Consumer Demand for Dairy Alternative Beverages in the United States and its Implications to U.S. Dairy Industry

Alicia Copeland

Department of Agricultural Economics

Texas A&M University

alicope@tamu.edu

Senarath Dharmasena

Department of Agricultural Economics

Texas A&M University

sdharmasena@tamu.edu

Consumer Demand for Dairy Alternative Beverages in the United States and its Implications to U.S. Dairy Industry

Alicia Copeland and Senarath Dharmasena

Abstract

Production and consumption of dairy alternative beverages in the United States has been on the rise as per capita consumption of fluid milk continues to fall. Almond milk and soymilk are the fastest growing categories in the U.S. dairy alternative marketplace. Using household-level purchase data from 2011 Nielsen Homescan panel and tobit econometric procedure, the conditional and unconditional own-price, cross-price and income elasticities for soymilk and almond milk will be estimated. Income, age, employment status, education level, race, ethnicity, region and presence of children are significant drivers affecting the demand for dairy alternative beverages, such as almond milk and soy milk. This paper investigates the growth of the dairy alternative beverage market in the United States and its implications for dairy farmer welfare.

Background and Justification

There are many different types of nonalcoholic beverages available in the United States today. Functionality and health dimensions of beverages have changed over the years. On top of conventional hydration and refreshment functions, beverages now are fortified with numerous vitamins, minerals, proteins, antioxidants, favorable fatty acids, etc. (BMC, 2010; 2011, 2012).

Currently, calcium and vitamin fortified dairy alternative beverages are entering the market to compete with dairy milk, providing consumers an alternative, specifically for those who are lactose intolerant. To strengthen the position of this, the new food guidelines developed under the "ChooseMyPlate", placed dairy alternatives such as soymilk, rice milk and almond milk in the "Dairy Group" (USDA, 2014). This placement raised eyebrows of dairy producers and marketers in the United States. Although, the dairy industry in the United States offers a wide array of milk and processed dairy products to consumers, per capita consumption of milk has been declining over the past 25 years ((Davis *et al.,* 2010; USDA-ERS, 2013). This decline in demand for dairy milk could probably be due to changing consumer perceptions as well as presence of wide array of dairy alternatives now available in the market.

Dairy-alternative products represented roughly five percent of dairy launches in 2012, with soy being the primary or secondary ingredient in 78 percent of them (Innova Market Insights, 2013). However, this trend with respect to soy is changing as interest is growing in dairy alternatives made with ingredients including almonds, rice, oats, barley, hazelnuts and walnuts.

According to Chicago based market research firm, Mintel, almond milk has overtaken soymilk over the past two years and has become America's most popular plant-based milk alternative accounting for 4.1% of total milk sales (KCT.org, 2014). Almond milk now

dominates dairy alternative beverage market with a staggering 60% market share, while soymilk has only about 30% share (Food Navigator, 2014). Growth in dairy alternatives has been attributed to improved health-related claims and consumer perceptions, a flurry of brands, appealing and convenient packaging, and a plethora of flavors available. Also, vegans, vegetarians and consumers concerned with the additives in dairy milk, such as growth hormones and antibiotics, are now opting to purchase dairy alternatives instead of dairy milk (Neville, 2015). Sales of dairy alternative beverages reached nearly $2 billion in 2013, driven up largely as a result of popularity of almond milk (The Washington Post, 2014).

This increasing demand for dairy alternative beverages and declining demand for dairy milk in the United States could negatively affect dairy poducers in terms of low prices for dairy milk as well as reduced farm income/welfare. Therefore, it is of interest for dairy producers in the United States to know the competitiveness of dairy alternatives in the dairy marketplace and their implications on dairy prices and farm income/welfare.

Objectives
Our study has three specific objectives. First, we estimate demand for almond milk, soymilk, dairy milk (white), dairy milk (flavored), other dairy alternative beverages (rice milk, coconut milk) in a demand system framework. Second, we estimate the economic and demographic profiles of dairy alternative beverage consumers in the United States. Lastly, we investigate the economic ramifications on U.S. milk producers in the event that demand for dairy alternative beverages continues to grow as well as if over-capacity occurs, and leads to declines in the dairy alternative price, the overall price received by dairy farmers.

