Consumer Choice of Private Label or National Brand: The case of organic and non-organic milk

Yan Zhuang
Graduate Research Assistant, Penn State University

Carolyn Dimitri
USDA/Economic Research Service

Edward C. Jaenicke
Associate Professor
Department of Agricultural Economics and Rural Sociology
Penn State University
208A Armsby
University Park, PA 16802
Tel: (814) 865-5282
Email: tjaenicke@psu.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association, Milwaukee, Wisconsin, July 26–28, 2009.

This project was supported by the National Research Initiative of the Cooperative State Research, Education and Extension Service, USDA, Grant # 2007-04473.

Introduction

Since the 1970s and 1980s, private label products (also known as store brands) have seen great improvements in product quality and large gains in market share. Once considered a low-quality, low-price alternative, some private label products evolved to compete with high-quality, market-leading brands, including organic brands (Burt, 2000). Citing a study in 2006 produced by the Private Label Manufacturers Association (PLMA), Haberkorn (2006) notes that 41 percent of shoppers buy private label goods frequently, up from 36 percent in 2001 and 12 percent in 1991. In 2000, the market share of private label brands exceeded most national brands in about 50 percent of categories, ranking first or second in 131 out of 266 product categories (German, 2001).

Dairy is a grocery category where private label has one of the largest market shares. For the 52 weeks ending May 19, 2007, total private label sales across all categories reached $46.5 billion. Among all categories, private label milk led the way with $6.5 billion, followed by bread and baked goods with $3.4 billion, and cheese with $2.9 billion (Progressive Grocer, 11/2007). Citing Information Resources Inc., Barstow (2005) claims that about 60 percent of milk is sold under a store brand. Bonanno and Lopez (2005) say that the expansion of private label in food industry has altered the competition between retailers and manufacturers over the last two decades.

These private label trends extend to the organic food market, which itself has rapidly grown annually since early 1990s, and now has a 2007 growth rate of 19 percent. Global organic food market is expected to reach $70.2 billion by the end of 2010 (Research and

Markets, 2008). According to the trade journal *Gourmet Retailer* (2008), private labels are responsible for 17.4 percent of all organic sales, with dairy and produce items having the highest shares. In the market for organic milk, two national brand milk companies, Organic Valley and Horizon Organic, have led the market from the late 1980s. As of May 2007, these two producers provide 75 percent of U.S. organic milk supply (Schultz, 2008). After these two brands, private label organic milk occupies third position nationally, comprising just under 10 percent of the market share (Ihde, 2002).

Most research that investigates the private label versus national brand choice focuses on consumers' demographics and perceptions. Richardson, Jain, and Dick (1996) provide an extensive list of factors that affect consumers' private label choice: 1) Demographic variables, such income and family size, where lower income and larger family size households are more likely to buy private label brands; 2) Extrinsic cues, such as name, price and package, where better extrinsic cues increase the likelihood of purchase; 3) Perceived factors, such as the perceived value for money, risk, and quality variation, where the perceived value for money measured as the ratio of quality and price has a positive relationship with private label preference; 4) Former experience, such as familiarity with store brand, where more familiarity with private label means lower perceived risk and quality variation associated with private label, which makes consumers less dependent on extrinsic cues, hence higher private label preference.

Batra and Sinha (2000) specifically study how perceived risk affects the success of private label brands. After assessing the purchasing preferences for national brands versus

private label brands across twelve different categories, they find that if the cost of making a mistake in a category is low, consumers are more likely to choose private label brands. If the category has more "experience" than "search" characteristics, consumers are likely to favor national brands over private label brands, because consumers can compare functional attributes by "search" characteristics, while cannot know the true quality of "experience" product unless actually use it.

To the best of our knowledge, there are no studies about individual level consumer choice that examines the potential differences between private label and national brand organic product purchase. Our study investigates the private label decision for organic milk, where the market share for private label organic milk is growing fast. With this growing market share – both for organic food generally and organic private label milk specifically – as background, our paper investigates whether there exist differences in the way organic and non-organic milk buyers approach the decision to purchase branded or private label milk. While the consumer's choice can be partially decomposed into two related decisions, one on organic and another on private label, an accurate investigation of this question must account for potential selection effects that separate organic from non-organic milk buyers. Therefore, this paper estimates two sample selection models. In the first stage, milk consumers decide whether or not to buy organic. Relative prices, promotional variables, consumption patterns, and demographic factors are assumed to influence this first-stage decision. In the second stage, consumption patterns, promotions, demographics and a different set of relative prices are assumed to influence the private-label or brand choice, conditional on the outcome of the first stage.

One of the first steps in our investigation is to define the decision variable that identifies whether branded or private label, and organic or non-organic milk was purchased. Because this variable needs to accommodate the fact that consumers might buy more than one type of milk on each trip, we define a variable that identifies a household's most prevalent milk category each week. This variable falls in one of four categories: organic private label, non-organic private label, organic national brand, and non-organic national brand.

Using this definition for weekly "main" milk purchases, our results show that a large number of factors affect the first-stage decision of whether a household buys organic or non-organic milk. Relative prices, promotional variables, consumption patterns, and demographic factors are found to significantly affect both the decision to buy organic or non-organic milk, and the conditional decision to buy private label or branded milk. In the second stage, however, when it comes to the private label or national brand decision, we find that most but not all of these factors influence organic and non-organic consumers in the same general way. Two of the differences we do find are that both the presence of children in the household and marriage make private label purchases less likely for organic consumers but more likely for non-organic consumers. We also find that cents-off coupons significantly increase national brand non-organic purchases but not organic purchase.[1] These and other results along with the econometric model are discussed in more detail below.

