meff_logprice_sportsdrink-capfz*(c_meff_logprice_sportsdrink))/cexpected_Q_CH;

pr_Meff_hsize=(meff_hsize-capfz*(c_meff_hsize))/cexpected_Q_CH;
pr_meff_agehh2529=(meff_agehh2529-capfz*(c_meff_agehh2529))/cexpected_Q_CH;
pr_meff_agehh3034=(meff_agehh3034-capfz*(c_meff_agehh3034))/cexpected_Q_CH;
pr_meff_agehh3544=(meff_agehh3544-capfz*(c_meff_agehh3544))/cexpected_Q_CH;
pr_meff_agehh4554=(meff_agehh4554-capfz*(c_meff_agehh4554))/cexpected_Q_CH;
pr_meff_agehh5564=(meff_agehh5564-capfz*(c_meff_agehh5564))/cexpected_Q_CH;
pr_meff_agehhgt64=(meff_agehhgt64-capfz*(c_meff_agehhgt64))/cexpected_Q_CH;
pr_meff_emphhpt=(meff_emphhpt-capfz*(c_meff_emphhpt))/cexpected_Q_CH;
pr_meff_emphhft=(meff_emphhft-capfz*(c_meff_emphhft))/cexpected_Q_CH;
pr_meff_eduhhhs=(meff_eduhhhs-capfz*(c_meff_eduhhhs))/cexpected_Q_CH;
pr_meff_eduhhu=(meff_eduhhu-capfz*(c_meff_eduhhu))/cexpected_Q_CH;
pr_meff_eduhhpc=(meff_eduhhpc-capfz*(c_meff_eduhhpc))/cexpected_Q_CH;
pr_meff_black=(meff_black-capfz*(c_meff_black))/cexpected_Q_CH;
pr_meff_Oriental=(meff_Oriental-capfz*(c_meff_Oriental))/cexpected_Q_CH;
pr_meff_other=(meff_other-capfz*(c_meff_other))/cexpected_Q_CH;
pr_meff_hisp_yes=(meff_hisp_yes-capfz*(c_meff_hisp_yes))/cexpected_Q_CH;
pr_meff_newengland=(meff_newengland-capfz*(c_meff_newengland))/cexpected_Q_CH;
pr_meff_middleatlantic=(meff_middleatlantic-capfz*(c_meff_middleatlantic))/cexpected_Q_CH;
pr_meff_eastnorthcentral=(meff_eastnorthcentral-capfz*(c_meff_eastnorthcentral))/cexpected_Q_CH;
pr_meff_westnorthcentral=(meff_westnorthcentral-capfz*(c_meff_westnorthcentral))/cexpected_Q_CH;
pr_meff_southatlantic=(meff_southatlantic-capfz*(c_meff_southatlantic))/cexpected_Q_CH;
pr_meff_eastsouthcentral=(meff_eastsouthcentral-capfz*(c_meff_eastsouthcentral))/cexpected_Q_CH;
pr_meff_westsouthcentral=(meff_westsouthcentral-capfz*(c_meff_westsouthcentral))/cexpected_Q_CH;
pr_meff_mountain=(meff_mountain-capfz*(c_meff_mountain))/cexpected_Q_CH;
pr_meff_aclt6_only=(meff_aclt6_only-capfz*(c_meff_aclt6_only))/cexpected_Q_CH;
pr_meff_ac6_12only=(meff_ac6_12only-capfz*(c_meff_ac6_12only))/cexpected_Q_CH;

```
pr_meff_ac13_17only=(meff_ac13_17only-capfz*(c_meff_ac13_17only))/cexpected_Q_CH;
pr_meff_aclt6_6_12only=(meff_aclt6_6_12only-
capfz*(c_meff_aclt6_6_12only))/cexpected_Q_CH;
pr_meff_aclt6_13_17only=(meff_aclt6_13_17only-
capfz*(c_meff_aclt6_13_17only))/cexpected_Q_CH;
pr_meff_ac6_12and13_17only=(meff_ac6_12and13_17only-
capfz*(c_meff_ac6_12and13_17only))/cexpected_Q_CH;
pr_meff_aclt6_6_12and13_17=(meff_aclt6_6_12and13_17-
capfz*(c_meff_aclt6_6_12and13_17))/cexpected_Q_CH;
pr_meff_fhonly=(meff_fhonly-capfz*(c_meff_fhonly))/cexpected_Q_CH;
pr_meff_mhonly=(meff_mhonly-capfz*(c_meff_mhonly))/cexpected_Q_CH;

*conditional elasticities;
C_Own_Price_chocolate=(-1008.419964*(1-z*fz/capfz-(fz/capfz)*(fz/capfz)))/423.0009228;
* 423.0009228 is the  condtional mean of total _ch_oz;
C_cross_price_energy=(-146.672036*(1-z*fz/capfz-(fz/capfz)*(fz/capfz)))/423.0009228;
* -125.059086 is the beta of logprice_energy;
C_cross_price_sportsdrink=(-161.245539*(1-z*fz/capfz-(fz/capfz)*(fz/capfz)))/423.0009228;
C_income_Ch=(-18.491725*(1-z*fz/capfz-(fz/capfz)*(fz/capfz)))/423.0009228;
*-0.471297 is the beta of hinc;

run;

proc means data=Yang.CH_marginal n mean median std min max;
var
Un_own_price_chocholate Un_cross_price_energy Un_cross_price_sportsdrink Un_income_Ch
meff_logprice_chocolate meff_price_chocolate

meff_logprice_energy
meff_price_energy
meff_logprice_sportsdrink
meff_price_sportsdrink

Meff_hsize  meff_agehh2529 meff_agehh3034 meff_agehh3544
meff_agehh4554 meff_agehh5564 meff_agehhgt64 meff_emphhpt meff_emphhft meff_eduhhhs
meff_eduhhu
meff_eduhhpc meff_black meff_oriental meff_other meff_hisp_yes meff_newengland
meff_middleatlantic
meff_eastnorthcentral meff_westnorthcentral meff_southatlantic meff_eastsouthcentral
meff_westsouthcentral
meff_mountain meff_aclt6_only meff_ac6_12only meff_ac13_17only meff_aclt6_6_12only
meff_aclt6_13_17only
meff_ac6_12and13_17only meff_aclt6_6_12and13_17  meff_fhonly meff_mhonly;
title "unconditional elasticity ch, unconditional marginal effects ch";
run;
data Yang.CH_marginal2 ;set Yang.CH_marginal;
if total_CH_oz>0;
```

```
run;*select all the household who buy ch,in this part we ignore whether the hh buy energy, sport ;
proc means data=Yang.CH_marginal2 n mean median std min max;
var
C_Own_Price_chocolate  C_cross_price_energy C_cross_price_sportsdrink C_income_Ch

c_meff_logprice_chocolate
c_meff_price_chocolate
c_meff_logprice_energy
c_meff_price_energy
c_meff_logprice_sportsdrink
c_meff_price_sportsdrink

c_meff_hinc

C_Meff_hsize  C_meff_agehh2529 c_meff_agehh3034 c_meff_agehh3544
c_meff_agehh4554 c_meff_agehh5564 C_meff_agehhgt64 C_meff_emphhpt C_meff_emphhft
c_meff_eduhhhs c_meff_eduhhu
c_meff_eduhhpc c_meff_black c_meff_oriental c_meff_other c_meff_hisp_yes
c_meff_newengland c_meff_middleatlantic
c_meff_eastnorthcentral c_meff_westnorthcentral c_meff_southatlantic c_meff_eastsouthcentral
c_meff_westsouthcentral
c_meff_mountain c_meff_aclt6_only c_meff_ac6_12only c_meff_ac13_17only
c_meff_aclt6_6_12only c_meff_aclt6_13_17only
c_meff_ac6_12and13_17only c_meff_aclt6_6_12and13_17  c_meff_fhonly c_meff_mhonly;
title "conditional marginal effects ch";
run;
proc means data=Yang.CH_marginal2 n mean median std min max;
var
pr_meff_logprice_chocolate
pr_meff_logprice_energy
pr_meff_logprice_sportsdrink

pr_Meff_hsize
pr_meff_agehh2529
pr_meff_agehh3034
pr_meff_agehh3544
pr_meff_agehh4554
pr_meff_agehh5564
pr_meff_agehhgt64
pr_meff_emphhpt
pr_meff_emphhft
pr_meff_eduhhhs
pr_meff_eduhhu
pr_meff_eduhhpc
pr_meff_black
pr_meff_Oriental
pr_meff_other
```

```
pr_meff_hisp_yes
pr_meff_newengland
pr_meff_middleatlantic
pr_meff_eastnorthcentral
pr_meff_westnorthcentral
pr_meff_southatlantic
pr_meff_eastsouthcentral
pr_meff_westsouthcentral
pr_meff_mountain
pr_meff_aclt6_only
pr_meff_ac6_12only
pr_meff_ac13_17only
pr_meff_aclt6_6_12only
pr_meff_aclt6_13_17only
pr_meff_ac6_12and13_17only
pr_meff_aclt6_6_12and13_17
pr_meff_fhonly
pr_meff_mhonly;
title "mean of the marginal change in probability-ch";
run;
```