Data and Methodology

Household purchases of soymilk, almond milk (expenditure and quantity) and socio-economic-demographic characteristics are generated for each household in the Nielsen Homescan panel for calendar year 2011 (a total of 62,092 households), the most recent year currently available to us. Only 6,776 households purchased soymilk, while 7,487 households purchased almond milk. 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 quantity of soymilk and almond milk purchased are: price of soymilk, price of almond milk; age, gender, employment and education status of the household head; region; race; Hispanic origin; age and presence of children, income of the household. We hypothesize that almond milk and soymilk are substitutes, hence positive cross-price elasticities. Also, we hypothesize that education status, hence the knowledge of the product, increases the consumption of each beverage; high income households consume more of each beverage; age and presence of children at home increases the consumption of each beverage; full-time employed households consume more away from-home, hence less soymilk and almond milk are consumed at home; households in the South Atlantic region of the U.S. consume more soymilk and almond milk; Whites consume more soymilk and almond milk.

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 may not purchase some beverages during the sampling period. The presence of these in the sample creates a zero consumption level for that observation, 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) gives 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)[1] 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's model (Tobin, 1958) to obtain both conditional and unconditional elasticity estimates pertaining to soymilk and almond milk. Also, we use the decomposition of the coefficient estimates of tobit model suggested by McDonald and Moffitt (1980) to shed light on changes in probability of being above the limit (the limit being zero in this analysis) and changes in the value of the dependent variable if it is already above the limit.

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 using price of each beverage category as explanatory variables in the tobit model, we have to impute prices 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[2]. These variables are used extensively in the price imputation literature to impute prices (Kyureghian, Nayga and Capps, 2011; Alviola and Capps, 2010). Estimated parameters from this auxiliary regression are then used to impute prices for

---

[1] Alternatively, the Heckman (1979) model only speaks to conditional demand estimates, although the first stage probit analysis provides information on the probability to purchase or not to purchase the product.

[2] Here we provide summary statistics for observed prices and imputed prices for each beverage category. According to means and standard deviations of observed and imputed prices for each beverage, it is clear that the prices and standard deviations were very consistent for with-in sample estimates as well as out-of-sample price imputations.

	Observed Price		Imputed Price	
	Mean	Standard deviation	Mean	Standard deviation
Almond Milk	0.0530	0.0130	0.0531	0.0020
Soymilk	0.0547	0.0167	0.0548	0.0017

those observations where price was not observed. This price imputation technique is well accepted in extant literature and a very common approach to deal with imputing (or forecasting) missing prices and price endogeneity issues (for example see Capps, et al, 1994; Alviola and Capps, 2010; Kyureghian, Nayga and Capps, 2011; and Dharmasena and Capps, 2012). Variability of demand for different quality of beverages is addressed via income variable in the auxiliary regression. Likewise, variability of socio-demographic conditions and its effect on price is approximated via household size variable. The variability in the location of the household and its effect on price is addressed through region variable in the auxiliary regression. Once the prices for each beverage concerned (soymilk and almond milk) are imputed, we use them and the other explanatory variables to estimate the tobit model pertaining to soymilk and almond milk consumption. Description of the explanatory variables used in the tobit analysis of soymilk and almond milk are shown in Table 1.

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 $i = 1,2,3, \ldots, N$, 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 effect is shown by,

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

Furthermore, the McDonald and Moffitt (1980) decomposition relating both change in

conditional expectations and unconditional expectations can be shown in equation (6). In other

words, the total change in unconditional expected value of the dependent variable, *y* can be

represented by the sum of the change in the expected value of *y* being above the limit, weighted

by the probability of being above the limit and the change in probability of being above the limit

weighted by the expected value of *y* being above the limit.

