

Is Chocolate Milk the New-Age Energy\Sports Drink in the United States?

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 prepared for presentation at the Southern Agricultural Economics Association Annual Meeting, Corpus Christi, TX, February 5-8, 2011

Copyright 2011 by Senarath Dharmasena and Oral Capps, Jr. All rights reserved. Readers may make verbatim copied of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Is Chocolate Milk the New-Age Energy\Sports Drink in the United States?

Senarath Dharmasena and Oral Capps, Jr.

Abstract

Data from U.S. households for calendar year 2008 were used in examining demographic and economic factors affecting demand for chocolate milk using Heckman two-step procedure. Price, income, age, education, region, race, Hispanic status, and presence of children were significant drivers of consumption of chocolate milk. Sample selection bias was statistically significant.

Key Words: Chocolate milk, Nielsen HomeScan data, Heckman two-step

JEL Classification: D11, D12

Is Chocolate Milk the New-Age Energy\Sports Drink in the United States?

Senarath Dharmasena and Oral Capps, Jr.

Background:

Energy and sports drinks are sold in the market as a functional beverage serving to boost on-the-go lifestyles and to rejuvenate individuals after workouts (Beverage Marketing Corporation, 2010). However, Karp *et al.*, (2006) and Thomas *et al.*, (2009) suggest consumption of chocolate milk *vis-à-vis* sports and energy drinks is an effective recovery aid after prolonged workouts. To strengthen the role of chocolate milk as a new-age sports/energy drink, the “Got Milk?” campaign with the participation of U.S Olympic celebrities promotes chocolate milk as an “easy, effective and cost efficient” way to fuel up the body after an intense workout (Brandweek, 2010). According to NPD Group (2010) and Nielsen (2010), U.S. consumption of chocolate milk is growing and servings of plain chocolate milk grew from 1.2 billion in 2009 to 1.4 billion in 2010.

Given this backdrop, 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 from a competitive intelligence perspective as well as from a strategic decision-making perspective. We could not find any past study pertaining to demand for chocolate milk in the extant literature. Therefore, to our knowledge, our study is the first to examine the economic and demographic factors determining U.S. demand for chocolate milk.

Objectives

A thorough and a complete analysis of demand for chocolate milk is important due to increasing growth in consumption in recent times as an alternative beverage to sports and energy drinks and to the lack of information in the literature. In this light, specific objectives are: (1) to determine the factors affecting the decision to purchase chocolate milk, and (2) once the decision to purchase chocolate milk is made, to determine the drivers of purchase volume.

Methodology

At first, household purchases of chocolate 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 15,078 unique households purchased chocolate 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). Using this data set, we estimate demand for drinkable yogurts with adjustment to sample selection bias (Heckman, 1979).

Factors hypothesized to affect the decision to buy chocolate milk and volume of chocolate milk purchased are: price of chocolate milk, and host of demographic characteristics such as, gender, employment and education status of the household head; region; race; Hispanic origin; age and presence of children, and income of the household head.

Model Development, Procedures and Variables

Choice to purchase or not to purchase chocolate milk could be affected by price of chocolate milk and various demographic factors. This type of choice is a dichotomous discrete (buy or not-to buy or “one” if buy and “zero” if do not buy) and a probit model is used generally to model such a choice decision. The dependent variable is a zero one type dummy variable which is created to reflect the non-purchase or purchase respectively of chocolate milk. It is regressed on price and a host of demographic factors. Probit analysis will provide statistically significant findings of the decision to purchase chocolate milk.

Demographic and economic factors hypothesized to be affecting the decision to buy chocolate milk are listed on Table 1. Also, we provide different categories used in each factor along with base category for dummy variables.

The probit model for chocolate milk can be written as follows:

$$\begin{aligned} \Pr(Y = 1 | x_i' \beta) = & \beta_1 + \beta_2 PRICE_i + \beta_3 AGEHH 2529_i + \beta_4 AGEHH 3034_i + \\ & \beta_5 AGEHH 3544_i + \beta_6 AGEHH 4554_i + \beta_7 AGEHH 5564_i + \beta_8 AGEHHGT 64_i + \\ & \beta_9 EMPHHPT_i + \beta_{10} EMPHHFT_i + \beta_{11} EDUHHHS_i + \beta_{12} EDUHHU_i + \\ & \beta_{13} EDUHHPC_i + \beta_{14} MIDWEST_i + \beta_{15} SOUTH_i + \\ & \beta_{16} WEST_i + \beta_{17} BLACK_i + \beta_{18} ASIAN_i + \\ & \beta_{19} OTHER_i + \beta_{20} HISP_YES_i + \beta_{21} AGEPCLT 6_ONLY_i + \\ & \beta_{22} AGEPC 6_12ONLY_i + \beta_{23} AGEPC13_17ONLY_i + \\ & \beta_{24} AGEPCLT 6_6_12ONLY_i + \beta_{25} AGEPCLT 6_13_17ONLY_i + \\ & \beta_{26} AGEPC 6_12AND13_17ONLY_i + \beta_{27} AGEPCLT 6_6_12AND13_17_i + \\ & \beta_{28} MHONLY_i + \beta_{29} FHONLY_i + \beta_{30} INCOME_i \end{aligned} \quad (1)$$

where $i = 1, \dots, n$ is the number of households. Y corresponds to the decision to buy chocolate milk. Variables are defined in Table 1.

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. 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. This phenomenon also is known as *sample selection bias*. Heckman (1979) stated that not adjusting for sample selection may result in biased estimates of the demand parameters. Furthermore, Heckman (1979), discussed the sample selection bias as a specification error, and developed a simple consistent estimation method that eliminates the specification error for the case of censored samples. It is known as Heckman-type correction procedure.

The first stage of the Heckman-two-step sample selection procedure, involves in decision to purchase chocolate milk. It is modeled through a probit model. A binary dependent variable is observed (purchase or not purchase), where purchase is represented by one (1) and not purchase is given by a zero (0). The latent selection equation can be written as follows;

$$Z_h = w'_h \gamma + \varepsilon_h \quad (2)$$

where Z_k represents a latent selection variable (buy or not to buy type dichotomous variable),

$$Z_h = \begin{cases} 1 & \text{if } Z_h > 0 \\ 0 & \text{if } Z_h < \text{or} = 0 \end{cases} \quad (3),$$

w_h is a vector of explanatory variables in the latent decision making variable, γ_h is a vector of parameters to be estimated in the decision making equation, ε_h is the error term, and $h = 1, 2, \dots, N$ is the number of observations (in our work the number of households in the sample) in the sample. Modeling above equation 2 through probit model gives us following relationships;

$$\Pr[Z_h = 1] = \phi(w_h, \gamma) \quad (4) \text{ and}$$

$$\Pr[Z_h = 0] = 1 - \phi(w_h, \gamma) \quad (5)$$

where ϕ is the normal cumulative probability distribution function (cdf). The first stage estimation provides estimates of γ and the inverse of the Mills Ratio (IMR hereinafter). We also generate the associated probability density function (pdf). Inverse of Mills Ratio is calculated taking the ratio of pdf to cdf. Mathematically, it is as follows;

$$\text{for } Z_k = 1, \text{ } IMR_h = \frac{\varphi(w_h \hat{\gamma})}{\phi(w_h \hat{\gamma})} \quad (6),$$

where φ represents the probability density function. Inverse mills ratio is a monotone decreasing function of the probability that an observation is selected into the sample, $\phi(w_k \hat{\gamma}_k)$ (Heckman, 1979). In particular,

$$\lim_{\phi(Z_{hi}) \rightarrow 1} IMR_h = 0 \quad (7)$$

$$\lim_{\phi(Z_{hi}) \rightarrow 0} IMR_h = \infty \quad (8)$$

$$\frac{\partial IMR_h}{\partial \phi(Z_h)} < 0 \quad (9)$$

The calculated IMR, will be used as an additional explanatory variable in the second stage volume equation, which takes care of the sample selection bias in the data. Second stage equation is given as follows;

$$E[Y_h | Z_h = 1] = X'_h \beta + \alpha \frac{\varphi(w_h \hat{\gamma})}{\phi(w_h \hat{\gamma})} \quad (10)$$

$$E[Y_h | Z_h = 1] = X'_h \beta + \alpha IMR_h \quad (11)$$

where X_k is a vector of explanatory variables considered in the second stage. Importantly, only observations associated with non-zero observations on Y_k are considered here. The IMR calculated using information retrieved from first stage probit model is used as an explanatory variable in the second stage (see equations 10 and 11 above). Presence of a sample selection bias in data will be communicated through statistical significance of the coefficient associated with IMR, i.e. α_k . If α_k is statistically not different from zero, we conclude that there is no sample selection bias in the data and result in the following regression model;

$$E[Y_h | Z_h = 1] = X'_h \beta_i \quad (12)$$

It is important to know that the explanatory variables in first stage and second stage equations may or may not be the same. In our work, the price variables in both equations do not. However, rest of the demographic variables is exactly the same in the first stage and second stage.