2.2 Energy Drinks:

```
libname Yang 'Z:\Yang\Yang';
proc means data= Yang.ed;
run;
data Yang.ed_dot;set Yang.ed;
price_energy = total_ed_exp/total_ed_oz;
if household_income=03 then  household_income=5;
if household_income=04 then  household_income=6.5;
if household_income=06 then  household_income=9;
if household_income=08 then  household_income=11;
if household_income=10 then  household_income=13.5;
if household_income=11 then  household_income=17.5;
if household_income=13 then  household_income=22.5;
if household_income=15 then  household_income=27.5;
if household_income=16 then  household_income=32.5;
if household_income=17 then  household_income=37.5;
if household_income=18 then  household_income=42.5;
if household_income=19 then  household_income=47.5;
if household_income=21 then  household_income=55;
if household_income=23 then  household_income=65;
if household_income=26 then  household_income=85;
if household_income=27 then  household_income=112.5;
run;

data Yang.hhhh;set Yang.ed_dot;
if price_energy ne 0 then delete;
```

```
run;

data Yang.ed_dot1;set Yang.ed_dot;
if index(household_code,"2099025")>0 then delete;
if index(household_code,"8073408")>0 then delete;
if index(household_code,"8210882")>0 then delete;
if index(household_code,"8236525")>0 then delete;
if index(household_code,"8315819")>0 then delete;
if index(household_code,"8515976")>0 then delete;
if index(household_code,"8657627")>0 then delete;
if index(household_code,"8674944")>0 then delete;
if index(household_code,"30034887")>0 then delete;
if index(household_code,"30059156")>0 then delete;
if index(household_code,"30562133")>0 then delete;
if index(household_code,"30732159")>0 then delete;
if index(household_code,"30745170")>0 then delete;
if index(household_code,"8644217")>0 then delete;
if index(household_code,"30643618")>0 then delete;
if index(household_code,"8593529")>0 then delete;

run;

proc means data=Yang.ed_dot1;
var price_energy  total_ed_oz  ;title "conditional mean if price_chocolate=0 then
price_chocolate=.";
run;

data Yang.ed_delete;set Yang.ed;
if index(household_code,"2099025")>0 then delete;
if index(household_code,"8073408")>0 then delete;
if index(household_code,"8210882")>0 then delete;
if index(household_code,"8236525")>0 then delete;
if index(household_code,"8315819")>0 then delete;
if index(household_code,"8515976")>0 then delete;
if index(household_code,"8657627")>0 then delete;
if index(household_code,"8674944")>0 then delete;
if index(household_code,"30034887")>0 then delete;
if index(household_code,"30059156")>0 then delete;
if index(household_code,"30562133")>0 then delete;
if index(household_code,"30732159")>0 then delete;
if index(household_code,"30745170")>0 then delete;
if index(household_code,"8644217")>0 then delete;
if index(household_code,"30643618")>0 then delete;
if index(household_code,"8593529")>0 then delete;
if index(household_code,"2001406")>0 then delete;
if index(household_code,"2001670")>0 then delete;
if index(household_code,"2001777")>0 then delete;
if index(household_code,"2018988")>0 then delete;
```

```
if index(household_code,"2029790")>0 then delete;
if index(household_code,"2049206")>0 then delete;
if index(household_code,"2050892")>0 then delete;
if index(household_code,"2057150")>0 then delete;
if index(household_code,"2066676")>0 then delete;
if index(household_code,"2083426")>0 then delete;
if index(household_code,"2097247")>0 then delete;
if index(household_code,"2098235")>0 then delete;
if index(household_code,"8005801")>0 then delete;
if index(household_code,"8030882")>0 then delete;
if index(household_code,"8045762")>0 then delete;
if index(household_code,"8046245")>0 then delete;
if index(household_code,"8074841")>0 then delete;
if index(household_code,"8096931")>0 then delete;
if index(household_code,"8123724")>0 then delete;
if index(household_code,"8162112")>0 then delete;
if index(household_code,"8176181")>0 then delete;
if index(household_code,"8191323")>0 then delete;
if index(household_code,"8193635")>0 then delete;
if index(household_code,"8203954")>0 then delete;
if index(household_code,"8231965")>0 then delete;
if index(household_code,"8245003")>0 then delete;
if index(household_code,"8251995")>0 then delete;
if index(household_code,"8253652")>0 then delete;
if index(household_code,"8270506")>0 then delete;
if index(household_code,"8277284")>0 then delete;
if index(household_code,"8314713")>0 then delete;

if index(household_code,"8327678")>0 then delete;
if index(household_code,"8331605")>0 then delete;
if index(household_code,"8343180")>0 then delete;
if index(household_code,"8350288")>0 then delete;
if index(household_code,"8370169")>0 then delete;

if index(household_code,"8420574")>0 then delete;
if index(household_code,"8529905")>0 then delete;
if index(household_code,"8579702")>0 then delete;
if index(household_code,"8599531")>0 then delete;
if index(household_code,"8607537")>0 then delete;
if index(household_code,"8633094")>0 then delete;
if index(household_code,"8634622")>0 then delete;
if index(household_code,"8644988")>0 then delete;
if index(household_code,"9005129")>0 then delete;
if index(household_code,"9150684")>0 then delete;
if index(household_code,"9170619")>0 then delete;
run;
data Yang.ed_first;set Yang.ed_delete;
if household_size=1 then hsize=1;
```

```
if household_size=2 then hsize=2;
if household_size=3 then hsize=3;
if household_size=4 then hsize=4;
if household_size=5 then hsize=5;
if household_size=6 then hsize=6;
if household_size=7 then hsize=7;
if household_size=8 then hsize=8;
if household_size=9 then hsize=9;
if household_income=03 then hinc=5;
if household_income=04 then hinc=6.5;
if household_income=06 then hinc=9;
if household_income=08 then hinc=11;
if household_income=10 then hinc=13.5;
if household_income=11 then hinc=17.5;
if household_income=13 then hinc=22.5;
if household_income=15 then hinc=27.5;
if household_income=16 then hinc=32.5;
if household_income=17 then hinc=37.5;
if household_income=18 then hinc=42.5;
if household_income=19 then hinc=47.5;
if household_income=21 then hinc=55;
if household_income=23 then hinc=65;
if household_income=26 then hinc=85;
if household_income=27 then hinc=112.5;
*age of household head "agehh",agef is agefemale;
agehh=female_head_age;
if female_head_age=0 then agehh=male_head_age;
if agehh=1 then agehhlt25=1; else agehhlt25=0;
if agehh=2 then agehh2529=1; else agehh2529=0;
if agehh=3 then agehh3034=1; else agehh3034=0;
if agehh=4 or agehh=5 then agehh3544=1; else agehh3544=0;
if agehh=6 or agehh=7 then agehh4554=1; else agehh4554=0;
if agehh=8 then agehh5564=1; else agehh5564=0;
if agehh=9 then agehhgt64=1; else agehhgt64=0;
*employment status of the household head "emphh";
emphh=female_head_employment;
if female_head_employment=0 then emphh=male_head_employment;
if emphh=1 or emphh=2 then emphhpt=1; else emphhpt=0;
if emphh=3 then emphhft=1; else emphhft=0;
if emphh=9 then emphhnfp=1; else emphhnfp=0;
*education of the household head "eduhh";
eduhh=female_head_education;
if female_head_education=0 then eduhh=male_head_education;
if eduhh=1 or eduhh=2 then eduhhlths=1; else eduhhlths=0;
if eduhh=3 then eduhhhs=1; else eduhhhs=0;
if eduhh=4 or eduhh=5 then eduhhu=1; else eduhhu=0;
if eduhh=6 then eduhhpc=1; else eduhhpc=0;
*race of the household;
```

```
if race=1 then White=1; else White=0;
if race=2 then Black=1; else Black=0;
if race=3 then Oriental=1; else Oriental=0;
if race=4 then Other=1; else Other=0;

*hispanic origin;
if hispanic_origin=1 then hisp_yes=1; else hisp_yes=0;
if hispanic_origin=2 then hisp_no=1; else hisp_no=0;

*region;
if region_code=1 then NewEngland=1; else NewEngland=0;
if region_code=2 then MiddleAtlantic=1; else  MiddleAtlantic=0;
if region_code=3 then EastNorthCentral=1; else EastNorthCentral=0;
if region_code=4 then WestNorthCentral=1; else WestNorthCentral=0;
if region_code=5 then SouthAtlantic=1; else SouthAtlantic=0;
if region_code=6 then EastSouthCentral=1; else EastSouthCentral=0;
if region_code=7 then WestSouthCentral=1; else WestSouthCentral=0;
if region_code=8 then Mountain=1; else Mountain=0;
if region_code=9 then Pacific=1; else Pacific=0;

*age and presence of children "ac";
if age_and_presence_of_children=1 then aclt6_only=1; else aclt6_only=0;
if age_and_presence_of_children=2 then ac6_12only=1; else ac6_12only=0;
if age_and_presence_of_children=3 then ac13_17only=1; else ac13_17only=0;
if age_and_presence_of_children=4 then aclt6_6_12only=1; else aclt6_6_12only=0;
if age_and_presence_of_children=5 then aclt6_13_17only=1; else aclt6_13_17only=0;
if age_and_presence_of_children=6 then ac6_12and13_17only=1; else ac6_12and13_17only=0;
if age_and_presence_of_children=7 then aclt6_6_12and13_17=1; else aclt6_6_12and13_17=0;
if age_and_presence_of_children=9 then no_child=1; else no_child=0;
*houdehold head's gender "fhonly=femal household head only" and "mhonly=male household
head only";
if male_head_age=0 then fhonly=1; else fhonly=0;
if female_head_age=0 then mhonly=1; else mhonly=0;
if female_head_age ne 0 and male_head_age ne 0 then fhmh=1; else fhmh=0;
price_energy = total_ed_exp/total_ed_oz;
loghinc=log(hinc);
*create a dummy variable for the quantity of soymilk;
if total_ed_oz>0 then Dummy_ed=1; else Dummy_ed=0;
*replace dots (.s) with a zero;

if total_ed_oz=.  then total_ed_oz=0;
if total_ed_exp=. then total_ed_exp=0;
if total_ed_coupon=. then total_ed_coupon=0;
*renaming quanitity purchased;
Q_ed=total_ed_oz;

keep
household_code
```

hinc
hsize

agehhlt25
agehh2529
agehh3034
agehh3544
agehh4554
agehh5564
agehhgt64

emphhpt
emphhft
emphhnfp

eduhhlths
eduhhhs
eduhhu
eduhhpc

white
black
oriental
other

hisp_yes
hisp_no

NewEngland
MiddleAtlantic
EastNorthCentral
WestNorthCentral
SouthAtlantic
EastSouthCentral
WestSouthCentral
Mountain
Pacific

aclt6_only
ac6_12only
ac13_17only
aclt6_6_12only
aclt6_13_17only
ac6_12and13_17only
aclt6_6_12and13_17
no_child