(6) $\frac{\partial E(y)}{\partial X} = F(z)\left(\frac{\partial Ey^*}{\partial X}\right) + E(y^*)\left(\frac{\partial F(z)}{\partial X}\right)$

Empirical Estimation

Single equation tobit models each for soymilk and almond milk are estimated. We are

expecting to try several functional forms such as linear, quadratic and semi-log to find which

model performs best based on the following criteria, model fit, significance of variables and loss

metrics such as the Akaike Information Criteria (AIC), Schwarz Information Criteria (SIC) and

Hannan-Quinn Information Criteria (HQC). Ultimately we will use the best functional form to

calculate both conditional and unconditional marginal effects associated with each explanatory

variable. The level of significance used in this study is 0.05 (*p*-value is 0.05). For preliminary

analysis we used the semi-log functional form and following derivations and results are based off

of this functional form. The equations for unconditional and conditional marginal effects for the

semi-log model and the corresponding unconditional and conditional own-price, cross-price and

income elasticity estimates are explained below.

The unconditional marginal effect for the price variable of the semi-log model is as follows,

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

where P^U is the average price of all observations (unconditional price) for each beverage

considered. The conditional marginal effect for the price variable for the linear-log model is as

follows,

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

where, p^C is the average price of censored sample (conditional price) for each beverage

considered. The unconditional income effect for each beverage for the linear-log model is

expressed in equation (9) and the conditional income effect for each beverage for the linear-log

model is shown in equation (10).

(9) $\quad \frac{\partial E(y)}{\partial I} = \frac{\beta}{I^U} F(z)$

(10) $\quad \frac{\partial E(y^*)}{\partial I} = \frac{\beta}{I^C} \left(1 - z\frac{f(z)}{F(z)} - \frac{f(z)^2}{F(z)^2}\right)$

where, I^U is the unconditional mean income and I^C is the conditional mean income. The

unconditional own- price, cross-price and income elasticities are represented by equations (11),

(12) and (13) respectively.

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

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

(13) $\varepsilon_I^U = \frac{\beta}{I_i^U} F(z) \frac{I_i^U}{Q_i^U}$

The conditional own-price, cross-price and income elasticities are represented by equations (14), (15), (16) respectively,

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

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

(16) $\varepsilon_I^C = \frac{\beta}{I_i^C} (1 - z \frac{f(z)}{F(z)} - \frac{f(z)^2}{F(z)^2}) \frac{I_i^C}{Q_i^C}$

The McDonald and Moffitt (1980) decomposition explained in equation (6) can be manipulated to obtain the expression shown in equation (17) to shed light on change in probability of being above the limit (for conditional sample) for consumption of each beverage category for a change in each explanatory variable, i.e. $\left(\frac{\partial F(z)}{\partial X}\right)$.

(17) $\left(\frac{\partial F(z)}{\partial X}\right) = \frac{1}{E(y^*)}$

Preliminary Results and Discussion

Preliminary analysis was performed used 2011 Nielsen Homescan data comprised of 62,092 households. The tobit model estimates are presented in Table 2. Currently we are in the process of calculating conditional and unconditional marginal effects and the elasticities. Some summary statistics results are discussed below.

Market penetration for soymilk was found to be 11%, while market penetration for almond milk was found to be 12%. The average price paid by households who purchased soymilk was $0.05 per ounce ($3.50 for 64 ounces; the most popular container size). The average

price paid by households who purchased almond milk was $0.05 per ounce ($3.39 for 64 ounces). The average consumption/purchase of soymilk by a consuming household was estimated to be 480 ounces per year (approximately eight half gallon containers per household per year). The average consumption/purchase of almond milk by a consuming household was estimated to be 424 ounces per year (approximately seven half gallon containers per household per year).

We also found that household composition and demographic characteristics played an important role in the demand for both almond and soymilk. Households in the South Atlantic region of the United States (Delaware, Washington DC, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, and West Virginia) consumed more soymilk and almond milk than those from other regions. Those who are classified as White consumed more soymilk as well as almond milk.

While the present analysis is somewhat limited with our focus on overall demand for almond milk and soymilk, this preliminary analysis puts us in position to estimate own-price, cross-price and expenditure elasticities for the separable food groups. Also, we will be profiling demographic characteristics of consumers with regards to these food groups. Lastly, using estimated elasticities we will be in position to discuss the welfare effects of the dairy alternative beverage boom on U.S. dairy farmers.

