Market Competitiveness and Demographic Profiles of Dairy Alternative Beverages in the United States: The Case of Soymilk

Senarath Dharmasena*

Oral Capps, Jr.*

*Agribusiness, Food and Consumer Economics Research Center (AFCERC)

Texas A&M University

Emails: sdharmasena@tamu.edu

ocapps@tamu.edu

Phone: (979) 845-8491

Selected Paper, Southern Agricultural Economics Association (SAEA) Annual Meeting, Birmingham, Alabama, February 4-7, 2012

Market Competitiveness and Demographic Profiles of Dairy Alternative Beverages in the United States: The Case of Soymilk

Abstract

Data from U.S. households for year 2008 were used in examining market competitiveness of soymilk using tobit procedure. Unconditional own- and cross-price elasticities are larger than their conditional counterparts. Income, age, employment status, education level, race, ethnicity, region and presence of children are significant drivers affecting the demand for soymilk.

Key Words: Soymilk, white milk, flavored milk, Nielsen HomeScan data, tobit procedure

JEL Classification: D11, D12

Market Competitiveness and Demographic Profiles of Dairy Alternative Beverages in the United States: The Case of Soymilk

Background:

Currently, calcium fortified soy based dairy alternative beverages are entering the market on one hand to compete with white and flavored milk in the marketplace and on the other providing consumers an alternative calcium fortified beverage, specifically for those who are having trouble consuming dairy based products (due to lactose intolerance and other health concerns). To strengthen the position of calcium fortified soy-based beverages in the U.S. market, the new food guidelines developed under the "ChooseMyPlate" of United States Department of Agriculture, placed calcium fortified soy beverages in the *"Dairy Group"* which is introduced as a side dish (USDA, 2011). This raised eyebrows of dairy producers and dairy marketers in the United States and it is of best interest for them to know the competitiveness of calcium fortified dairy alternative beverages, in particular, soy beverages in the U.S. diary marketplace.

According to Beverage Marketing Corporation (2010), soymilk has been one of the fastest growing categories in the general beverage marketplace and has had a much higher growth rate than the dairy milk segment over the last decade. Growth in soymilk has been attributed to improved health-related claims and consumer perceptions, flurry of soymilk brands, appealing and convenient packaging and multitude on flavors available. Soy beverage retail sales topped to $1.7 billion in 2008 and continue to grow adding flavored soymilk products such as chocolate, vanilla and strawberries, hence directly competing with flavored diary milk (Beverage Marketing Corporation, 2010). As far as brand specific information

is concerned, Silk® soymilk brand has the highest market share (62%), followed by Rice Dream® (6%), 8th Continent® (6%), Lifeway® (2%), and Odwalla® (1%) (Beverage Marketing Corporation, 2010).

Given this backdrop, knowledge of price sensitivity, substitutes and complements and demographic profiles with respect to consumption of soymilk is important for manufacturers, retailers and advertisers of soymilk and dairy milk from a competitive intelligence perspective as well as from a strategic decision-making perspective. We did not find any past study pertaining to demand for soymilk in the extant literature. Therefore, to the best our knowledge, our study is the first to examine the market competitiveness and demographic factors determining U.S. demand for soymilk.

A thorough and a complete analysis of demand for soymilk is important due to increasing growth in consumption in recent times as an alternative beverage to dairy based milk in the United States and to the lack of information in the literature. In this light, specific objectives are: (1) to determine the conditional factors affecting the volume of purchase of soymilk; (2) to determine the unconditional factors affecting the volume of purchase of soymilk; (3) to determine the conditional and unconditional own- and cross-price demand elasticities of soymilk and its competitors; (4) to determine retail level pricing strategies for soymilk in competitive marketplace.

Data and Methodology

Household purchases of soymilk, white milk and flavored milk (expenditure and quantity) and socio-economic-demographic characteristics are generated for each household in the Nielsen Homescan Panel for calendar year 2008 (total of

61,440 households). Only 7,729 households purchased soymilk, while 58,268 households purchased white milk. Flavored milk was purchased by 16,468 households. Quantity data are standardized in terms of liquid ounces and expenditure data are expressed in terms of dollars. Then taking the ratio of expenditure to volume, we generate unit values (prices in dollars per ounce) for each beverage category.