fhonly

```
mhonly
fhmh

price_energy
loghinc

Q_ed
Dummy_ed
total_ed_oz
total_ed_exp
total_ed_coupon

run;
proc means data=Yang.ed_first; title " mean";
run;*  mean;
data Yang.ed_income_cond;set yang.ed_first;
if index(total_ed_oz or total_ed_exp,"0")>0  then delete;
run;* conditional mean";

proc means data=Yang.ed_income_cond; title " conditional mean";
run;

*to get the auxiliary regression;
Proc reg data= Yang.ed_first;
model price_energy=loghinc hsize NewEngland MiddleAtlantic EastNorthCentral
WestNorthCentral SouthAtlantic EastSouthCentral WestSouthCentral Mountain;
RUN;*  this price in "first_first" replace the 0 with .;
data Yang.ed_second; set Yang.ed_first;
if (total_ed_oz ne 0 or total_ed_exp ne 0) then price_energy=total_ed_exp/total_ed_oz;
if (total_ed_oz = 0 or total_ed_exp = 0)
then price_energy=0.10870 + 0.00516*loghinc-0.00160*hsize+0.01148*NewEngland
+0.01848*MiddleAtlantic +0.00012149*EastNorthCentral-0.00007598*westNorthCentral
+0.00853*SouthAtlantic+0.00836*EastSouthCentral+0.00632*WestSouthCentral
+0.00327*Mountain;
RUN;
proc means data=Yang.ed_second;
var price_energy   Q_ed   Dummy_ed  hinc;title "unconditional mean of consumption";
run;

data Yang.imputed_price_test2; set Yang.ed_first;
if (total_ed_oz ne 0 or total_ed_exp ne 0) then price_energy=. ;
if (total_ed_oz = 0 or total_ed_exp = 0)
then price_energy=0.10870 + 0.00516*loghinc-0.00160*hsize+0.01148*NewEngland
+0.01848*MiddleAtlantic +0.00012149*EastNorthCentral-0.00007598*westNorthCentral
+0.00853*SouthAtlantic+0.00836*EastSouthCentral+0.00632*WestSouthCentral
+0.00327*Mountain;
RUN;
```

```
proc means data= Yang.imputed_price_test2;
var price_energy ; title "mean of imputed price";
run;

libname Yang 'Z:\Yang\Yang';
data Yang.tobit; merge Yang.ch_second Yang.ed_second Yang.sp_second;
by household_code;
run;

data Yang.tobit2_energy ; set Yang.tobit;
logprice_chocolate=log(price_chocolate);
logprice_energy=log(price_energy);
logprice_sportsdrink=log(price_sportsdrink);
loghinc=log(hinc);
run;

Proc QLIM data=Yang.tobit2_energy ;
	model Q_ed= logprice_energy  price_chocolate logprice_sportsdrink
						hsize
						loghinc agehh2529 agehh3034 agehh3544 agehh4554
agehh5564 agehhgt64
emphhpt emphhft eduhhhs eduhhu eduhhpc
						black oriental other hisp_yes newengland
middleatlantic eastnorthcentral
westnorthcentral southatlantic eastsouthcentral westsouthcentral
mountain
aclt6_only ac6_12only ac13_17only aclt6_6_12only aclt6_13_17only
ac6_12and13_17only aclt6_6_12and13_17
						fhonly mhonly;
	endogenous Q_ed ~ censored(lowerbound=0);
	nloptions maxiter=500; /*maximum number of iterations set at 300*/
	*hetero Q_ed ~ P_chocolate P_energy p_sp hinc agehh2529 agehh3034 agehh3544
agehh4554 agehh5564 agehhgt64 emphhpt emphhft eduhhhs eduhhu eduhhpc
			east midwest south black oriental other hisp_yes aclt6_only ac6_12only
ac13_17only aclt6_6_12only aclt6_13_17only
			ac6_12and13_17only aclt6_6_12and13_17 fhonly mhonly;
	output out=Yang.Tobit_out_energy conditional expected marginal xbeta;*conditional
expected value of Q_ed,
	unconditional expected value of Q_ed,
	unconditional marginal effect of all variables, xbeta;
run;

Proc means data=Yang.Tobit_out_energy;
var   Meff_logPrice_energy  Meff_Price_chocolate Meff_logPrice_sportsdrink;

run;
*************************************************************************
*****************
```

Following is the calculation of components needed to generate the McDonald and Moffitt Decomposition. SAS does not provide all information to calculate that. Therefore, we have to manually calculate some of measures. Also, if there is any non-linearity in the Tobit model, SAS does not use that information in calculating marginal effects. Therefore, we need to make adjustments for that too.
***;

```
data Yang.ed_marginal;set Yang.Tobit_out_energy;
* z is the normalized index,sas does not provide z directly,  z=xbeta_y/sigma;
z=xbeta_Q_ed/1531.565121;
*capfz is the cdf standard normal,sas does not provide capfz directly;
capfz=probnorm(z);*F(z);
*fz is the standard normal density function,sas does not provide capfz directly, * 2.5066272 is the result of (1/sqrt(2*pi));;
fz=exp(-z**2/2)/2.5066272;
*expected_Q_ed is the unconditional expected value of the dependent variable;
*expected_Q_ed also serve as the predicted value of the dependent variable;
*sas captures the unconditional expected values;
expected_Q_ed=xbeta_Q_ed*capfz+1531.565121*fz;
*cexpected_Q_ed is the conditional expected value of the dependent variable;
*sas captures the condtional expected values;
cexpected_Q_ed=xbeta_Q_ed+1531.565121*fz/capfz;

*unconditional marginal effects of price and household income;*sas captures the unconditional marginal effects of variables(logprice,loghinc);
meff_logprice_energy=(-1506.665989)*capfz;
meff_price_energy=(-1506.665989)*capfz/0.1309351;*0.1309351 is the unconditional mean of energy price;
meff_logprice_sportsdrink=(-179.832642)*capfz;*this is captured by sas automatically;
meff_price_sportsdrink=(-179.832642)*capfz/0.0527959;*0.05247959 is the unconditional mean of chocolate price;
meff_price_chocolate=(2460.871145)*capfz;*this is captured by sas automatically;

meff_loghinc=(8.986075)*capfz;
meff_hinc=(8.986075)*capfz/58.3209305;*58.3209305 is the unconditional mean of income;

Un_own_price_energy=((-1506.665989)*capfz)/31.8735011;* 31.8735011 is the uncondtional mean of Q_ed;
Un_cross_price_chocolate=((2460.871145)*capfz*0.0503423)/31.8735011;* 2460.871145 is the beta of logprice_energy, unconditional mean for Price_chocolate is 0.0503423;
Un_cross_price_sportsdrink=((-179.832642)*capfz)/31.8735011;* -179.832642 is the beta of logprice_sportsdrink;
Un_income_ed=((8.986075)*capfz)/31.8735011;*8.986075 is the beta of hinc;

* conditional marginal effects for price;
c_meff_logprice_energy=(-1506.665989)*(1-z*fz/capfz-(fz/capfz)*(fz/capfz));
```

c_meff_price_energy=(**-1506.665989**)*(**1**-z*fz/capfz-
(fz/capfz)*(fz/capfz))/**0.1294626**;*0.1294626 is the conditional mean of energy price;
c_meff_logprice_sportsdrink=(**-179.832642**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz));
c_meff_price_sportsdrink=(**-179.832642**)*(**1**-z*fz/capfz-
(fz/capfz)*(fz/capfz))/**0.0523869**;*0.0523869 is the conditional mean of sportsdrink price;
c_meff_price_chocolate= (**2460.871145**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz));
*0.0485468 is the conditional mean of chocolate price;

c_meff_loghinc=(**8.986075**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz));
c_meff_hinc=(**8.986075**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))/**59.8461624**; *59.8461624 is the conditional mean of hinc;
*conditional marginal effects;
c_Meff_hsize=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))***145.634098**;
c_meff_agehh2529=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-136.991553**);
c_meff_agehh3034=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-205.928381**);

c_meff_agehh3544=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-493.844408**);
c_meff_agehh4554=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-614.886354**);

c_meff_agehh5564=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-962.168620**);
c_meff_agehhgt64=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-1369.218589**);
c_meff_emphhpt=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-48.396489**);

c_meff_emphhft=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**73.052735**);
c_meff_eduhhhs=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-182.899801**);
c_meff_eduhhu=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-328.257701**);
c_meff_eduhhpc=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-576.505921**);

c_meff_black=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-29.923925**);
c_meff_Oriental=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-261.000309**);
c_meff_other=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**126.411478**);

c_meff_hisp_yes=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**105.142756**);
c_meff_newengland=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-451.728286**);
c_meff_middleatlantic=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))* (**-374.220025**);
c_meff_eastnorthcentral=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-401.715441**);

c_meff_westnorthcentral=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-349.840486**);
c_meff_southatlantic=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-337.746668**);
c_meff_eastsouthcentral=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-298.102209**);
c_meff_westsouthcentral=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))* (**-128.101268**);
c_meff_mountain=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-78.475689**);

c_meff_aclt6_only=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-198.692854**);
c_meff_ac6_12only=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-104.005961**);

c_meff_ac13_17only=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))* **265.723313**;

c_meff_aclt6_6_12only=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-460.371311**);
c_meff_aclt6_13_17only=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-173.197619**);
c_meff_ac6_12and13_17only=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-152.664320**);

c_meff_aclt6_6_12and13_17= (**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-321.226799**);
c_meff_fhonly=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**49.423494**);
c_meff_mhonly=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**291.318822**);

**now to calculate the marginal change in probability, we use the McDonald and Moffitt Decomposition
Reference:
McDonald, J.F., and R. A. Moffitt, "The Uses of Tobit Analysis" The Review of Economics and Statistics, May 1980, 62(2):318-321;

pr_meff_logprice_energy=(meff_logprice_energy-capfz*(c_meff_logprice_energy))/cexpected_Q_ed;
pr_meff_logprice_sportsdrink=(meff_logprice_sportsdrink-capfz*(c_meff_logprice_sportsdrink))/cexpected_Q_ed;
pr_meff_price_chocolate=(meff_price_chocolate-capfz*(c_meff_logprice_chocolate))/cexpected_Q_ed;
pr_Meff_hsize=(meff_hsize-capfz*(c_meff_hsize))/cexpected_Q_ed;
pr_meff_agehh2529=(meff_agehh2529-capfz*(c_meff_agehh2529))/cexpected_Q_ed;
pr_meff_agehh3034=(meff_agehh3034-capfz*(c_meff_agehh3034))/cexpected_Q_ed;
pr_meff_agehh3544=(meff_agehh3544-capfz*(c_meff_agehh3544))/cexpected_Q_ed;
pr_meff_agehh4554=(meff_agehh4554-capfz*(c_meff_agehh4554))/cexpected_Q_ed;
pr_meff_agehh5564=(meff_agehh5564-capfz*(c_meff_agehh5564))/cexpected_Q_ed;
pr_meff_agehhgt64=(meff_agehhgt64-capfz*(c_meff_agehhgt64))/cexpected_Q_ed;
pr_meff_emphhpt=(meff_emphhpt-capfz*(c_meff_emphhpt))/cexpected_Q_ed;
pr_meff_emphhft=(meff_emphhft-capfz*(c_meff_emphhft))/cexpected_Q_ed;
pr_meff_eduhhhs=(meff_eduhhhs-capfz*(c_meff_eduhhhs))/cexpected_Q_ed;
pr_meff_eduhhu=(meff_eduhhu-capfz*(c_meff_eduhhu))/cexpected_Q_ed;
pr_meff_eduhhpc=(meff_eduhhpc-capfz*(c_meff_eduhhpc))/cexpected_Q_ed;
pr_meff_black=(meff_black-capfz*(c_meff_black))/cexpected_Q_ed;
pr_meff_Oriental=(meff_Oriental-capfz*(c_meff_Oriental))/cexpected_Q_ed;
pr_meff_other=(meff_other-capfz*(c_meff_other))/cexpected_Q_ed;
pr_meff_hisp_yes=(meff_hisp_yes-capfz*(c_meff_hisp_yes))/cexpected_Q_ed;
pr_meff_newengland=(meff_newengland-capfz*(c_meff_newengland))/cexpected_Q_ed;
pr_meff_middleatlantic=(meff_middleatlantic-capfz*(c_meff_middleatlantic))/cexpected_Q_ed;
pr_meff_eastnorthcentral=(meff_eastnorthcentral-capfz*(c_meff_eastnorthcentral))/cexpected_Q_ed;
pr_meff_westnorthcentral=(meff_westnorthcentral-capfz*(c_meff_westnorthcentral))/cexpected_Q_ed;
pr_meff_southatlantic=(meff_southatlantic-capfz*(c_meff_southatlantic))/cexpected_Q_ed;
pr_meff_eastsouthcentral=(meff_eastsouthcentral-capfz*(c_meff_eastsouthcentral))/cexpected_Q_ed;
pr_meff_westsouthcentral=(meff_westsouthcentral-capfz*(c_meff_westsouthcentral))/cexpected_Q_ed;
pr_meff_mountain=(meff_mountain-capfz*(c_meff_mountain))/cexpected_Q_ed;