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.

Balagtas, J. and D. Sumner. 2003,"The Effect of the Northeast Dairy Compact on Producers and Consumers with Implications of Compact Contagion." Review of Agricultural Economics 25(1): 123-144.

Banks, J., R. Blundell, and A. Lewbel, 1997, "Quadratic Engle Curves and Consumer Demand." *The Review of Economics and Statistics,* 79(4):527-539

BMC, 2010, 2011, 2012, Beverage Marketing Corporation Multiple Beverage Marketplace Reports

Capps, Jr. O., R. Tsai, 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.

Davis, Christopher G., Diansheng Dong, Don P. Blayney, and Ashley Owens. An Analysis of U.S. Household Dairy Demand, TB-1928, U.S. Dept. of Agr.,Econ. Res. Serv. December 2010.

Dharmasena, S., O.Capps, Jr. 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, first published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/hec.1738.

Food Navigator, (2014), "Almond Milk Accounts for Two-Thirds of Sales in Plant-Based Milk Category." Internet access: http://www.foodnavigator-usa.com/Manufacturers/Almond-milk-accounts-for-two-thirds-of-sales-in-plant-based-milk-category-says-WhiteWave-Foods (accessed on September 3, 2014).

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

Innova Market Insights, FoodBusinessNews.net, accessed January 7, 2013.

KCT.org, (2014) "Demand for Almond Milk is Growing, but is it Bad for the Planet?" Internet access: http://www.kcet.org/living/food/the-nosh/demand-for-almond-milk-is-soaring-but-is-it-bad-for-the-planet.html (accessed on September 3, 2014)

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

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

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

Neville, Antal (2015). "Soy and Almond Milk Production in the US" . IBISWorld, (internet access on April 1, 2014: http://clients1.ibisworld.com/reports/us/industry/default.aspx?entid=4195)

Sam, A. G., and Y. Zheng, 2010, "Semiparametric Estimation of Consumer Demand Systems with Micro Data." "American Journal of Agricultural Economics, 92(1):246-257

The Washington Post, (2014), "Got Milk? From a Cow or a Plant?" Internet access: http://www.washingtonpost.com/news/to-your-health/wp/2014/06/10/got-milk-from-a-cow-or-a-plant/ (acceded on September 3, 2014).

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

USDA, (2014). "What foods are included in the Dairy Group?". ChooseMyPlate, (internet access on September 3, 2014: http://www.choosemyplate.gov/food-groups/dairy.html)

USDA-ERS Economic Research Service (ERS). 2013. Diary Data: http://www.ers.usda.gov/data-products/dairy-data.aspx#.UXSd6yvwLTV. Internet access December 27, 2013.

Table 1: Description of the Explanatory Variables Used in the Tobit Analysis of Soymilk and Almond Milk

Explanation
Price of Soymilk and Almond Milk (all in $/oz)
Household Income (dollars)
Age of Household Head less than 25 years (Base category)
Age of Household Head between 25-29 years
Age of household Head between 30-34 years
Age of household Head between 35-44 years
Age of household Head between 45-54 years
Age of household Head between 55-64 years
Age of household Head greater than 64 years
Household Head not employed for full pay (Base category)
Household Head Part-time Employed
household Head Full-time Employed
Education of Household Head: Less than high school (Base category)
Education of Household Head: High school only
Education of Household Head: Undergraduate only
Education of Household Head: Some post-college
Region: East (Base category)
Region: Central (Midwest)
Region South
Region West
Race White (Base category)
Race Black
Race Oriental
Race Other (non-Black, non-White, non-Oriental)
Non-Hispanic Ethnicity (Base category)
Hispanic Ethnicity
No Child less than 18 years (Base category)
Age and Presence of Children less than 6-years
Age and Presence of Children between 6-12 years
Age and Presence of Children between 13-17 years
Age and Presence of Children less than 6 and 6-12 years
Age and Presence of Children less than 6 and 13-17 years
Age and Presence of Children between 6-12 and 13-17 years
Age and Presence of Children less than 6, 6-12 and 13-17 years
Household Head both Male and Female (Base category)
Household Head Male only
Household Head Female only

Source: Constructed by authors; base category of dummy variables are printed in italics.