Factors hypothesized to affect the volume of soymilk purchased are: price of soymilk, price of competing beverages such as regular white milk and flavored milk; age, gender, employment and education status of the household head; region; race; Hispanic origin; age and presence of children, income of the household.

A common characteristic in micro level data (data gathered at consumer level such as at the individual or household level) is a situation where some consumers do not purchase some items during the sampling period and presence of them in the sample creates a zero consumption level for that data period. The data used in this study are gathered at household level and due to that it suffers from zero consumption data, hence zero expenditure. As such we face a censored sample of data. Application of ordinary least squares (OLS) to estimate a regression with a limited dependent variable (such as in a censored sample like ours) usually give rise to biased estimates, even asymptotically (Kennedy, 2003). Removing all observations pertaining to zero purchases and estimating regression functions only for non-zero purchases too creates a bias in the estimates (Kennedy, 2003). This phenomenon also is known as *sample selection bias*. Tobin (1958) and Heckman (1979) suggested alternative models to deal with sample selection bias in estimating regression models in the presence of censored data. In this paper,we

center attention on Tobin (1958) model to glean both conditional and unconditional demand estimates pertaining to soymilk. Heckman (1979) model only will be able to speak to conditional demand estimates, although in the first stage probit analysis will provide information on consumer's probability to purchase or not to purchase the product. Also, we use the decomposition of the "beta'" coefficient estimates of tobit model suggested by McDonald and Moffitt (1980) to shed light on changes in probability of being above the limit (limit being zero in this paper) and changes in the value of the dependent variable if it is already above the limit. This is the McDonald and Moffitt decomposition associated with tobit parameter estimates.

For all those transactions associated with zero quantities and hence zero expenditures, we do not observe any unit value or price. However, since we are expecting to use price of each beverage category as explanatory variables in the tobit model, we have to impute price for those observations where no price is observed. Price imputation is done using an auxiliary regression, where observed prices for each beverage are regressed on household income, household size and region where the household is located. These variables are used extensively in the price imputation literature as good instruments in imputing prices. Once the prices for each beverage concerned (soymilk, white milk, and flavored milk) are imputed, we use them and aforementioned explanatory variables to estimate the tobit model pertaining to soymilk consumption. Table 1 shows different categories of explanatory variables used in this study along with base categories for dummy variables.

The Tobit Model

The stochastic model underlying the tobit model can be expressed as follows:

(1) $$y_i = \begin{cases} X_i\beta + u_i, & X_i\beta + u_i > 0 \\ 0, & X_i\beta + u_i \leq 0 \end{cases}$$

where $i = 1,2,3, \ldots\ldots, N$, the number of observations. y_i is the censored dependent variable; X_i is the vector of explanatory variables; β is the vector of unknown parameters to be estimated; $E[u_i|X] = 0$ and $u_i \sim N(0, \sigma^2)$. The unconditional expected value for y_i is expressed in equation (2) and the corresponding conditional expected value for y_i is shown in equation (3), where the normalized index value z is shown as $z = \frac{X\beta}{\sigma}$. Also, $F(z)$ is the cumulative distribution function (CDF) associated with z and $f(z)$ is the corresponding probability density function (pdf).

(2) $$E(y) = X\beta F(z) + \sigma f(z)$$

(3) $$E(y^*) = X\beta + \sigma \frac{f(z)}{F(z)}$$

The unconditional marginal effect is represented by,

(4) $$\frac{\partial E(y)}{\partial X} = \beta F(z)$$

The conditional marginal effete is shown by,

(5) $$\frac{\partial E(y^*)}{\partial X} = \beta\left(1 - z\frac{f(z)}{F(z)} - \frac{f(z)^2}{F(z)^2}\right)$$