pr_meff_aclt6_only=(meff_aclt6_only-capfz*(c_meff_aclt6_only))/cexpected_Q_ed;
pr_meff_ac6_12only=(meff_ac6_12only-capfz*(c_meff_ac6_12only))/cexpected_Q_ed;
pr_meff_ac13_17only=(meff_ac13_17only-capfz*(c_meff_ac13_17only))/cexpected_Q_ed;
pr_meff_aclt6_6_12only=(meff_aclt6_6_12only-
capfz*(c_meff_aclt6_6_12only))/cexpected_Q_ed;
pr_meff_aclt6_13_17only=(meff_aclt6_13_17only-
capfz*(c_meff_aclt6_13_17only))/cexpected_Q_ed;
pr_meff_ac6_12and13_17only=(meff_ac6_12and13_17only-
capfz*(c_meff_ac6_12and13_17only))/cexpected_Q_ed;
pr_meff_aclt6_6_12and13_17=(meff_aclt6_6_12and13_17-
capfz*(c_meff_aclt6_6_12and13_17))/cexpected_Q_ed;
pr_meff_fhonly=(meff_fhonly-capfz*(c_meff_fhonly))/cexpected_Q_ed;
pr_meff_mhonly=(meff_mhonly-capfz*(c_meff_mhonly))/cexpected_Q_ed;

*conditional elasticities;
C_Own_Price_energy=((**-1506.665989**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz)))/**441.1158858**;
* 441.1158858 is the condtional mean of total _ed_oz;
C_cross_price_chocolate=(**2460.871145***(**1**-z*fz/capfz-
(fz/capfz)*(fz/capfz))***0.0485519**)/**441.1158858**;
* 2460.871145 is the beta of logprice_energy,,0.0485519 is the conditional mean for chocolate milk price   ;
C_cross_price_sportsdrink=((**-179.832642**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz)))/**441.1158858**;
C_income_ed=**8.986075***(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))/**441.1158858**;
*8.986075 is the beta of loghinc;
run;

proc means data=Yang.ed_marginal n mean median std min max;
var
Un_own_price_energy
Un_cross_price_chocolate
Un_cross_price_sportsdrink
Un_income_ed;
title "unconditional elasticity for energy drinks";
run;
proc means data=Yang.ed_marginal n mean median std min max;
var
meff_logprice_energy
meff_price_energy
meff_logprice_sportsdrink
meff_price_sportsdrink
meff_price_chocolate

Meff_hsize meff_agehh2529 meff_agehh3034 meff_agehh3544
meff_agehh4554 meff_agehh5564 meff_agehhgt64 meff_emphhpt meff_emphhft meff_eduhhhs meff_eduhhu
meff_eduhhpc meff_black meff_oriental meff_other meff_hisp_yes meff_newengland meff_middleatlantic

```
meff_eastnorthcentral meff_westnorthcentral meff_southatlantic meff_eastsouthcentral
meff_westsouthcentral
meff_mountain meff_aclt6_only meff_ac6_12only meff_ac13_17only meff_aclt6_6_12only
meff_aclt6_13_17only
meff_ac6_12and13_17only meff_aclt6_6_12and13_17  meff_fhonly meff_mhonly;
title ", unconditional marginal effects energy drinks";
run;
data Yang.ed_marginal2;set Yang.ed_marginal;
if total_ed_oz>0;
run;
 proc means data=Yang.ed_marginal2 n mean median std min max;
var
C_Own_Price_energy  C_cross_price_chocolate C_cross_price_sportsdrink C_income_ed;
title "conditional elasticity of energy";
run;
proc means data=Yang.ed_marginal2 n mean median std min max;
var
c_meff_logprice_energy
c_meff_price_energy
c_meff_logprice_sportsdrink
c_meff_price_sportsdrink
c_meff_price_chocolate
c_meff_hinc
C_Meff_hsize  C_meff_agehh2529 c_meff_agehh3034 c_meff_agehh3544
c_meff_agehh4554 c_meff_agehh5564 C_meff_agehhgt64 C_meff_emphhpt C_meff_emphhft
c_meff_eduhhhs c_meff_eduhhu
c_meff_eduhhpc c_meff_black c_meff_oriental c_meff_other c_meff_hisp_yes
c_meff_newengland c_meff_middleatlantic
c_meff_eastnorthcentral c_meff_westnorthcentral c_meff_southatlantic c_meff_eastsouthcentral
c_meff_westsouthcentral
c_meff_mountain c_meff_aclt6_only c_meff_ac6_12only c_meff_ac13_17only
c_meff_aclt6_6_12only c_meff_aclt6_13_17only
c_meff_ac6_12and13_17only c_meff_aclt6_6_12and13_17  c_meff_fhonly c_meff_mhonly;
run;
proc means data=Yang.ed_marginal2 n mean median std min max;
var
pr_meff_logprice_energy
pr_meff_logprice_sportsdrink
pr_meff_price_chocolate
pr_Meff_hsize
pr_meff_agehh2529
pr_meff_agehh3034
pr_meff_agehh3544
pr_meff_agehh4554
pr_meff_agehh5564
pr_meff_agehhgt64
pr_meff_emphhpt
pr_meff_emphhft
```

pr_meff_eduhhhs
pr_meff_eduhhu
pr_meff_eduhhpc
pr_meff_black
pr_meff_Oriental
pr_meff_other
pr_meff_hisp_yes
pr_meff_newengland
pr_meff_middleatlantic
pr_meff_eastnorthcentral
pr_meff_westnorthcentral
pr_meff_southatlantic
pr_meff_eastsouthcentral
pr_meff_westsouthcentral
pr_meff_mountain
pr_meff_aclt6_only
pr_meff_ac6_12only
pr_meff_ac13_17only
pr_meff_aclt6_6_12only
pr_meff_aclt6_13_17only
pr_meff_ac6_12and13_17only
pr_meff_aclt6_6_12and13_17
pr_meff_fhonly
pr_meff_mhonly;
title "mean of the marginal change in probability-ed";
run;