Table 2: Tobit Regression Results for Soymilk and Almond Milk

Variable	Soymilk			Almond Milk		
	Estimate	Std Error	*p*-Value	Estimate	Std Error	*p*-Value
Intercept	-9278.2072	397.0687	<.0001	-6770.1670	329.2582	<.0001
Log price soymilk	-1723.4721	68.5338	<.0001	-545.8587	77.3053	<.0001
Log price almond milk	-996.6089	109.6160	<.0001	-1210.6443	65.4958	<.0001
Household income	2.1374	0.3364	<.0001	2.3043	0.2736	<.0001
Age of household head 25-29	-397.0727	183.5472	0.0305	218.1864	180.0338	0.2255
Age of household head 30-34	-470.7336	178.6003	0.0084	148.8665	176.8362	0.3999
Age of household head 35-44	-500.3157	174.8920	0.0042	125.6555	174.4868	0.4714
Age of household head 45-54	-550.4484	174.1740	0.0016	9.9455	174.0606	0.9544
Age of household head 55-64	-563.1433	174.1171	0.0012	-31.5616	174.0094	0.8561
Age of household head >64	-620.1727	174.7188	0.0004	-141.7247	174.4095	0.4164
Employment status part-time	67.4579	25.6930	0.0087	70.0099	20.8655	0.0008
Employment status full-time	-38.8605	23.0105	0.0913	-59.7592	18.7477	0.0014
Education: high school	1.0461	64.8797	0.9871	107.5217	56.8678	0.0587
Education: undergraduate	154.7967	63.3361	0.0145	272.3864	55.6492	<.0001
Education post-college	218.4007	67.9804	0.0013	315.8985	59.1747	<.0001
New England	-75.9570	50.2566	0.1307	-95.5184	40.4384	0.0182
Middle Atlantic	-29.9272	35.6022	0.4006	-48.0795	28.7430	0.0944
East North Central	-191.5047	33.8203	<.0001	-236.9577	27.7256	<.0001
West North Central	-207.8269	41.4558	<.0001	-230.2061	33.9590	<.0001
South Atlantic	-104.9438	32.6981	0.0013	-49.1965	26.1086	0.0595
East South Central	-230.2380	46.2173	<.0001	-227.7682	37.3055	<.0001
West South Central	-170.4134	38.2607	<.0001	-250.7022	31.4628	<.0001
Mountain	-100.7484	41.4034	0.0150	-59.5046	33.2067	0.0731
Black	306.4138	28.9526	<.0001	160.9915	24.0667	<.0001
Asian	380.6302	47.4880	<.0001	187.7611	39.8391	<.0001
Other	153.7278	45.9466	0.0008	83.4705	38.1870	0.0288
Hispanic	218.7281	40.3453	<.0001	121.3651	33.5999	0.0003
Children less than 6 years	39.3359	54.9568	0.4741	-56.6692	45.4745	0.2127
Children 6-12years	17.2180	40.9713	0.6743	-57.3301	33.8431	0.0903
Children 13-17years	-23.0820	36.8255	0.5308	-94.9405	30.4853	0.0018
Children < 6 & 6-12 years	-152.6524	60.6536	0.0118	-44.6119	47.3075	0.3457
Children <6 & 13-17years	-75.6982	136.7221	0.5798	-190.2520	113.2898	0.0931

Variable	Soymilk			Almond Milk		
	Estimate	Std Error	*p*-Value	Estimate	Std Error	*p*-Value
Children 6-12&13-17years	-174.6827	53.1200	0.0010	-194.8537	43.1861	<.0001
Children <6 & 6-12&13-17	-133.9031	124.3219	0.2815	-115.2011	99.8369	0.2485
Female head only	-1.1237	24.0332	0.9627	84.6910	19.3962	<.0001
Male head only	-179.1168	34.9760	<.0001	-205.0115	29.2588	<.0001
Sigma	1338.6091	13.0729	<.0001	1124.4188	10.5130	<.0001