[1] We find this result despite the fact that not all states allow cents off coupons

Model Development and Specification

We assume that consumers weigh the organic/non-organic and the private label/national brand choice in a way that is consistent with utility maximization, but that an analyst may not be able to observe all elements that influence utility. Random utility models, widely used in consumer choice models, incorporate uncertainty into utility models by allowing decision-makers but not analysts to have complete information about the consumer's utility function. Uncertainty may come from unobserved product attributes or unobserved consumer heterogeneity. A stochastic error term is modeled in utility function to reflect uncertainty. Utility of a consumer i choosing alternative $\mathcal{X}$ is $U_{\mathcal{X}}^{i} = V_{\mathcal{X}}^{i} + \varepsilon_{\mathcal{X}}^{i}$, where $V_{\mathcal{X}}^{i}$ is the determinant part, and $\varepsilon_{\mathcal{X}}^{i}$is the stochastic part. A decision maker maximizes his or her utility, so the probability of alternative $\mathcal{X}$ being chosen by consumer i is $P_A^i(\mathcal{X}) = P\left(U_{\mathcal{X}}^{i} = \max_{z \in A} U_z^i\right)$.

Random utility-based consumer choice models using scanner data have often estimated by logit or probit regressions. Following McFadden's (1974) multinomial logit model, the random utility framework was first used to scanner data brand choice model together with multinomial logit estimation by Guadagni and Little (1983). The determinant part of random utility function is often a linear function of observed attributes such as price and income. The stochastic part is an independently double exponential distributed variable. A consumer chooses the alternative with the highest utility, and the probability of choosing alternative k has the multinomial logit form of $p_k = {e^{v_k}}/{\sum_{j \in S_i} e^{v_j}}$, where v is the determinant part. In Guadagni and Little's (1983) paper, a multinomial logit brand choice model estimation on regular ground coffee purchase shows that brand loyalty, size loyalty, store promotion, regular

price and promotional price cut are statistically significant. Since this model calculates brand choice probability given that a purchase has been made, a nested logit model of coffee purchase (Guadagni and Little, 1998) was introduced. Random utility model-based consumer choice models have also incorporated heterogeneity, state dependence, and similar factors (see Keane 1997, Bucklin and Gupta 1992, among others), but the basic framework remains the same. The nature of random utility framework and logit estimation method fits any discrete consumer choice situation satisfying utility maximization. Therefore, we can use random utility framework to analyze consumer choice between organic and non-organic milk, and between private label and national brand milk.

In this paper, we will address the question of whether organic and non-organic milk buyers approach the choice between private label and national brand differently. In order to answer this question, we need to compare the private label versus national brand choice in two separate settings – organic and non-organic. Take organic milk buyers for example: a household chooses organic milk if the utility of organic milk to the household exceeds the utility of non-organic milk. Then, organic milk buyers choose from organic private label milk and organic national brand milk. The process is similar for non-organic milk buyers. However, analysts are not able to observe the actual utility function; only the final choices of milk buyers are observed instead. From the final choice, we can infer which choice brings a household maximum utility. But the observed data is truncated because we can only study how organic milk buyers approach the choice between private label and national brand when a household chooses organic milk, and similarly, we can only study how non-organic milk buyers approach the choice between private label and national brand when a household chooses non-organic

milk. One way to solve this problem, used in this paper, is sample selection framework. By using a sample selection framework, we can account for the hazard of a household choosing organic milk or non-organic milk, and the joint distribution of disturbances which may occur from sample selection bias (see, for example, Stordal, Lien and Baardsen, 2008).

To study whether organic and non-organic milk buyers approach the choice between private label and national brand differently, we therefore need to compare the private label versus national brand choice in two separate settings. Hence the sample selection model will be used twice, once for organic milk and the other for non-organic milk. Since the model specification and calibration are very similar for the two settings, we will focus on private label versus national brand choice for organic buyers in the following discusson. Private label versus national brand choice for non-organic buyers can be derived in the same way.

The choice between organic milk and non-organic milk

In each weekly time period, we assume a household's main milk purchase decision can be split into two stages. In the first stage, households decide to buy mainly organic milk or non-organic milk.[2] In the second stage, they then choose to buy mainly private label milk or national brand milk. Let U_{ijt} denote the utility to household *i* of purchasing *j* at time *t*, where *j* = organic or non-organic milk. This utility will depend on observed and unobserved characteristics of brands, observed and unobserved characteristics of households.

$$U_{ijt} = V_{ijt} + \varepsilon_{ijt},$$

[2] We use the word "mainly" here to (i) acknowledge that households might buy more than one type of milk in a single week (e.g., some private label organic milk and some private label non-organic milk. Our study focuses on each week's main purchase, which we define as the one category out of four with the highest expenditure.