Choice of explanatory variables in the first and second stage has an implication on the derivation and interpretation of marginal effects associated with variables in the second stage. This is because in the second stage, we have the IMR term augmenting the regular regression function with other explanatory variables. Therefore, in calculating marginal

effects, the influence of IMR and its associated regression coefficient on other regression coefficients have to be taken into consideration.

Suppose X_{kj} denote the j th regressor that is common to both first stage regressors, w_k and, second stage regressors, X_j . Differentiating equation 11 with respect to j th regressor, the marginal effect is given by the following relationship (following explanation is borrowed from Saha, Capps and Byrne (1997));

$$\frac{\partial E[Y_h | Z_h = 1]}{\partial X_{kj}} = \beta_{ij} + \alpha_i \frac{\partial(IM\hat{R}_{hi})}{\partial X_{hj}} \quad (13)$$

It is evident from 13 that marginal effect of the j th regressor on Y_{ki} consists of two parts: a change in X_j which affects the probability of consuming the commodity (this effect is

represented by $\frac{\partial(IM\hat{R}_{hi})}{\partial X_{hj}}$ in 13); a change in X_j which affects the level of consumption (or

expenditure of consumption) which is conditional upon the household choosing to consume

the i th commodity (this is represented by β_{ij} in 13). The former of the above two expression is

important, because the sign and magnitude of the marginal effect depends not only on the β_{ij} ,

but also that of the $\frac{\partial(IM\hat{R}_{hi})}{\partial X_{hj}}$. According to Saha, Capps and Byrne (1997), after some

simplification we get arrive at the following relationship for the Heckman second stage marginal effects,

$$ME_{kj}^{\hat{E}} = \frac{\partial E[y_k | Z = 1]}{\partial X_{kj}} = \beta_j - \alpha \gamma_{ij} \{W\lambda MR_k + (IMR_k)^2\} \quad (14)$$

In general the marginal effect $ME_{kj}^{\hat{E}} \neq \hat{\beta}_j$; however the only case where $ME_{kj}^{\hat{E}} = \hat{\beta}_j$ is where

$\hat{\alpha} = 0$ which is a situation where the errors in the first-stage and second-stage estimation

equations have zero covariance. It must be noted that the ME_{kj} estimation depends on a local set of co-ordinates. Therefore, we estimate the ME_{kj} at the sample means. Following equation 14 shows this result. For simplicity, let us denote IMR in the letter λ .

$$ME_{kj} |_{samplemean} = \hat{\beta}_j - \hat{\alpha}_i \hat{\gamma}_j \{ (\overline{W} \hat{\gamma}) \hat{\lambda} + \hat{\lambda}^2 \} \quad (15)$$

where \overline{W} denotes the vector of regressor sample means in the probit equation (the first stage equation of the Heckman two-step model and

$$\hat{\lambda} = \frac{\varphi(\overline{W} \hat{\gamma})}{\phi(\overline{W} \hat{\gamma})} \quad (16)$$

is the inverse Mills ratio evaluated at those means.

The Heckman two-step demand model for chocolate milk can be written as follows:

$$\begin{aligned} q_i = & \beta_1 + \beta_2 P_i + \beta_3 AGEHH2529_i + \beta_4 AGEHH3034_i + \\ & \beta_5 AGEHH3544_i + \beta_6 AGEHH4554_i + \beta_7 AGEHH5564_i + \beta_8 AGEHHGT64_i + \\ & \beta_9 EMPHHPT_i + \beta_{10} EMPHHFT_i + \beta_{11} EDUHHHS_i + \beta_{12} EDUHHU_i + \\ & \beta_{13} EDUHHPC_i + \beta_{14} REG_CENTRAL_i + \beta_{15} REG_SOUTH_i + \\ & \beta_{16} REG_WEST_i + \beta_{17} RACE_BLACK_i + \beta_{18} RACE_ORIENTAL_i + \\ & \beta_{19} RACE_OTHER_i + \beta_{20} HISP_YES_i + \beta_{21} AGEPCLT6_ONLY_i + \\ & \beta_{22} AGEPC6_12ONLY_i + \beta_{23} AGEPC13_17ONLY_i + \\ & \beta_{24} AGEPCLT6_6_12ONLY_i + \beta_{25} AGEPCLT6_13_17ONLY_i + \\ & \beta_{26} AGEPC6_12AND13_17ONLY_i + \beta_{27} AGEPCLT6_6_12AND13_17_i + \\ & \beta_{28} MHONLY_i + \beta_{29} FHONLY_i + \beta_{30} INCOME_i + \alpha_i IMR + \varepsilon_i \end{aligned} \quad (17)$$