Empirical Estimation

We tried several functional forms such as liner, quadratic and linear-log to find that Linear-Log model (we used logged price variables in the model) outperforms other functional forms as far as the model fit, significance of variables and loss matrices such as AIC and Schwarz criteria are concerned. The tobit model for soymilk can be represented as follows,

$$
\begin{aligned}
(Q\_Soy\_Milk) &= \beta_1 + \beta_2 \log PRICE\_SOY_i + \beta_2 \log PRICE\_SOY_i + \\
&\beta_3 \log PRICE\_WMILK_i + \beta_5 PRICE\_FMILK_i + \beta_6 AGEHH2529_i + \\
&\beta_7 AGEHH3034_i + \beta_8 AGEHH3544_i + \beta_9 AGEHH4554_i + \beta_{10} AGEHH5564_i + \\
&\beta_{11} AGEHHGT64_i + \beta_{12} EMPHHPT_i + \beta_{13} EMPHHFT_i + \beta_{14} EDUHHHS_i + \\
&\beta_{15} EDUHHU_i + \beta_{16} EDUHHPC_i + \beta_{17} MIDWEST_i + \beta_{18} SOUTH_i + \\
&\beta_{19} WEST_i + \beta_{20} BLACK_i + \beta_{21} ASIAN_i + \\
&\beta_{22} OTHER_i + \beta_{23} HISP\_YES_i + \beta_{24} AGEPCLT6\_ONLY_i + \\
&\beta_{25} AGEPC6\_12ONLY_i + \beta_{26} AGEPC13\_17ONLY_i + \\
&\beta_{27} AGEPCLT6\_6\_12ONLY_i + \beta_{28} AGEPCLT6\_13\_17ONLY_i + \\
&\beta_{29} AGEPC6\_12AND13\_17ONLY_i + \beta_{30} AGEPCLT6\_6\_12AND13\_17_i + \\
&\beta_{31} MHONLY_i + \beta_{32} FHONLY_i + \beta_{33} INCOME_i
\end{aligned} \tag{6}
$$

As such, we will calculate both conditional and unconditional marginal effects associated with each explanatory variable. The level of significance we will be using in this study is 0.05. We further conduct an *F*-test for demographic variable categories to find statistically significant demographics. The equations for unconditional and conditional marginal effects for the Linear-Log model and corresponding unconditional and conditional own- and cross-price elasticity estimates are explained below.

The unconditional marginal effect for the Linear-Log model is as follows,

$$\frac{\partial E(y)}{\partial X} = \frac{\beta}{P^U} F(z) \tag{7}$$

where P^U is the average price of all observations (unconditional price) considered. The conditional marginal effect for the Linear-Log model is as follows,

(8) $$\frac{\partial E(y^*)}{\partial X} = \frac{\beta}{P^C}\left(1 - z\frac{f(z)}{F(z)} - \frac{f(z)^2}{F(z)^2}\right)$$

Where, P^C is the average price of non-censored sample (conditional price).
The unconditional own- and cross-price demand elasticities are represented by equations (9) and (10) respectively.

(9) $$\varepsilon_{ii}^U = \frac{P_i^U}{Q_i^U}\frac{\beta}{P^U}F(z)$$

(10) $$\varepsilon_{ij}^U = \frac{P_j^U}{Q_i^U}\frac{\beta}{P^U}F(z)$$

The conditional own- and cross-price demand elasticities are represented by equations (11) and (12) respectively,

(11) $$\varepsilon_{ii}^C = \frac{P_i^C}{Q_i^C}\frac{\beta}{P^C}\left(1 - z\frac{f(z)}{F(z)} - \frac{f(z)^2}{F(z)^2}\right)$$

(12) $$\varepsilon_{ij}^C = \frac{P_j^C}{Q_i^C}\frac{\beta}{P^C}\left(1 - z\frac{f(z)}{F(z)} - \frac{f(z)^2}{F(z)^2}\right)$$

Results and Discussion

Our analysis reveals that market penetration for soymilk, white milk and flavored milk is 12.6%, 94.8% and 26.8% respectively. The unconditional own-price elasticity of demand for soymilk is estimated to be -1.98 and its conditional counterpart is -0.41. It must be noted when the whole sample of observations are concerned, soymilk shows an elastic demand vis-à-vis an inelastic demand for the truncated sample with those who actually purchase soymilk. In other words, the sample of consumers who actually bought soymilk is not very price sensitive, whereas the pooled sample (with those who bought and did not buy) would move

away from soymilk to its closest substitute in the event of price increase (more price sensitive).