where V_{ijt} is the determinant part of household *i*'s utility, calculated from observed variables, and ε_{ijt} is the random part of household *i*'s utility, capturing unobserved variables. Household *i* makes decision to maximize utility, so the probability of choosing organic milk at time *t* is $P_{iOt} = P(U_{iOt} \geq U_{iNOt})$, where U_{iOt} is the utility of household I choosing organic milk at time *t*, and U_{iNOt} is the utility of household *i* choosing non-organic milk at time *t*. Similarly, the probability of choosing non-organic milk at time *t* is $P_{iNOt} = P(U_{iNOt} \geq U_{iOt})$. Following Guadagni and Little (1983), the deterministic component of a household's utility for alternative *j* is expressed as a linear function of observed variables, including attributes of the product (e.g., price, coupon) and attributes of the household (e.g., income, age, or education). In general

$$V_{ijt} = X_{ijt}\beta \quad j = \text{Organic or Non-organic,}$$

where X_{ijt} is a vector of observed characteristics and attributes of product of household *i* at time *t*, and β is a corresponding vector of coefficients capturing how these attributes affect households' evaluation. Under the assumption of an *i.i.d*. error term, the probability of household *i* choosing product *j* at time *t* takes the form of (McFadden 1974):

$$P_{ijt} = \frac{e^{V_{ijt}}}{\sum_j e^{V_{ijt}}}$$

In practice, we observe households' actual choice and attributes values instead of utility and probability. So the data consist of observed choices, where O_{it} denotes the observed choice between organic milk and non-organic milk:

$$O_{it} = \begin{cases} 1 & If\ household\ i\ chooses\ organic\ milk\ in\ week\ t \\ 0 & Otherwise \end{cases}$$

In the first step, household *i*'s decision of whether or not to buy organic milk can be expressed with the latent variable O_{it}^*,

$$O_{it}^* = U_{iOt} - U_{iNOt} ,$$

where U_{iOt} is the utility for household *i* to choose organic milk at time *t*, and U_{iNOt} is the utility for household *i* to choose non-organic milk at time *t*. If the utility of choosing organic milk at time *t* is greater than choosing non-organic milk for household *i*, O_{it}^* is positive, and O_{it} equals 1. Otherwise, O_{it}^* is negative, and O_{it} equals 0. Since the sign of O_{it}^* is determined by U_{iOt} and U_{iNOt}, and U_{ijt} depends on a vector of household and product characteristics, O_{it}^* is determined by a vector of household and product characteristics. The relationship between the latent utility function and observed decision is:

$$O_{it} = \begin{cases} 1 \;\; if \; O_{it}^* > 0 \\ 0 \;\; Otherwise \end{cases} ,$$

where

(1) $$O_{it} = X_{it}\beta + \mu_{it} ,$$

where X_{it} is the vector consisted of observed household *i*'s characteristics and observed product attributes available to household *i* at time *t*, and β is a vector a coefficients to be estimated.

Let Φ denote the cumulative density function of the standard normal distribution, then

$$Prob(O_{it} = 1|X_{it}) = \Phi(X_{it}\beta)$$

$$Prob(O_{it} = 0|X_{it}) = 1 - \Phi(X_{it}\beta)$$

Therefore, in the first step, we can use a probit model to estimate coefficient vector β, and find out how the factors in X_{it} influence consumers' choice between organic milk and non-organic milk.

The choice between private label milk and national brand milk

In order to find out whether organic and non-organic milk buyers approach the choice between private label and national brand differently, we need to know how organic buyers and non-organic buyers choose between private label and national brand milk separately. Sample selection bias appears because a household can only choose from organic private label milk and organic national brand milk if this household is an organic milk buyer. Similarly, a household can only choose from non-organic private label milk and non-organic national brand milk if this household is a non-organic milk buyer.

To correct for sample selection bias, we use Heckman's two step sample selection model. In the first step, as discussed above, a Probit model of whether a household is an organic buyer is estimated. Based on the result of estimation, an inverse Mill's ratio can be calculated. In the second step, this inverse Mill's ratio is included in the independent variables to count for the hazard of not being selected. As mentioned above, the model specification and calibration are very similar for organic and non-organic milk settings, so we will focus on private label versus national brand choice for organic buyers in the second step, private label versus national brand choice for non-organic buyers can be derived in the same way.

The data have information of whether household *i* chooses private label milk at time *t*, let

$$C_{it} = \begin{cases} 1 & If\ household\ i\ chooses\ private\ label\ milk\ in\ week\ t \\ 0 & Otherwise \end{cases}$$

and

$$C_{it}^{O} = \begin{cases} 1 & If\ household\ i\ chooses\ organic\ private\ label\ milk\ in\ week\ t \\ 0 & Otherwise \end{cases}$$

Then C_{it}^{O} can only be observed if household *i* buys organic milk in week *t*. Since the choice between organic private label milk and organic national brand milk is not available for those households that do not choose organic milk, the model does not account for selection bias will produce biased estimation. Therefore, to get the unbiased estimation, we need to identify $Prob(C_{it}^{O} = 1|X_{it}, O_{it} = 1)$ and $Prob(O_{it} = 1|X_{it})$, not $Prob(C_{it}^{O} = 1|X_{it})$.

The choice between organic private label milk and organic national brand milk is also based on random utility framework, so, like O_{it}, C_{it}^{O} also has a continuous latent variable $C_{it}^{O^*} = U_{iOt}^{PL} - U_{iOt}^{NB}$, where U_{iOt}^{PL} is the utility of organic buyer *i* to choose private label milk at time *t*, and U_{iOt}^{NB} is the utility of organic buyer *i* to choose national brand milk at time *t*. C_{it}^{O} satisfies

$$C_{it}^{O} = \begin{cases} 1 & if\ C_{it}^{O^*} > 0 \\ 0 & Otherwise \end{cases}$$

The sign of $C_{it}^{O^*}$depends on households and product attributes, as well as the probability of a household not buying organic milk:

(2) $$C_{it}^{O} = X_{it}\gamma + \theta M_{it} + \varepsilon_{it}\ ,$$

where X_{it} are observable characteristics of households and products, and inverse Mill's ratio $M_{it} = \phi(X_{it}\beta)/\Phi(X_{it}\beta)$ is calculated by the estimation results from the first step. ϕ and Φ

denote the probability density function and cumulative density function of the standard normal distribution respectively. $Prob\left(C_{it}^{O}=1\middle|X_{it},O_{it}=1\right)=\Phi(X_{it}\gamma+\theta M_{it})$ can be estimated using probit estimation. Heckman's model is consistent but not efficient, so a robust procedure correcting for heterogeneity will be used to get more efficient standard errors.