where $i = 1, \dots, n$ is the number of observations (households in our work) in the model. q_i corresponds to the quantity of purchase of chocolate milk and P_i variable represent the price of chocolate milk. We have defined the variables in the above equation 17 in Table 1. In the equation 17, IMR stands for the inverse Mills ratio and α_i corresponds to the coefficient associated with IMR . Presence of sample selection bias is determined looking at the

significance of α_i . If we have sample selection bias, we have to do an adjustment to the coefficient estimates in the second stage estimation in trying to get at correct marginal effects. Procedure to adjust for marginal effects was elaborated in the preceding section.

As such, we will calculate 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.

Results and Discussion

Market penetration for chocolate milk is 25 percent. The average at-home quantity of chocolate milk consumed is 404 ounces per household per year and the average price is \$0.04 per ounce. Factors affecting the probability of purchase of chocolate milk (the decision to buy) are, price of chocolate milk, household income, age of household head, education status of household head, region, race, Hispanic household head, age and presence of children, and gender of household head. The factors affecting the volume of purchase of chocolate milk are price of chocolate milk, household income, age of household head, education status of the household head, region, race, Hispanic household head, age and presence of children in the household, and gender of household head. The own-price elasticity of demand for chocolate milk was estimated to be -0.04. Sample selection bias was statistically significant.

References:

Beverage Marketing Corporation, 2010

Brandweek. (2010). "Got Chocolate Milk? Apolo Ohno Does". www.brandweek.com
(internet accessed on August 13, 2010)

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

Karp, J.S., J.D. Johnston, S. Tecklenburg, T.M. Mickleborough, A.D. Fly, and J.M. Stager.
(2006). "Chocolate Milk as a Post-Exercise Recovery Aid". *International Journal
of Sport Nutrition and Exercise Metabolism* 16: 78-91

Nielsen. (2010). www.nielsen.com (internet accessed on September 13, 2010)

NPD Group. (2010). www.npd.com (internet accessed on September 13, 2010)

Thomas, K., P. Morris, and Emma Stevenson. (2009). "Improved Endurance Capacity
following Chocolate Milk Consumption compared with 2 Commercially Available
Sport Drinks". *Applied Physiological Nutrition Metabolism* 34: 78-82

Saha, A., O.Capps. Jr, et al. (1997). "Calculating Marginal Effects in Dichotomous-
Continuous Models." *Applied Economic Letters* 4(3): 181-185.

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

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

Variable	Explanation
PRICE	Price of Chocolate Milk
<i>AGEHHLT25</i>	<i>Age of Household Head less than 25 years (Base category)</i>
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
<i>EMPHHNF</i>	<i>Household Head not employed for full pay (Base category)</i>
EMPHHPT	Household Head Part-time Employed
EMPHHFT	household Head Full-time Employed
<i>EDUHHLTHS</i>	<i>Education of Household Head: Less than high school (Base category)</i>
EDUHHHS	Education of Household Head: High school only
EDUHHU	Education of Household Head: Undergraduate only
EDUHHPC	Education of Household Head: Some post-college
<i>EAST</i>	<i>Region: East (Base category)</i>
MIDWEST	Region: Central (Midwest)
SOUTH	Region South
WEST	Region West
<i>WHITE</i>	<i>Race White (Base category)</i>
BLACK	Race Black
ASIAN	Race Oriental
RACE_OTHER	Race Other (non-Black, non-White, non-Oriental)
<i>HISP_NO</i>	<i>Non-Hispanic Ethnicity (Base category)</i>
HISP_YES	Hispanic Ethnicity

Table 1 Continued....

Variable	Explanation
<i>NPCLT_18</i>	<i>No Child less than 18 years (Base category)</i>
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
<i>FHMH</i>	<i>Household Head both Male and Female (Base category)</i>
MHONLY	Household Head Male only
FHONLY	Household Head Female only