The unconditional cross-price elasticity of demand with respect to soymilk and white milk is 1.26 and the conditional counterpart associated with this cross-price elasticity is estimated to be 0.25. Similar trend is observed with the cross-price elasticity associated with flavored milk, where the unconditional cross-price elasticity with respect to flavored milk is 0.16 and the conditional cross-price elasticity is estimated to be 0.03. In all, both white milk and flavored milk are substitutes in consumption for soymilk.

Income, age of the household head, employment status of household head, education level of household head, race, ethnicity, region and presence of children in the household are found to be important in affecting the demand for soymilk.

References:

Beverage Marketing Corporation, (2010), Soy Beverages in the U.S.

Heckman, J. J., (1979). "Sample Selection Bias as a Specification Error". *Econometrica* 47: 153-161

Kennedy, P. (2003). Limited Dependent Variables. A Guide to Econometrics, MIT Press.

McDonald, J. F., and R. A. Moffitt, (1980), "The Uses of Tobit Analysis." *The Review of Economics and Statistics,* 62(2): 318-321

Tobin, J., (1958), "Estimation of Relationships for Limited Dependent Variables." *Econometrica,* 26(1): 24-36

USDA, (2011). "What foods are included in the Dairy Group?". ChooseMyPlate, (internet access on September 9, 2011: http://www.choosemyplate.gov/foodgroups/dairy.html)

Table 1 Description of the Right-Hand Side Variables Used in the Econometric Analysis

Variable	Explanation
PRICE	Price of Drinkable yogurt
AGEHHLT25	*Age of Household Head less than 25 years (Base category)*
AGEHH2529	Age of Household Head between 25-29 years
AGEHH3034	Age of household Head between 30-34 years
AGEHH3544	Age of household Head between 35-44 years
AGEHH4554	Age of household Head between 45-54 years
AGEHH5564	Age of household Head between 55-64 years
AGEHHGT64	Age of household Head greater than 64 years
EMPHHNFP	*Household Head not employed for full pay (Base category)*
EMPHHPT	Household Head Part-time Employed
EMPHHFT	household Head Full-time Employed
EDUHHLTHS	*Education of Household Head: Less than high school (Base category)*
EDUHHHS	Education of Household Head: High school only
EDUHHU	Education of Household Head: Undergraduate only
EDUHHPC	Education of Household Head: Some post-college
EAST	*Region: East (Base category)*
MIDWEST	Region: Central (Midwest)
SOUTH	Region South
WEST	Region West
WHITE	*Race White (Base category)*
BLACK	Race Black
ASIAN	Race Oriental
RACE_OTHER	Race Other (non-Black, non-White, non-Oriental)
HISP_NO	*Non-Hispanic Ethnicity (Base category)*
HISP_YES	Hispanic Ethnicity

Table 1 Continued....

Variable	Explanation
NPCLT_18	*No Child less than 18 years (Base category)*
AGEPCLT6_ONLY	Age and Presence of Children less than 6-years
AGEPC6_12ONLY	Age and Presence of Children between 6-12 years
AGEPC13_17ONLY	Age and Presence of Children between 13-17 years
AGEPCLT6_6_12ONLY	Age and Presence of Children less than 6 and 6-12 years
AGEPCLT6_13_17ONLY	Age and Presence of Children less than 6 and 13-17 years
AGEPC6_12AND13_17ONLY	Age and Presence of Children between 6-12 and 13-17 years
AGEPCLT6_6_12AND13_17	Age and Presence of Children less than 6, 6-12 and 13-17 years
FHMH	*Household Head both Male and Female (Base category)*
MHONLY	Household Head Male only
FHONLY	Household Head Female only