Data

This study uses Nielsen Homescan data, which are collected from individual households. It provides market-related information such as purchase date, dollars paid, promotion type, and brand information. It also provides demographic information, including household size, education, age, race, and much more. We specifically use data on milk for all U.S. markets in 2004, 2005, and 2006. We use a week as discrete time interval because most households do grocery shopping each week. Because a household may buy different types of milk in one week, we loosely follow Rhee and Bell (2002) and define "main milk category" that captures the highest expense in one of four milk categories.[3] To get the "main milk category", expenditures on organic private label milk, organic national brand milk, non-organic private label milk and non-organic national brand milk are calculated, and the category with the highest expenditure is defined as the main milk category. After aggregating the data set on a weekly basis, we have 283,728 weekly trips with milk purchases. Among these trips, 67.35 percent (191,103 weekly trips) match with a main purchase of private label milk, and 32.65 percent (92,625 weekly trips) match with a main purchase of national brand milk. Within the four milk categories, non-

[3] Rhee and Bell (2002) identify a household's "main" store based on a weekly allocation of expenditure at a number of stores. We identity a main category of milk based on a weekly allocation of expenditure across four milk categories.

organic private label milk has the highest share (64.42 percent), and organic private label milk has the lowest share (2.94 percent). Table 1 shows the frequencies of main milk purchase.

The sample selection model framework is applied to all Nielsen data from 2004 to 2006 after dropping observations with missing county values. As explained below, we calculate county-based prices for each of the four milk categories, and missing county codes necessitate dropping 1,067 observations. Table 2 presents the definitions of the 29 variables used on the analysis.

Demographic variables in the dataset include income, education, household size, race and other descriptors. For the most part, these data are used as is without any transformation of the Nielsen data. In some cases, categorical variables are converted in binary dummy variables; in other cases, some of the categorical variables are combined. For example, the Nielson data divides the ages of both female household heads and male heads into 10 categories, where one category is for no male or female household head, another is for an under 25 year-old head, a third is for over 25 but under 30 year-old head, and so on. We generate a new "maxage" variable to represent the maximum age of household heads. Over 50% of the weekly milk purchases are made by households with heads over 55 years old. Like our new age variable, we also generate a variable representing the maximum education level of household heads. Slightly over 51 percent of weekly milk purchases are made by households with at least college education heads. Age and presence of children information is categorized by nine intervals, from children under age 6 only, to no children under age 18. We make the appearance of children in a family a binary dummy variable, denoting whether or not there is a

child under 18 in a household. Just over 72 percent of weekly milk purchases are from households that do not have any children. Of all the weekly milk purchases, 5.48 percent are households that Nielsen identifies as black, 3.26 percent are from households identified s oriental, and 8.26 percent are from households identified as Hispanic. Another variable, household income, requires a non-trivial transformation. The Nielson data contains a categorical code for a particular income range. To provide a clearer interpretation without losing the underlying information, we transform the categorical variable into income midpoints. The top income category is equal to or more than $100,000 a year, we use $170,000 a year to represent this category.

Table 3 provides a brief summary of how private label or national brand purchases vary with household size. It shows that households with four or fewer members compose 91.7 percent of the weekly milk purchase trips. Small households with one or two members occupy over 60 percent of the weekly milk purchase trips. Measured by milk purchase trips, private label purchases exceed national brand purchases across all household sizes.

In addition to demographic information on the purchasing household, the Nielsen data provide some detailed information on the actual purchase event. For example, the Nielsen data contain a code for a promotion type used in the purchase. One of the codes is for coupons, and by interacting this variable with the brand information, we can tell if the coupon was issued by a manufacturer or a retailer and if it was on a private label or national brand.[4] Keep in mind, however, that Nielsen's coupon variable is only observable on the coupon's redemption, not it

[4] Some states have regulations and laws that prohibit pricing milk below cost. In some cases, these regulations and laws may prohibit the use of coupons on milk products.

issuance. Another code is for other store promotions that include special store features, trials and displays. Ultimately, we create two binary dummy variables, one for coupon use, and the other for other promotion use. These dummy variables allow for accessibility to coupons and their redemption may to potentially influence the milk purchase choice. In our data, there are 6,463 coupon redemption observations (which represents 2.29 percent of the observations) and 46,451 other observations of promotion use (which represents 16.43 pervent of the observations).

A major task associated with using the Neilson Homescan data is the construction of weekly price vectors facing each household because only the price on items bought are observed. Some prices must be inferred. Because we assume that households usually do grocery shopping in a certain area near their home, we calculate weekly category prices by county. The algorithm is as follows: 1) we calculate the realized unit price of each milk purchase; 2) For each category, we calculate the mean price by county for each week. In this way, we obtain a county-based organic private label price, organic national brand price, non-organic private label price and non-organic national brand price for each week. By a similar algorithm, we can get county-based prices for more aggregate categories of organic and non-organic milk (combining private label and national brand).