2.3 Sports drinks:

libname Yang 'k:\Yang\Yang';
*this part, just check the mean, and calculate the conditional mean;
proc means data= Yang.sports_drink;
run;
data Yang.sp_dot;set Yang.sports_drink;
price_sportsdrink = total_sp_exp/total_sp_oz;
if household_income=**03** then household_income=**5**;
if household_income=**04** then household_income=**6.5**;
if household_income=**06** then household_income=**9**;
if household_income=**08** then household_income=**11**;
if household_income=**10** then household_income=**13.5**;
if household_income=**11** then household_income=**17.5**;
if household_income=**13** then household_income=**22.5**;
if household_income=**15** then household_income=**27.5**;
if household_income=**16** then household_income=**32.5**;
if household_income=**17** then household_income=**37.5**;
if household_income=**18** then household_income=**42.5**;
if household_income=**19** then household_income=**47.5**;
if household_income=**21** then household_income=**55**;
if household_income=**23** then household_income=**65**;

```
if household_income=26 then  household_income=85;
if household_income=27 then  household_income=112.5;
run;

data Yang.sss;set Yang.sp_dot;
if price_sportsdrink ne 0 then delete;
run;* to check household's sports price is 0;
data Yang.sp_dot1;set Yang.sp_dot;
if index(household_code,"2099025")>0 then delete;
if index(household_code,"8073408")>0 then delete;
if index(household_code,"8210882")>0 then delete;
if index(household_code,"8236525")>0 then delete;
if index(household_code,"8315819")>0 then delete;
if index(household_code,"8515976")>0 then delete;
if index(household_code,"8657627")>0 then delete;
if index(household_code,"8674944")>0 then delete;
if index(household_code,"30034887")>0 then delete;
if index(household_code,"30059156")>0 then delete;
if index(household_code,"30562133")>0 then delete;
if index(household_code,"30732159")>0 then delete;
if index(household_code,"30745170")>0 then delete;
if index(household_code,"8644217")>0 then delete;
if index(household_code,"30643618")>0 then delete;
if index(household_code,"8593529")>0 then delete;

run;* DELETE THE HOUSEHOLD WHO CONSUME CHOCOLATE BUT EXPENDITURE
IS 0;


proc means data=Yang.sp_dot1;
var household_code price_sportsdrink  total_sp_oz  ;title "conditional mean ";
run;
data yang.sp_dot2; set yang.sp_dot1;
if price_sportsdrink < 0.1009913+5*1.1581719 then delete;
run;
*this part, I want to get the imputed price, and the final dataset which will be used in tobit model;
data Yang.sp_delete;set Yang.sports_drink;
if index(household_code,"2099025")>0 then delete;
if index(household_code,"8073408")>0 then delete;
if index(household_code,"8210882")>0 then delete;
if index(household_code,"8236525")>0 then delete;
if index(household_code,"8315819")>0 then delete;
if index(household_code,"8515976")>0 then delete;
if index(household_code,"8657627")>0 then delete;
if index(household_code,"8674944")>0 then delete;
if index(household_code,"30034887")>0 then delete;
if index(household_code,"30059156")>0 then delete;
if index(household_code,"30562133")>0 then delete;
```

```
if index(household_code,"30732159")>0 then delete;
if index(household_code,"30745170")>0 then delete;
if index(household_code,"8644217")>0 then delete;
if index(household_code,"30643618")>0 then delete;
if index(household_code,"8593529")>0 then delete;

if index(household_code,"2001406")>0 then delete;
if index(household_code,"2001670")>0 then delete;
if index(household_code,"2001777")>0 then delete;
if index(household_code,"2018988")>0 then delete;
if index(household_code,"2029790")>0 then delete;
if index(household_code,"2049206")>0 then delete;
if index(household_code,"2050892")>0 then delete;
if index(household_code,"2057150")>0 then delete;
if index(household_code,"2066676")>0 then delete;
if index(household_code,"2083426")>0 then delete;
if index(household_code,"2097247")>0 then delete;
if index(household_code,"2098235")>0 then delete;
if index(household_code,"8005801")>0 then delete;
if index(household_code,"8030882")>0 then delete;
if index(household_code,"8045762")>0 then delete;
if index(household_code,"8046245")>0 then delete;
if index(household_code,"8074841")>0 then delete;
if index(household_code,"8096931")>0 then delete;
if index(household_code,"8123724")>0 then delete;
if index(household_code,"8162112")>0 then delete;
if index(household_code,"8176181")>0 then delete;
if index(household_code,"8191323")>0 then delete;
if index(household_code,"8193635")>0 then delete;
if index(household_code,"8203954")>0 then delete;
if index(household_code,"8231965")>0 then delete;
if index(household_code,"8245003")>0 then delete;
if index(household_code,"8251995")>0 then delete;
if index(household_code,"8253652")>0 then delete;
if index(household_code,"8270506")>0 then delete;
if index(household_code,"8277284")>0 then delete;
if index(household_code,"8314713")>0 then delete;

if index(household_code,"8327678")>0 then delete;
if index(household_code,"8331605")>0 then delete;
if index(household_code,"8343180")>0 then delete;
if index(household_code,"8350288")>0 then delete;
if index(household_code,"8370169")>0 then delete;

if index(household_code,"8420574")>0 then delete;
if index(household_code,"8529905")>0 then delete;
if index(household_code,"8579702")>0 then delete;
if index(household_code,"8599531")>0 then delete;
```

```
if index(household_code,"8607537")>0 then delete;
if index(household_code,"8633094")>0 then delete;
if index(household_code,"8634622")>0 then delete;
if index(household_code,"8644988")>0 then delete;
if index(household_code,"9005129")>0 then delete;
if index(household_code,"9150684")>0 then delete;
if index(household_code,"9170619")>0 then delete;
run;

data Yang.sp_first;set Yang.sp_delete;
if household_size=1 then hsize=1;
if household_size=2 then hsize=2;
if household_size=3 then hsize=3;
if household_size=4 then hsize=4;
if household_size=5 then hsize=5;
if household_size=6 then hsize=6;
if household_size=7 then hsize=7;
if household_size=8 then hsize=8;
if household_size=9 then hsize=9;

if household_income=03 then hinc=5;
if household_income=04 then hinc=6.5;
if household_income=06 then hinc=9;
if household_income=08 then hinc=11;
if household_income=10 then hinc=13.5;
if household_income=11 then hinc=17.5;
if household_income=13 then hinc=22.5;
if household_income=15 then hinc=27.5;
if household_income=16 then hinc=32.5;
if household_income=17 then hinc=37.5;
if household_income=18 then hinc=42.5;
if household_income=19 then hinc=47.5;
if household_income=21 then hinc=55;
if household_income=23 then hinc=65;
if household_income=26 then hinc=85;
if household_income=27 then hinc=112.5;
*age of household head "agehh",agef is agefemale;
agehh=female_head_age;
if female_head_age=0 then agehh=male_head_age;
if agehh=1 then agehhlt25=1; else agehhlt25=0;
if agehh=2 then agehh2529=1; else agehh2529=0;
if agehh=3 then agehh3034=1; else agehh3034=0;
if agehh=4 or agehh=5 then agehh3544=1; else agehh3544=0;
if agehh=6 or agehh=7 then agehh4554=1; else agehh4554=0;
if agehh=8 then agehh5564=1; else agehh5564=0;
if agehh=9 then agehhgt64=1; else agehhgt64=0;
*employment status of the household head "emphh";
emphh=female_head_employment;
```

```
if female_head_employment=0 then emphh=male_head_employment;
if emphh=1 or emphh=2 then emphhpt=1; else emphhpt=0;
if emphh=3 then emphhft=1; else emphhft=0;
if emphh=9 then emphhnfp=1; else emphhnfp=0;
*education of the household head "eduhh";
eduhh=female_head_education;
if female_head_education=0 then eduhh=male_head_education;
if eduhh=1 or eduhh=2 then eduhhlths=1; else eduhhlths=0;
if eduhh=3 then eduhhhs=1; else eduhhhs=0;
if eduhh=4 or eduhh=5 then eduhhu=1; else eduhhu=0;
if eduhh=6 then eduhhpc=1; else eduhhpc=0;
*race of the household;
if race=1 then White=1; else White=0;
if race=2 then Black=1; else Black=0;
if race=3 then Oriental=1; else Oriental=0;
if race=4 then Other=1; else Other=0;

*hispanic origin;
if hispanic_origin=1 then hisp_yes=1; else hisp_yes=0;
if hispanic_origin=2 then hisp_no=1; else hisp_no=0;

*region;
if region_code=1 then NewEngland=1; else NewEngland=0;
if region_code=2 then MiddleAtlantic=1; else  MiddleAtlantic=0;
if region_code=3 then EastNorthCentral=1; else EastNorthCentral=0;
if region_code=4 then WestNorthCentral=1; else WestNorthCentral=0;
if region_code=5 then SouthAtlantic=1; else SouthAtlantic=0;
if region_code=6 then EastSouthCentral=1; else EastSouthCentral=0;
if region_code=7 then WestSouthCentral=1; else WestSouthCentral=0;
if region_code=8 then Mountain=1; else Mountain=0;
if region_code=9 then Pacific=1; else Pacific=0;

*age and presence of children "ac";
if age_and_presence_of_children=1 then aclt6_only=1; else aclt6_only=0;
if age_and_presence_of_children=2 then ac6_12only=1; else ac6_12only=0;
if age_and_presence_of_children=3 then ac13_17only=1; else ac13_17only=0;
if age_and_presence_of_children=4 then aclt6_6_12only=1; else aclt6_6_12only=0;
if age_and_presence_of_children=5 then aclt6_13_17only=1; else aclt6_13_17only=0;
if age_and_presence_of_children=6 then ac6_12and13_17only=1; else ac6_12and13_17only=0;
if age_and_presence_of_children=7 then aclt6_6_12and13_17=1; else aclt6_6_12and13_17=0;
if age_and_presence_of_children=9 then no_child=1; else no_child=0;
*houdehold head's gender "fhonly=femal household head only" and "mhonly=male household
head only";
if male_head_age=0 then fhonly=1; else fhonly=0;
if female_head_age=0 then mhonly=1; else mhonly=0;
if female_head_age ne 0 and male_head_age ne 0 then fhmh=1; else fhmh=0;

price_sportsdrink = total_sp_exp/total_sp_oz;
```

```
loghinc=log(hinc);
*create a dummy variable for the quantity of sp;
if total_sp_oz>0 then Dummy_sp=1; else Dummy_sp=0;
*replace dots (.s) with a zero;

if total_sp_oz=.  then total_sp_oz=0;
if total_sp_exp=. then total_sp_exp=0;
if total_sp_coupon=. then total_sp_coupon=0;
*renaming quanitity purchased;
Q_sp=total_sp_oz;

keep
household_code
hinc
hsize

agehhlt25
agehh2529
agehh3034
agehh3544
agehh4554
agehh5564
agehhgt64

emphhpt
emphhft
emphhnfp

eduhhlths
eduhhhs
eduhhu
eduhhpc

white
black
oriental
other

hisp_yes
hisp_no

NewEngland
MiddleAtlantic
EastNorthCentral
WestNorthCentral
SouthAtlantic
EastSouthCentral
WestSouthCentral
```

```
Mountain
Pacific

aclt6_only
ac6_12only
ac13_17only
aclt6_6_12only
aclt6_13_17only
ac6_12and13_17only
aclt6_6_12and13_17
no_child

fhonly
mhonly
fhmh

price_sportsdrink
loghinc

Q_sp
Dummy_sp
total_sp_oz
total_sp_exp
total_sp_coupon

run;
proc means data=Yang.sp_first; title " mean";
run;
data yang.income_sp;set yang.sp_first;
if index(total_sp_oz or total_sp_exp,"0")>0 then delete;
run;* to get the conditional mean of income;
proc means data=yang.income_sp;
run;* to get the conditional mean of income;
* to get the auxiliary regression;
Proc reg data= Yang.sp_first;
model price_sportsdrink=hinc hsize NewEngland MiddleAtlantic EastNorthCentral
WestNorthCentral SouthAtlantic EastSouthCentral WestSouthCentral Mountain;
RUN;*  this price in "first_first" replace the 0 with .;
data Yang.sp_second; set Yang.sp_first;
if (total_sp_oz ne 0 or total_sp_exp ne 0) then price_sportsdrink=total_sp_exp/total_sp_oz;
if (total_sp_oz = 0 or total_sp_exp = 0)
then price_sportsdrink=0.05678 --0.00001276*hinc -0.00095727*hsize+0.00432*NewEngland
-0.00465*MiddleAtlantic -0.00218*EastNorthCentral-0.00166*westNorthCentral
-0.00272*SouthAtlantic+0.00493*EastSouthCentral+0.00333*WestSouthCentral
-0.00177*Mountain;
RUN;
proc means data=Yang.sp_second;
var price_sportsdrink   Q_sp   Dummy_sp  hinc;title "unconditional mean of consumption";
```

```
run;

*to test wheather the imputed prices are fit our market price;
data Yang.imputed_price_sp; set Yang.sp_first;
if (total_sp_oz ne 0 or total_sp_exp ne 0) then price_sportsdrink=. ;
if (total_sp_oz = 0 or total_sp_exp = 0)
then price_sportsdrink=0.05678 -0.00001276*hinc -0.00095727*hsize+0.00432*NewEngland
-0.00465*MiddleAtlantic -0.00218*EastNorthCentral-0.00166*westNorthCentral
-0.00272*SouthAtlantic+0.00493*EastSouthCentral+0.00333*WestSouthCentral
-0.00177*Mountain;
RUN;
proc means data= Yang.imputed_price_sp;
var price_sportsdrink; title "mean of imputed price";
run;
libname Yang 'k:\Yang\Yang';
data Yang.tobit; merge Yang.ch_second Yang.ed_second Yang.sp_second;
by household_code;
run;
data Yang.tobit_sp ; set Yang.tobit;
logprice_chocolate=log(price_chocolate);
logprice_energy=log(price_energy);
logprice_sportsdrink=log(price_sportsdrink);
loghinc=log(hinc);
run;

/*Following is the tobit model for chocolate*/
Proc QLIM data=Yang.tobit_sp ;
	model Q_sp= logprice_chocolate logprice_energy logprice_sportsdrink
					hsize
					loghinc agehh2529 agehh3034 agehh3544 agehh4554
agehh5564 agehhgt64
emphhpt emphhft eduhhhs eduhhu eduhhpc
					black oriental other hisp_yes newengland
middleatlantic eastnorthcentral
westnorthcentral southatlantic eastsouthcentral westsouthcentral
mountain
aclt6_only ac6_12only ac13_17only aclt6_6_12only aclt6_13_17only
ac6_12and13_17only aclt6_6_12and13_17
					fhonly mhonly;
	endogenous Q_sp ~ censored(lowerbound=0);
	nloptions maxiter=500; /*maximum number of iterations set at 300*/
	*hetero Q_sp ~ P_chocolate P_energy p_sp hinc agehh2529 agehh3034 agehh3544
agehh4554 agehh5564 agehhgt64 emphhpt emphhft eduhhhs eduhhu eduhhpc
			east midwest south black oriental other hisp_yes aclt6_only ac6_12only
ac13_17only aclt6_6_12only aclt6_13_17only
			ac6_12and13_17only aclt6_6_12and13_17 fhonly mhonly;
	output out=Yang.Tobit_out_sp conditional expected marginal xbeta;*conditional
expected value of Q_sp,
```

unconditional expected value of Q_sp,
unconditional marginal effect of all variables, xbeta;
run;