At least two empirical issues complicate the construction of the price vectors. First, because organic private label milk only has a small market share, some counties have no purchase records in certain weeks; hence, prices are missing for those weeks. Previous studies (such as Keane 1997 and Gupta 1988) have used prices in adjacent weeks to approximate the

missing prices. Because of the large number of observations in the Neilson data, we choose instead to eliminate the observations with missing prices. A second issue concerning prices is the choice between the shelf price and the realized price. The shelf price is the listed price in store and includes most price deductions except for coupons or other promotions that are deducted at the register. On the other hand, the realized price is the final price households paid for the purchase and this price accounts for coupon use or any other register-based promotions. Coupon availability, which we do not observe, affects this decision. Because we assume that households live in the same county have similar access to coupons, and because we construct county-based prices, we believe that a realized price (rather than a shelf price) provides an accurate representation of the actual price.

Table 4 summarizes the average annual prices of the four milk categories. It shows that organic milk prices increase and non-organic milk prices decrease over the years. Furthermore, in each year the private label categories have lower prices than national brand for both organic and non-organic milk, as expected.

Finally, we want our two-stage model to account for consumers' shopping habits that might affect the milk purchase choice. We therefore calculate households' weekly expenditure on total dairy products, total non-milk organic dairy, and total private label dairy. We also utilized data that contains consumers' shopping behavior in the fresh produce and meat departments from 2004 to 2006 and include total weekly expenditure, total weekly organic expenditure, and total weekly private label expenditures in fresh produce and meat categories.

Results

We use a two-step procedure to estimate Probits specified in (1) and (2). The estimates of the two-stage decisions of the organic and non-organic private label and national brand milk decisions are shown in Tables 5 and 6, respectively. In both tables, the coefficients on the left are first-stage results of households' propensity to buy organic or non-organic milk, and the coefficients on the right are second-stage results about of households' choice between private label milk and national brand milk, conditional on the decision made in the first stage. Comparing the first-stage estimates in both tables is trivial: the results are by design identical except reversed in sign. On the other hand, comparing the first- and second-step results on each table is nontrivial and shows how identical factors may influence the organic decision and the private label decision in different ways. For the most part, separate examinations of Tables 5 and 6 show that, for the large majority of cases, individual factors influence the two decisions in similar ways. A second nontrivial comparison, and the main focus of our investigation, involves comparing the second-stage results in Table 5 against the second-stage results in Table 6. While these results show more similarities than differences, a few important differences emerge as factors influencing the private label decision. These results are discussed in more detail next.

Results from the first-stage choice

Upon examination of Table 5 (and trivially Table 6), results from first stage show that demographic variables play an important role in affecting a household's choice between organic and non-organic milk. The following factors are shown to significantly increase the probability of a household choosing organic milk as their main weekly purchase: higher income, better education, having children at home, being Oriental or Hispanic, having a male head

working 30-34 hours per week, having a female head working under 30 hours per week, and spending more total weekly expenditure on dairy, non-milk dairy, private label fresh produce and meat, and organic fresh produce and meat. Conversely, an older household head, a larger family size, being Black, having a male head working under 30 hours per week or over 35 hours per week, having no female head or a female head working over 35 hours per week, and spending more on private label dairy or on fresh produce and meat will decrease the probability of a household choosing organic milk. One interesting finding is that the employment status of male head and female head affects the probability of choosing organic milk in different ways. This result might be explained by the distinct roles of male and female head in milk purchase decision making in a household.

All marketing variables significantly affect the choice between organic and non-organic milk. As we expected, a higher non-organic milk price makes a household more likely to buy organic milk, and a higher organic milk price makes a household more likely to buy non-organic milk. Households that use coupons are more likely to buy organic milk, while households that use other store-based promotions are more likely to buy non-organic milk.

Results from the second-stage: private label milk vs. national brand milk

Examining the second-step results in detail allows us to compare the ways that organic and non-organic milk buyers approach the choice between private label and national brand milk. From the second-step estimates in Table 5 and Table 6, one can see that some demographic and marketing variables affect organic and non-organic milk buyers in a similar way:

1) For both organic buyers and non-organic buyers, higher income and better educated household heads make a household more likely to buy private label milk, while older household head and larger household size make a household less likely to buy private label milk.
2) Also for both organic and non-organic buyers, Black households are more likely to buy a national brand, and Hispanic households are more likely to buy private label milk.
3) A household with a male head working 30-34 hours per week is more likely to buy private label milk. Alternatively, a household without a female head, or with a female head working over 35 hours per week is more likely to buy a national brand milk.
4) Increasing the price of private label milk will make consumers more likely to choose national brand milk, as we expected. Non-coupon store promotions, such as store features and displays, increase the probability of choosing private label milk. This result makes sense as retailers may have greater motivation to promote private label brands.
5) The significantly positive sign of time trend variable "week" for both organic and non-organic milk buyers shows that private label milk purchases have increased over time, all else equal.
6) All the shopping behavior or habit variables affect the choice between private label and national brand similarly for organic and non-organic buyers. Households that spend more on dairy, non-milk organic dairy, organic fresh produce and meat, and private label fresh produce and meat are more likely to buy private label milk. Households with more total expenditure on fresh produce and meat are less likely to buy private label milk.

The above similarities are probably due to the very general differences between private label and national brands for milk. On the other hand, the following distinctions may stem from the specific properties related to organic and non-organic milk.

1) Oriental households, households with married heads, and households have children under eighteen are less likely to buy private label milk if they are organic buyers, and more likely to buy private label milk if they are non-organic buyers.
2) Households with no male head are more likely to buy national brand milk if they are organic buyers, and more likely to buy private label if they are non-organic buyers. It is opposite for households with a male head working under 30 hours or over 35 hours per week; i.e., they prefer private label milk if they are organic milk buyers, and they prefer a national brand milk if they are non-organic milk buyers. However, these results are only significant for non-organic milk buyers, not for organic ones. This finding may suggest that employment levels affect loyalty to private label or national brand of non-organic milk buyers, but not of organic milk buyers.
3) Coupon use makes a household is more likely to buy national brand milk in the non-organic case. This effect is not significant for organic milk purchase. This result is may be due to the greater use of coupons for national brands than for private label brands. There are 3,743 observations in our data for national brand coupon redemptions, and 2,766 observations for private label coupon redemption.
4) When the national brand price increases, non-organic buyers are more likely to buy private label milk. This result does not hold for organic milk buyers, thereby implying that organic milk buyers could be less price sensitive and more loyal to private label milk.

5) The significant values of inverse Mill's ratios in both models indicate that the sample selection effect is important.

Conclusion and Discussion

This research is among the first efforts that investigate whether organic and non-organic milk buyers approach the choice between private label and national brand differently. We use a two-stage sample selection estimation procedure to correct for sample selection bias and model the purchase decision in two steps. Households first decide whether to buy organic milk, and then decide whether to buy private label milk. Since previous studies show that demographic variables and marketing variables affect consumers' choice of organic product and private label (Hammarlund, 2002; Thompson & Kidwell, 1998; Batra & Sinha, 2000), we include demographic variables and marketing variables in both the first and second steps. However, we include different relative price variables in the two steps. In the first stage, households compare the aggregated prices of organic milk and non-organic milk. In the second stage, households compare the prices of organic private label milk and organic national brand milk if they choose organic milk in the first step, or compare the prices of non-organic private label milk and non-organic national brand milk if they choose non-organic milk in the first step.

Using the Nielsen Homescan data, we find that most variables affect the choice between private label and national brand similarly for organic and non-organic milk buyers, due to the differences of private label and national brand. Income, education, promotions, and purchase habits affect organic and non-organic milk buyers in similar ways. But there are some differences between the ways of organic and non-organic buyers approaching the choice

between private label and national brand milk, due to the specific property related to organic milk. Marital status, children, employment hours, coupon use and price changes affect organic and non-organic milk buyers differently.

From the point of view of food manufacturers and retailers, the results of this paper help managers understand who is buying organic private label milk, and how marketing actions (prices, coupons, etc.) affect consumers' decision making. Although this paper focuses on milk purchases, this sample selection method can be extended to organic private label purchase of other categories. Managers can use the results to design marketing strategies focusing on organic and non-organic buyers respectively.

Although this paper gives promising results of how organic and non-organic milk buyers approach the choice between private label and national brand differently, there are some limitations that can be improved on with future work. Advertising affects consumers' choice, but there is no advertising information available in our data set. So future research may incorporate how advertisement affects consumers' choice between private label and national brand. Finally, since the market for private label organic milk is still growing rapidly, the data set is unbalanced. With the development of private label organic milk, more balanced data set may provide better estimation results.

References

Barstow, Cynthia. "Natural and Organic Private Labels Become Recognizable Brand Names." *Natural Grocery Buyer*, no. Fall (2005).

Batra, Rajeev, and Indrajit Sinha. "Consumer-Level Factors Moderating the Success of Private Label Brands." *Journal of Retailing* 76, no. 2 (2000): 175-91.

Bonanno, Alessandro, and Rigoberto A. Lopez. "Private Label Expansion and Supermarket Milk Prices." *Journal of Agricultural & Food Industrial Organization* 3, no. SPECIAL ISSUE: Retailer Strategies in the Food Marketing Chain (2005).

Bucklin, Randolph E., and Sunil Gupta. "Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach." *Journal of Marketing Research* 29 (1992): 201-15.

Burt, Steve. "The Strategic Role of Retail Brands in British Grocery Retailing." *European Journal of Marketing* 34, no. 8 (2000).

Dimitri, Carolyn, and Kathryn M. Venezia. "Retail and Consumer Aspects of the Organic Milk Market." 1-18: Economic Research Service/USDA, 2007.

German, Gene. "Are Consumers Buying More Private Label (or Store Brand) Products?" *Smart Marketing*, no. June (2001).

Guadagni, Peter M., and John D. C. Little. "A Logit Model of Brand Choice Calibrated on Scanner Data." *Marketing Science* 2, no. 3 (1983): 203-38.

———. "When and What to Buy: A Nested Logit Model of Coffee Purchase." *Journal of Forecasting* 27 (1988): 303-26.

Haberkorn, Jen. "Grocers' Private Labels Gain Market Share: Store Brands Getting Upgrade, Going Upscale." *The Washington Times*, November 30, 2006 Thursday 2006, C08.

Hammarlund, Ray. "A Study of Marketing Issues with Organic Milk." Agricultural Marketing Resource Center, 2002.

Ihde, Erin Speiser. *Milking the Organic Market* 25-Nov-2002 2002 [cited. Available from http://www.brandchannel.com/features_effect.asp?pf_id=133.

Keane, Michael P. "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior." *Journal of Business & Economic Statistics* 15, no. 3 (1997): 310-27.

McFadden, Daniel. *Conditional Logit Analysis of Qualitative Choice Behavior*. Edited by Paul Zarembka, *Frontiers in Econometrics*. New York: Academic Press, 1974.

Research&Markets. *Emerging Organic Food Markets* 2008 [cited 10/12 2008]. Available from http://www.researchandmarkets.com/research/9c0e57/emerging_organic_f.