Proc means data=Yang.Tobit_out_sp;
var Meff_logPrice_chocolate Meff_logPrice_energy
Meff_logPrice_sportsdrink;*unconditional marginal effect of logprice_chocolate and logprice_energy;
run; *meff represents the unconditional marginal effect;
****
Following is the calculation of components needed to generate the McDonald and Moffitt Decomposition. SAS does not provide all information to calculate that. Therefore, we have to manually calculate some of measures. Also, if there is any non-linearity in the Tobit model, SAS does not use that information in calculating marginal effects. Therefore, we need to make adjustments for that too.
***;

data Yang.sp_marginal;set Yang.Tobit_out_sp;
* z is the normalized index,sas does not provide z directly,z=xbeta_y/sigma;
z=xbeta_Q_sp/ **1541.710879**;
*capfz is the cdf standard normal,sas does not provide capfz directly;
capfz=probnorm(z);*F(z);
*fz is the standard normal density function,sas does not provide capfz directly, * 2.5066272 is the result of (1/sqrt(2*pi));;
fz=exp(-z****2**/**2**)/**2.5066272**;
*expected_Q_sp is the unconditional expected value of the dependent variable;
*expected_Q_sp also serve as the predicted value of the dependent variable;
*sas captures the unconditional expected values;
expected_Q_sp=xbeta_Q_sp*capfz+ **1541.710879***fz;
*cexpected_Q_sp is the conditional expected value of the dependent variable;
*sas captures the condtional expected values;
cexpected_Q_sp=xbeta_Q_sp+ **1541.710879***fz/capfz;

*unconditional marginal effects of price and household income;*sas captures the unconditional marginal effects of variables(logprice,loghinc);
meff_logprice_chocolate=(**-78.120369**)*capfz;*this is captured by sas automatically;
meff_price_chocolate=(**-78.120369**)*capfz/**0.0503423**;*0.05034123 is the unconditional mean of chocolate price;
meff_logprice_energy=(**-319.242230**)*capfz;
meff_price_energy=(**-319.242230**)*capfz/**0.1309351**;*0.1309351 is the unconditional mean of energy price;
meff_logprice_sportsdrink=(**-1639.646361**)*capfz;*this is captured by sas automatically;
meff_price_sportsdrink=(**-1639.646361**)*capfz/ **0.0537267**;* 0.0537267is the unconditional mean of sports drinks
meff_loghinc=(38.038082)*capfz;
meff_hinc=(**38.038082**)*capfz/**58.3209305**;*58.3209305 is the unconditional mean of income;

*unconditional Elasticities for sports drink;
Un_own_price_sportsdrink=(**-1639.646361**)*(capfz)/**270.7298976**; * 270.7298976 is the uncondtional mean of Q_sp;
Un_cross_price_energy=(**-319.242230**)*(capfz)/**270.7298976**;
Un_cross_price_chocolate=(**-78.120369**)*(capfz)/**270.7298976**;
Un_income_sp=((**38.038082**)*capfz)/**270.7298976**; * 38.038082is the beta of hinc;

* conditional marginal effects for price;
c_meff_logprice_chocolate= (**-78.120369**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz));
c_meff_price_chocolate=(**-78.120369**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))/**0.0485519**;*0.0485468 is the conditional mean of chocolate price;
c_meff_logprice_energy=(**-319.242230**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz));
c_meff_price_energy=(**-319.242230**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))/**0.1294626**;*0.1294626 is the conditional mean of energy price;
c_meff_logprice_sportsdrink=(**-1639.646361**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz));
c_meff_price_sportsdrink=(**-1639.646361**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))/ **0.0523869**;* 0.0523869 is the conditional mean of sportsdrink price;

c_meff_loghinc=(**38.038082**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz));
c_meff_hinc=(**38.038082**)*(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))/**61.3868090**; *61.3868090 is the conditional mean of hinc--sp;

*conditional marginal effects;
c_Meff_hsize=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))* **163.536444**;
c_meff_agehh2529=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-83.884740**);
c_meff_agehh3034=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-121.917448**);

c_meff_agehh3544=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**- 163.604880**);
c_meff_agehh4554=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-238.373453**);

c_meff_agehh5564=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-509.206683**);
c_meff_agehhgt64=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-720.419668**);
c_meff_emphhpt=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-15.139101**);

c_meff_emphhft=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**37.612471**);
c_meff_eduhhhs=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**68.045329**);
c_meff_eduhhu=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-6.097570**);
c_meff_eduhhpc=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-145.523864**);

c_meff_black=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-17.813593**);
c_meff_Oriental=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**-161.998185**);
c_meff_other=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**82.721471**);

c_meff_hisp_yes=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**19.670463**);
c_meff_newengland=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**76.861736**);
c_meff_middleatlantic=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))* **36.204422**;
c_meff_eastnorthcentral=(**1**-z*fz/capfz-(fz/capfz)*(fz/capfz))*(**5.093450**);

```
c_meff_westnorthcentral=(1-z*fz/capfz-(fz/capfz)*(fz/capfz))*( -24.359815);
c_meff_southatlantic=(1-z*fz/capfz-(fz/capfz)*(fz/capfz))*(  165.198508);
c_meff_eastsouthcentral=(1-z*fz/capfz-(fz/capfz)*(fz/capfz))*( 296.758590);
c_meff_westsouthcentral=(1-z*fz/capfz-(fz/capfz)*(fz/capfz))*  252.844360;
c_meff_mountain=(1-z*fz/capfz-(fz/capfz)*(fz/capfz))*(  115.997787);

c_meff_aclt6_only=(1-z*fz/capfz-(fz/capfz)*(fz/capfz))*( -164.755368);
c_meff_ac6_12only=(1-z*fz/capfz-(fz/capfz)*(fz/capfz))*( 95.665944);

c_meff_ac13_17only=(1-z*fz/capfz-(fz/capfz)*(fz/capfz))* 480.682390;

c_meff_aclt6_6_12only=(1-z*fz/capfz-(fz/capfz)*(fz/capfz))*(  -320.620339);
c_meff_aclt6_13_17only=(1-z*fz/capfz-(fz/capfz)*(fz/capfz))*( 21.360341);
c_meff_ac6_12and13_17only=(1-z*fz/capfz-(fz/capfz)*(fz/capfz))*( 293.436511);

c_meff_aclt6_6_12and13_17= (1-z*fz/capfz-(fz/capfz)*(fz/capfz))*( -13.478918);
c_meff_fhonly=(1-z*fz/capfz-(fz/capfz)*(fz/capfz))*( -233.505810);
c_meff_mhonly=(1-z*fz/capfz-(fz/capfz)*(fz/capfz))*(  -54.664418);

**now to calculate the marginal change in probability, we use the McDonald and Moffitt
Decomposition
Reference:
McDonald, J.F., and R. A. Moffitt, "The Uses of Tobit Analysis" The Review of Economics and
Statistics, May 1980, 62(2):318-321;

*probability of being above the limit for chocolate,formula 17;
pr_meff_logprice_chocolate=(meff_logprice_chocolate-
capfz*(c_meff_logprice_chocolate))/cexpected_Q_sp;
pr_meff_logprice_energy=(meff_logprice_energy-
capfz*(c_meff_logprice_energy))/cexpected_Q_sp;
pr_meff_logprice_sportsdrink=(meff_logprice_sportsdrink-
capfz*(c_meff_logprice_sportsdrink))/cexpected_Q_sp;

pr_Meff_hsize=(meff_hsize-capfz*(c_meff_hsize))/cexpected_Q_sp;
pr_meff_agehh2529=(meff_agehh2529-capfz*(c_meff_agehh2529))/cexpected_Q_sp;
pr_meff_agehh3034=(meff_agehh3034-capfz*(c_meff_agehh3034))/cexpected_Q_sp;
pr_meff_agehh3544=(meff_agehh3544-capfz*(c_meff_agehh3544))/cexpected_Q_sp;
pr_meff_agehh4554=(meff_agehh4554-capfz*(c_meff_agehh4554))/cexpected_Q_sp;
pr_meff_agehh5564=(meff_agehh5564-capfz*(c_meff_agehh5564))/cexpected_Q_sp;
pr_meff_agehhgt64=(meff_agehhgt64-capfz*(c_meff_agehhgt64))/cexpected_Q_sp;
pr_meff_emphhpt=(meff_emphhpt-capfz*(c_meff_emphhpt))/cexpected_Q_sp;
pr_meff_emphhft=(meff_emphhft-capfz*(c_meff_emphhft))/cexpected_Q_sp;
pr_meff_eduhhhs=(meff_eduhhhs-capfz*(c_meff_eduhhhs))/cexpected_Q_sp;
pr_meff_eduhhu=(meff_eduhhu-capfz*(c_meff_eduhhu))/cexpected_Q_sp;
pr_meff_eduhhpc=(meff_eduhhpc-capfz*(c_meff_eduhhpc))/cexpected_Q_sp;
pr_meff_black=(meff_black-capfz*(c_meff_black))/cexpected_Q_sp;
pr_meff_Oriental=(meff_Oriental-capfz*(c_meff_Oriental))/cexpected_Q_sp;
```

```
pr_meff_other=(meff_other-capfz*(c_meff_other))/cexpected_Q_sp;
pr_meff_hisp_yes=(meff_hisp_yes-capfz*(c_meff_hisp_yes))/cexpected_Q_sp;
pr_meff_newengland=(meff_newengland-capfz*(c_meff_newengland))/cexpected_Q_sp;
pr_meff_middleatlantic=(meff_middleatlantic-capfz*(c_meff_middleatlantic))/cexpected_Q_sp;
pr_meff_eastnorthcentral=(meff_eastnorthcentral-
capfz*(c_meff_eastnorthcentral))/cexpected_Q_sp;
pr_meff_westnorthcentral=(meff_westnorthcentral-
capfz*(c_meff_westnorthcentral))/cexpected_Q_sp;
pr_meff_southatlantic=(meff_southatlantic-capfz*(c_meff_southatlantic))/cexpected_Q_sp;
pr_meff_eastsouthcentral=(meff_eastsouthcentral-
capfz*(c_meff_eastsouthcentral))/cexpected_Q_sp;
pr_meff_westsouthcentral=(meff_westsouthcentral-
capfz*(c_meff_westsouthcentral))/cexpected_Q_sp;
pr_meff_mountain=(meff_mountain-capfz*(c_meff_mountain))/cexpected_Q_sp;
pr_meff_aclt6_only=(meff_aclt6_only-capfz*(c_meff_aclt6_only))/cexpected_Q_sp;
pr_meff_ac6_12only=(meff_ac6_12only-capfz*(c_meff_ac6_12only))/cexpected_Q_sp;
pr_meff_ac13_17only=(meff_ac13_17only-capfz*(c_meff_ac13_17only))/cexpected_Q_sp;
pr_meff_aclt6_6_12only=(meff_aclt6_6_12only-
capfz*(c_meff_aclt6_6_12only))/cexpected_Q_sp;
pr_meff_aclt6_13_17only=(meff_aclt6_13_17only-
capfz*(c_meff_aclt6_13_17only))/cexpected_Q_sp;
pr_meff_ac6_12and13_17only=(meff_ac6_12and13_17only-
capfz*(c_meff_ac6_12and13_17only))/cexpected_Q_sp;
pr_meff_aclt6_6_12and13_17=(meff_aclt6_6_12and13_17-
capfz*(c_meff_aclt6_6_12and13_17))/cexpected_Q_sp;
pr_meff_fhonly=(meff_fhonly-capfz*(c_meff_fhonly))/cexpected_Q_sp;
pr_meff_mhonly=(meff_mhonly-capfz*(c_meff_mhonly))/cexpected_Q_sp;

*conditional elasticities;
C_Own_Price_sportsdrink=((-1639.646361)*(1-z*fz/capfz-(fz/capfz)*(fz/capfz)))/756.5483992;
* 756.5483992 is the  condtional mean of total _sp-_oz;
C_cross_price_energy=((-319.242230)*(1-z*fz/capfz-(fz/capfz)*(fz/capfz)))/756.5483992;
*  is the beta of logprice_energy;
C_cross_price_chocolate=(-78.120369*(1-z*fz/capfz-(fz/capfz)*(fz/capfz)))/756.5483992;
C_income_sp=(38.038082*(1-z*fz/capfz-(fz/capfz)*(fz/capfz)))/756.5483992;
*38.038082is the beta of hinc;

run;


proc means data=Yang.sp_marginal n mean median std min max;
var
Un_own_price_sportsdrink Un_cross_price_energy  Un_cross_price_chocolate  Un_income_sp
meff_logprice_chocolate
meff_price_chocolate
meff_logprice_energy
meff_price_energy
meff_logprice_sportsdrink
```

```
meff_price_sportsdrink

Meff_hsize  meff_agehh2529 meff_agehh3034 meff_agehh3544
meff_agehh4554 meff_agehh5564 meff_agehhgt64 meff_emphhpt meff_emphhft meff_eduhhhs
meff_eduhhu
meff_eduhhpc meff_black meff_oriental meff_other meff_hisp_yes meff_newengland
meff_middleatlantic
meff_eastnorthcentral meff_westnorthcentral meff_southatlantic meff_eastsouthcentral
meff_westsouthcentral
meff_mountain meff_aclt6_only meff_ac6_12only meff_ac13_17only meff_aclt6_6_12only
meff_aclt6_13_17only
meff_ac6_12and13_17only meff_aclt6_6_12and13_17  meff_fhonly meff_mhonly;
title "unconditional elasticity sp, unconditional marginal effects sp";
run;
data Yang.sp_marginal2 ;set Yang.sp_marginal;
if total_sp_oz>0;
run;*select all the household who buy ch,in this part we ignore whether the hh buy energy,
sport ;
proc means data=Yang.sp_marginal2 n mean median std min max;
var
C_Own_Price_sportsdrink  C_cross_price_energy C_cross_price_chocolate C_income_sp

c_meff_logprice_chocolate
c_meff_price_chocolate
c_meff_logprice_energy
c_meff_price_energy
c_meff_logprice_sportsdrink
c_meff_price_sportsdrink

c_meff_hinc

C_Meff_hsize  C_meff_agehh2529 c_meff_agehh3034 c_meff_agehh3544
c_meff_agehh4554 c_meff_agehh5564 C_meff_agehhgt64 C_meff_emphhpt C_meff_emphhft
c_meff_eduhhhs c_meff_eduhhu
c_meff_eduhhpc c_meff_black c_meff_oriental c_meff_other c_meff_hisp_yes
c_meff_newengland c_meff_middleatlantic
c_meff_eastnorthcentral c_meff_westnorthcentral c_meff_southatlantic c_meff_eastsouthcentral
c_meff_westsouthcentral
c_meff_mountain c_meff_aclt6_only c_meff_ac6_12only c_meff_ac13_17only
c_meff_aclt6_6_12only c_meff_aclt6_13_17only
c_meff_ac6_12and13_17only c_meff_aclt6_6_12and13_17  c_meff_fhonly c_meff_mhonly;
title "conditional marginal effects sp";
run;
proc means data=Yang.sp_marginal2 n mean median std min max;
var
pr_meff_logprice_chocolate
pr_meff_logprice_energy
pr_meff_logprice_sportsdrink
```

```
pr_Meff_hsize
pr_meff_agehh2529
pr_meff_agehh3034
pr_meff_agehh3544
pr_meff_agehh4554
pr_meff_agehh5564
pr_meff_agehhgt64
pr_meff_emphhpt
pr_meff_emphhft
pr_meff_eduhhhs
pr_meff_eduhhu
pr_meff_eduhhpc
pr_meff_black
pr_meff_Oriental
pr_meff_other
pr_meff_hisp_yes
pr_meff_newengland
pr_meff_middleatlantic
pr_meff_eastnorthcentral
pr_meff_westnorthcentral
pr_meff_southatlantic
pr_meff_eastsouthcentral
pr_meff_westsouthcentral
pr_meff_mountain
pr_meff_aclt6_only
pr_meff_ac6_12only
pr_meff_ac13_17only
pr_meff_aclt6_6_12only
pr_meff_aclt6_13_17only
pr_meff_ac6_12and13_17only
pr_meff_aclt6_6_12and13_17
pr_meff_fhonly
pr_meff_mhonly;
title "mean of the marginal change in probability-sp";
    run;
```