Rhee, Hongjai, and David R. Bell. "The Inter-Store Mobility of Supermarket Shoppers." *Journal of Retailing* 78 (2002): 225-37.

Richardson, Paul S., Arun K. Jain, and Alan Dick. "Household Store Brand Proneness: A Framework." *Journal of Retailing* 72, no. 2 (1996): 159-85.

Schultz, Madeline. *Organic Dairy Profile* May 2008 2008 [cited 2008]. Available from http://www.agmrc.org/commodities__products/livestock/dairy/organic_dairy_profile.cfm.

Størdal, Stale, Gudbrand Lien, and Sjur Baardsen. "Analyzing Determinants of Forest Owners' Decision-Making Using a Sample Selection Framework." *Journal of Forest Economics* 14 (2008): 159-76.

"Store Brands by the Numbers." *Progressive Grocer*, November 14, 2007 2007.

Thompson, Gary D., and Julia Kidwell. "Explaining the Choice of Organic Produce: Cosmetic Defects, Prices, and Consumer Preferences." *American Journal of Agricultural Economics* 80, no. 2 (1998): 277-87.

What's Driving Private Label Growth? Sept 29, 2008 2008 [cited. Available from http://www.gourmetretailer.com/gourmetretailer/content_display/news/e3if965d7b68052140232124b92b509326e.

Table 1: Frequency Main Milk Purchase Categories by Weekly Trips

	Non-Organic	Organic	Total
National Brand	60,723 (21.4%)	31,902 (11.2%)	92,625
Private Label	182,775 (64.4%)	8,328 (2.9%)	191,103
Total	243,498	40,230	283,728

Table 2: Variable Descriptions

Variable Name	Description
Private Label = 1	Dummy variable for private label. Equals 1 if the product is private label milk, 0 otherwise
Organic = 1	Dummy variable for organic. Equals 1 if the product is organic milk, 0 otherwise
Non-organic = 1	Dummy variable for non-organic. Equals 1 if the product is non-organic milk, 0 otherwise
income	Household income
maxage	The maximum age of male and female household head
dumedu	Equals 1 if at least one household head has at least college education
hhsize	Household size
black	Equals 1 if the household head is black
oriental	Equals 1 if the household head is oriental
hispanic	Equals 1 if the household head is Hispanic
children	Equals 1 if the household has children under 18
_lmemp_0	Equals 1 if there is no male head in the household
_lmemp_1	Equals 1 if male head works under 30 hours per week
_lmemp_2	Equals 1 if male head works 30-34 hours per week
_lmemp_3	Equals 1 if male head works over 35 hours per week
_lfemp_0	Equals 1 if there is no female head in the household
_lfemp_1	Equals 1 if female head works under 30 hours per week
_lfemp_2	Equals 1 if female head works 30-34 hours per week
_lfemp_3	Equals 1 if female head works over 35 hours per week
married	Equals 1 if household heads are married, 0 otherwise
dumcpn	Equals 1 if coupon is used, 0 otherwise
otherpro	Equals 1 if other promotion is used, 0 otherwise
logorgprice	Logrithm value of organic milk price
lognorgprice	Logrithm value of non-organic milk price
logorgctlprice	Logrithm value of organic private label milk price
logorgnctlprice	Logrithm value of organic national brand milk price
lognorgctlprice	Logrithm value of non-organic private label milk price
lognorgnctlprice	Logrithm value of non-organic national brand milk price
smdairy	Total weekly expenditure on dairy
smorgdairynomilk	Total weekly expenditure on non-milk organic dairy
smctldairy	Total weekly expenditure on private label dairy
smfpm	Total weekly expenditure on fresh product and meat
smorgfpm	Total weekly expenditure on organic fresh product and meat
smctlfpm	Total weekly expenditure on private label fresh product and meat

Table 3: Household Size Distribution

HHSize	National Brand	Private Label	Freq.	Freq. Percent
1	22,676	35,700	58,376	20.65
2	38,524	78,576	117,100	41.43
3	14,403	30,511	44,914	15.89
4	11,159	27,737	38,896	13.76
5	3,749	11,918	15,667	5.54
6	1,189	3,655	4,844	1.71
7	365	1,507	1,872	0.66
8	227	391	618	0.22
9	82	292	374	0.13
Total	92,374	190,287	282,661	100

Table 4: Average annual prices for four milk categories based on weekly trips

	2004	2005	2006
Frequency	63,608	90,923	129,197
Organic Private Label Price ($/gallon)	6.22	6.90	7.29
Organic National Brand Price ($/gallon)	6.64	6.93	7.47
Non-Organic Private Label Price ($/gallon)	3.70	3.51	3.31
Non-Organic National Brand Price ($/gallon)	4.59	4.53	4.43

Table 5:Choice Between Organic Private Label and Organic National Brand Milk

First Stage			**Second Stage**		
Organic = 1	Coef.	Robust Std. Err.	Private Label = 1	Coef.	Robust Std. Err.
income	0.0027518***	0.0000827	income	0.005668***	0.0002758
maxage	-0.1185757***	0.0023374	maxage	-0.2246079***	0.0082826
dumedu	0.3001549***	0.0079063	dumedu	0.5117916***	0.030275
hhsize	-0.1449804***	0.0049801	hhsize	-0.2051234***	0.0185243
black	-0.0661094***	0.0158155	black	-0.2849562***	0.0472835
oriental	0.0489412***	0.0177933	oriental	-0.074489**	0.0414186
hispanic	0.0866916***	0.0126914	hispanic	0.1999453***	0.0357509
children	0.0482622***	0.0122309	children	-0.0511656	0.0353719
_lmemp_0	0.010892	0.0199516	_lmemp_0	-0.0619245	0.0592565
_lmemp_1	-0.0668597***	0.0223943	_lmemp_1	0.0891758	0.0664482
_lmemp_2	0.1369785***	0.0247081	_lmemp_2	0.2614156***	0.0761049
_lmemp_3	-0.0560067***	0.0109137	_lmemp_3	0.0262762	0.0367807
_lfemp_0	-0.3708646***	0.0217256	_lfemp_0	-1.034325***	0.0664925
_lfemp_1	0.1020496***	0.0118098	_lfemp_1	0.0138896	0.0337988
_lfemp_2	0.0250356	0.0170637	_lfemp_2	0.0022501	0.0487058
_lfemp_3	-0.0633198***	0.0087453	_lfemp_3	-0.2808625***	0.0255324
married	-0.0238373	0.0177477	married	-0.1917634***	0.0491187
dumcpn	0.1284482***	0.0212944	dumcpn	0.0580692	0.0620037
otherpro	-0.1249918***	0.010545	otherpro	0.3271655***	0.0312565
logorgprice	-0.2481633***	0.0256232	logorgctlprice	-0.2410777***	0.037153
lognorgprice	0.2526005***	0.0234252	logorgnctlprice	-0.3893459***	0.0885597
week	-0.0008573***	0.0000863	week	0.0005326***	0.0002697
smdairy	0.0359616***	0.0005953	smdairy	0.0110401***	0.0016215
smorgdairynomilk	0.1676145***	0.0046321	smorgdairynomilk	0.1243183***	0.0086568
smctldairy	-0.0869314***	0.0013563	smfpm	-0.0116978***	0.0009999
smfpm	-0.0081195***	0.0003105	smorgfpm	0.0963792***	0.0080788
smorgfpm	0.1305988***	0.0033151	smctlfpm	0.0187486***	0.0039835
smctlfpm	0.0020053***	0.0009515	invmills	2.194453***	0.068567
_cons	0.0209169	0.1168063	_cons	-4.182198***	0.3194223

Note:1) *** means significant at 1%, ** means significant at 5%, and * means significant at 10%.

2) There are 239668 observations in first step and 21365 observations in second step.

Table 6:Choice Between Non-Organic Private Label and Non-Organic National Brand Milk

First Stage			**Second Stage**		
Non-organic = 1	Coef.	Robust Std. Err.	Private Label = 1	Coef.	Robust Std. Err.
income	-0.0027518***	0.0000827	income	0.0078751***	0.0001769
maxage	0.1185757***	0.0023374	maxage	-0.2766826***	0.0062316
dumedu	-0.3001549***	0.0079063	dumedu	0.6429092***	0.0142998
hhsize	0.1449804***	0.0049801	hhsize	-0.2475325***	0.0076613
black	0.0661094***	0.0158155	black	-0.2693811***	0.0148074
oriental	-0.0489412***	0.0177933	oriental	0.1423241***	0.0211785
hispanic	-0.0866916***	0.0126914	hispanic	0.1901876***	0.0145005
children	-0.0482622***	0.0122309	children	0.1577878***	0.0132933
_lmemp_0	-0.010892	0.0199516	_lmemp_0	0.1344143***	0.0187893
_lmemp_1	0.0668597***	0.0223943	_lmemp_1	-0.2113582***	0.0201277
_lmemp_2	-0.1369785***	0.0247081	_lmemp_2	0.3098279***	0.0293107
_lmemp_3	0.0560067***	0.0109137	_lmemp_3	-0.202928***	0.0112938
_lfemp_0	0.3708646***	0.0217256	_lfemp_0	-0.6894784***	0.0243488
_lfemp_1	-0.1020496***	0.0118098	_lfemp_1	0.1591219***	0.0129717
_lfemp_2	-0.0250356	0.0170637	_lfemp_2	-0.0115713	0.0177379
_lfemp_3	0.0633198***	0.0087453	_lfemp_3	-0.0953217***	0.0093897
married	0.0238373	0.0177477	married	0.0644621***	0.01676
dumcpn	-0.1284482***	0.0212944	dumcpn	-0.2785598***	0.0234062
otherpro	0.1249918***	0.010545	otherpro	0.1299951***	0.0117901
logorgprice	0.2481633***	0.0256232	lognorgctlprice	-0.0883397***	0.0223622
lognorgprice	-0.2526005***	0.0234252	lognorgnctlprice	0.3139714***	0.0149381
week	0.0008573***	0.0000863	week	0.0019008***	0.000092
smdairy	-0.0359616***	0.0005953	smdairy	0.0337127***	0.0011104
smorgdairynomilk	-0.1676145***	0.0046321	smorgdairynomilk	0.7562462***	0.0205387
smctldairy	0.0869314***	0.0013563	smfpm	-0.0160527***	0.0003838
smfpm	0.0081195***	0.0003105	smorgfpm	0.4879599***	0.0152591
smorgfpm	-0.1305988***	0.0033151	smctlfpm	0.0214011***	0.0011563
smctlfpm	-0.0020053***	0.0009515	invmills	-7.153551***	0.1331806
_cons	-0.0209169	0.1168063	_cons	4.530659***	0.1411562

Note:1) *** means significant at 1%, ** means significant at 5%, and * means significant at 10%.

2) There are 239668 observations in first step and 193359 observations in second step.