3. SAS Codes for Correlation Test：

```
libname Yang 'z:\Yang\Yang';
proc corr data=yang.tobit;
var price_chocolate price_energy price_sportsdrink;
run;
```

4. SAS Codes for the Energy Drinks Tobit Model in Different Functions：

4.1 Tobit Model of Energy Drinks in Quadratic Funtions:

```
libname Yang 'Z:\Yang\Yang';
data Yang.tobit; merge Yang.ch_second Yang.ed_second Yang.sp_second;
by household_code;
run;

data Yang.tobit2_energy ; set Yang.tobit;
logprice_chocolate=log(price_chocolate);
logprice_energy=log(price_energy);
logprice_sportsdrink=log(price_sportsdrink);
loghinc=log(hinc);
price_ch2=price_chocolate*price_chocolate;
price_en2=price_energy*price_energy;
price_sp2=price_sportsdrink*price_sportsdrink;
run;

Proc QLIM data=Yang.tobit2_energy ;
        model Q_ed=  price_energy price_chocolate  price_sportsdrink price_ch2 price_en2
price_sp2 hsize hinc agehh2529 agehh3034 agehh3544 agehh4554 agehh5564 agehhgt64
emphhpt emphhft eduhhhs eduhhu eduhhpc
black oriental other hisp_yes newengland middleatlantic eastnorthcentral westnorthcentral
southatlantic eastsouthcentral westsouthcentral mountain aclt6_only ac6_12only ac13_17only
aclt6_6_12only aclt6_13_17only  ac6_12and13_17only aclt6_6_12and13_17
fhonly mhonly;
        endogenous Q_ed ~ censored(lowerbound=0);
        nloptions maxiter=500;
        output out=Yang.Tobit_out_energy conditional expected marginal xbeta;
run;
```

4.2 Tobit Model of Energy Drinks in Semi-Log Funtions:

4.2.1 log price, linear income:

```
libname Yang 'Z:\Yang\Yang';
data Yang.tobit; merge Yang.ch_second Yang.ed_second Yang.sp_second;
by household_code;
run;

data Yang.tobit2_energy ; set Yang.tobit;
logprice_chocolate=log(price_chocolate);
logprice_energy=log(price_energy);
logprice_sportsdrink=log(price_sportsdrink);
loghinc=log(hinc);
run;

Proc QLIM data=Yang.tobit2_energy ;
        model Q_ed= logprice_energy logprice_chocolate logprice_sportsdrink hsize hinc
agehh2529 agehh3034 agehh3544 agehh4554
```

```
agehh5564 agehhgt64 emphhpt emphhft eduhhhs eduhhu eduhhpc black oriental other hisp_yes
newengland middleatlantic eastnorthcentral
westnorthcentral southatlantic eastsouthcentral westsouthcentral
mountain aclt6_only ac6_12only ac13_17only aclt6_6_12only aclt6_13_17only
ac6_12and13_17only aclt6_6_12and13_17 fhonly mhonly;
        endogenous Q_ed ~ censored(lowerbound=0);
        nloptions maxiter=500; /*maximum number of iterations set at 300*/
        output out=Yang.Tobit_out_energy conditional expected marginal xbeta;
run;
```

4.2.2 linear price, log income:

```
libname Yang 'Z:\Yang\Yang';
data Yang.tobit; merge Yang.ch_second Yang.ed_second Yang.sp_second;
by household_code;
run;

data Yang.tobit2_energy ; set Yang.tobit;
logprice_chocolate=log(price_chocolate);
logprice_energy=log(price_energy);
logprice_sportsdrink=log(price_sportsdrink);
loghinc=log(hinc);
run;

/*Following is the tobit model for chocolate*/
Proc QLIM data=Yang.tobit2_energy ;
        model Q_ed=  price_energy price_chocolate  price_sportsdrink
hsize loghinc agehh2529 agehh3034 agehh3544 agehh4554 agehh5564 agehhgt64 emphhpt
emphhft eduhhhs eduhhu eduhhpc black oriental other hisp_yes newengland middleatlantic
eastnorthcentral westnorthcentral southatlantic eastsouthcentral westsouthcentral mountain
aclt6_only ac6_12only ac13_17only aclt6_6_12only aclt6_13_17only ac6_12and13_17only
aclt6_6_12and13_17 fhonly mhonly;
        endogenous Q_ed ~ censored(lowerbound=0);
        nloptions maxiter=500; /*maximum number of iterations set at 300*/
        output out=Yang.Tobit_out_energy conditional expected marginal xbeta;
run;
```

4.2.3 log price, log income:

```
libname Yang 'Z:\Yang\Yang';
data Yang.tobit; merge Yang.ch_second Yang.ed_second Yang.sp_second;
by household_code;
run;

data Yang.tobit2_energy ; set Yang.tobit;
logprice_chocolate=log(price_chocolate);
logprice_energy=log(price_energy);
logprice_sportsdrink=log(price_sportsdrink);
```

```
loghinc=log(hinc);
run;

/*Following is the tobit model for chocolate*/
Proc QLIM data=Yang.tobit2_energy ;
	model Q_ed= logprice_energy logprice_chocolate logprice_sportsdrink hsize loghinc
agehh2529 agehh3034 agehh3544 agehh4554 agehh5564 agehhgt64 emphhpt emphhft eduhhhs
eduhhu eduhhpc black oriental other hisp_yes newengland middleatlantic  eastnorthcentral
westnorthcentral southatlantic eastsouthcentral westsouthcentral mountain aclt6_only
ac6_12only ac13_17only  aclt6_6_12only aclt6_13_17only ac6_12and13_17only
aclt6_6_12and13_17 fhonly mhonly;
	endogenous Q_ed ~ censored(lowerbound=0);
	nloptions maxiter=500; /*maximum number of iterations set at 300*/
	output out=Yang.Tobit_out_energy conditional expected marginal xbeta;
run;
```

4.2.4 linear price of choclate milk, others are in log forms:

```
libname Yang 'Z:\Yang\Yang';
data Yang.tobit; merge Yang.ch_second Yang.ed_second Yang.sp_second;
by household_code;
run;

data Yang.tobit2_energy ; set Yang.tobit;
logprice_chocolate=log(price_chocolate);
logprice_energy=log(price_energy);
logprice_sportsdrink=log(price_sportsdrink);
loghinc=log(hinc);
run;



/*Following is the tobit model for chocolate*/
Proc QLIM data=Yang.tobit2_energy ;
	model Q_ed= logprice_energy  price_chocolate logprice_sportsdrink hsize loghinc
agehh2529 agehh3034 agehh3544 agehh4554 agehh5564 agehhgt64 emphhpt emphhft eduhhhs
eduhhu eduhhpc black oriental other hisp_yes newengland middleatlantic eastnorthcentral
westnorthcentral southatlantic eastsouthcentral westsouthcentral mountain
aclt6_only ac6_12only ac13_17only aclt6_6_12only aclt6_13_17only
ac6_12and13_17only aclt6_6_12and13_17 fhonly mhonly;
	endogenous Q_ed ~ censored(lowerbound=0);
	nloptions maxiter=500; /*maximum number of iterations set at 300*/
	output out=Yang.Tobit_out_energy conditional expected marginal xbeta;
run;
```

4.2.5 linear energy drinks price, others are in log forms:

```
libname Yang 'Z:\Yang\Yang';
data Yang.tobit; merge Yang.ch_second Yang.ed_second Yang.sp_second;
```

```
by household_code;
run;

data Yang.tobit2_energy ; set Yang.tobit;
logprice_chocolate=log(price_chocolate);
logprice_energy=log(price_energy);
logprice_sportsdrink=log(price_sportsdrink);
loghinc=log(hinc);
run;

Proc QLIM data=Yang.tobit2_energy ;
        model Q_ed= price_energy  logprice_chocolate logprice_sportsdrink  hsize loghinc
agehh2529 agehh3034 agehh3544 agehh4554 agehh5564 agehhgt64 emphhpt emphhft eduhhhs
eduhhu eduhhpc black oriental other hisp_yes newengland middleatlantic eastnorthcentral
westnorthcentral southatlantic eastsouthcentral westsouthcentral mountain aclt6_only
ac6_12only ac13_17only aclt6_6_12only aclt6_13_17only c6_12and13_17only
aclt6_6_12and13_17 fhonly mhonly;
        endogenous Q_ed ~ censored(lowerbound=0);
        nloptions maxiter=500; /*maximum number of iterations set at 300*/
        output out=Yang.Tobit_out_energy conditional expected marginal xbeta;
run;
```

4.2.6 linear sports drinks price, others are in log forms:

```
libname Yang 'Z:\Yang\Yang';
data Yang.tobit; merge Yang.ch_second Yang.ed_second Yang.sp_second;
by household_code;
run;

data Yang.tobit2_energy ; set Yang.tobit;
logprice_chocolate=log(price_chocolate);
logprice_energy=log(price_energy);
logprice_sportsdrink=log(price_sportsdrink);
loghinc=log(hinc);
run;

Proc QLIM data=Yang.tobit2_energy ;
        model Q_ed= logprice_energy  logprice_chocolate price_sportsdrink  hsize loghinc
agehh2529 agehh3034 agehh3544 agehh4554 agehh5564 agehhgt64 emphhpt emphhft eduhhhs
eduhhu eduhhpc black oriental other hisp_yes newengland middleatlantic eastnorthcentral
westnorthcentral southatlantic eastsouthcentral westsouthcentral mountain aclt6_only
ac6_12only ac13_17only aclt6_6_12only aclt6_13_17only ac6_12and13_17only
aclt6_6_12and13_17 fhonly mhonly;
        endogenous Q_ed ~ censored(lowerbound=0);
        nloptions maxiter=500; /*maximum number of iterations set at 300*/
        output out=Yang.Tobit_out_energy conditional expected marginal xbeta;
run;
```

5. Chi-square Test:

5.1 Chocolate milk:

```
Proc QLIM data=Yang.tobit1 ;
        model Q_CH= logprice_chocolate logprice_energy logprice_sportsdrink
                                        hsize
                                        loghinc agehh2529 agehh3034 agehh3544 agehh4554
agehh5564 agehhgt64
emphhpt emphhft eduhhhs eduhhu eduhhpc
                                        black oriental other hisp_yes newengland
middleatlantic eastnorthcentral
westnorthcentral southatlantic eastsouthcentral westsouthcentral
mountain
aclt6_only ac6_12only ac13_17only aclt6_6_12only aclt6_13_17only
ac6_12and13_17only aclt6_6_12and13_17
                                        fhonly mhonly;
        endogenous Q_CH ~ censored(lowerbound=0);
        nloptions maxiter=500; /*maximum number of iterations set at 300*/
        *hetero Q_CH ~ P_chocolate P_energy hinc agehh2529 agehh3034 agehh3544
agehh4554 agehh5564 agehhgt64 emphhpt emphhft eduhhhs eduhhu eduhhpc
                east midwest south black oriental other hisp_yes aclt6_only ac6_12only
ac13_17only aclt6_6_12only aclt6_13_17only
                ac6_12and13_17only aclt6_6_12and13_17 fhonly mhonly;
        output out=Yang.Tobit_out conditional expected marginal xbeta;*conditional expected
value of Q_CH,
        unconditional expected value of Q_CH,
        unconditional marginal effect of all variables, xbeta;
test agehh2529=0, agehh3034=0, agehh3544=0, agehh4554=0,
agehh5564=0, agehhgt64=0;
 test emphhpt=0, emphhft=0;
 test eduhhhs=0, eduhhu=0, eduhhpc=0;
 test black=0, oriental=0, other-0;
 test newengland=0,
middleatlantic=0, eastnorthcentral=0,
westnorthcentral=0, southatlantic=0, eastsouthcentral=0, westsouthcentral=0,
mountain=0;
test aclt6_only=0, ac6_12only=0, ac13_17only=0, aclt6_6_12only=0, aclt6_13_17only=0,
ac6_12and13_17only=0, aclt6_6_12and13_17=0;
test fhonly=0, mhonly=0;
run;
```

5.2 Energy drinks:

```
Proc QLIM data=Yang.tobit2_energy ;
        model Q_ed= logprice_energy  price_chocolate logprice_sportsdrink
                                        hsize
                                        loghinc agehh2529 agehh3034 agehh3544 agehh4554
```

```
agehh5564 agehhgt64
emphhpt emphhft eduhhhs eduhhu eduhhpc
                                        black oriental other hisp_yes newengland
middleatlantic eastnorthcentral
westnorthcentral southatlantic eastsouthcentral westsouthcentral
mountain
aclt6_only ac6_12only ac13_17only aclt6_6_12only aclt6_13_17only
ac6_12and13_17only aclt6_6_12and13_17
                                        fhonly mhonly;
     endogenous Q_ed ~ censored(lowerbound=0);
     nloptions maxiter=500; /*maximum number of iterations set at 300*/
     *hetero Q_ed ~ P_chocolate P_energy p_sp hinc agehh2529 agehh3034 agehh3544
agehh4554 agehh5564 agehhgt64 emphhpt emphhft eduhhhs eduhhu eduhhpc
          east midwest south black oriental other hisp_yes aclt6_only ac6_12only
ac13_17only aclt6_6_12only aclt6_13_17only
          ac6_12and13_17only aclt6_6_12and13_17 fhonly mhonly;
     output out=Yang.Tobit_out_energy conditional expected marginal xbeta;*conditional
expected value of Q_ed,
     unconditional expected value of Q_ed,
     unconditional marginal effect of all variables, xbeta;
     test agehh2529=0, agehh3034=0, agehh3544=0, agehh4554=0,
agehh5564=0, agehhgt64=0;
 test emphhpt=0, emphhft=0;
 test eduhhhs=0, eduhhu=0, eduhhpc=0;
 test black=0, oriental=0, other-0;
 test newengland=0,
middleatlantic=0, eastnorthcentral=0,
westnorthcentral=0, southatlantic=0, eastsouthcentral=0, westsouthcentral=0,
mountain=0;
test aclt6_only=0, ac6_12only=0, ac13_17only=0, aclt6_6_12only=0, aclt6_13_17only=0,
ac6_12and13_17only=0, aclt6_6_12and13_17=0;
test fhonly=0, mhonly=0;
run;
```

5.3 Sports drinks:

```
Proc QLIM data=Yang.tobit_sp ;
     model Q_sp= logprice_chocolate logprice_energy logprice_sportsdrink
                                        hsize
                                        loghinc agehh2529 agehh3034 agehh3544 agehh4554
agehh5564 agehhgt64
emphhpt emphhft eduhhhs eduhhu eduhhpc
                                        black oriental other hisp_yes newengland
middleatlantic eastnorthcentral
westnorthcentral southatlantic eastsouthcentral westsouthcentral
mountain
aclt6_only ac6_12only ac13_17only aclt6_6_12only aclt6_13_17only
ac6_12and13_17only aclt6_6_12and13_17
```

```
fhonly mhonly;
	endogenous Q_sp ~ censored(lowerbound=0);
	nloptions maxiter=500; /*maximum number of iterations set at 300*/
	*hetero Q_sp ~ P_chocolate P_energy p_sp hinc agehh2529 agehh3034 agehh3544
agehh4554 agehh5564 agehhgt64 emphhpt emphhft eduhhhs eduhhu eduhhpc
		east midwest south black oriental other hisp_yes aclt6_only ac6_12only
ac13_17only aclt6_6_12only aclt6_13_17only
		ac6_12and13_17only aclt6_6_12and13_17 fhonly mhonly;
	output out=Yang.Tobit_out_sp conditional expected marginal xbeta;*conditional
expected value of Q_sp,
	unconditional expected value of Q_sp,
	unconditional marginal effect of all variables, xbeta;
test agehh2529=0, agehh3034=0, agehh3544=0, agehh4554=0,
agehh5564=0, agehhgt64=0;
 test emphhpt=0, emphhft=0;
 test eduhhhs=0, eduhhu=0, eduhhpc=0;
 test black=0, oriental=0, other-0;
 test newengland=0,
middleatlantic=0, eastnorthcentral=0,
westnorthcentral=0, southatlantic=0, eastsouthcentral=0, westsouthcentral=0,
mountain=0;
test aclt6_only=0, ac6_12only=0, ac13_17only=0, aclt6_6_12only=0, aclt6_13_17only=0,
ac6_12and13_17only=0, aclt6_6_12and13_17=0;
test fhonly=0, mhonly=0;
run;
```