Factors Affecting Consumer Choice and Willingness to Pay for Milk Attributes

Daria J. Bernard

Alan Mathios

Department of Policy Analysis and Management
Cornell University

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Providence, Rhode Island, July 24-27, 2005

Abstract

This study used weekly scanner data to determine within the milk market the factors that affect consumer choice of non-rBST and organic products and the implications for the development of niche markets. This was accomplished by first understanding what product attributes affected demand for milk and then determining how much consumers were willing to pay for these attributes. The former was done within the multinomial logit framework while the latter used the hedonic price model to infer WTP. Results showed the price effect for rBST-free was the largest while the price effect for organic was the smallest and that consumers were willing to pay more for both rBST-free and organic milk compared to conventionally produced milk.

Key words: multinomial logit, scanner data, milk, rBST, organic, hedonic

Introduction

Food choices available to consumers have been changing primarily due to increased organic production and the introduction of genetically modified foods. While these innovations have affected numerous food product categories, this research focused on the influence of these factors on the milk industry. The main change here was the addition of recombinant bovine somatotropin (rBST), with was approved for use in milk production after a lengthy and controversial process by the Food and Drug Administration (FDA) in 1993.

At the time that rBST was approved, there were increasing public policy concerns on issues such as standardization, labeling, health risks, and consumer welfare. Its success therefore would depend not only on the benefits but also the concerns associated with its use. For rBST, the direct benefits were to farmers in increased milk production and feed efficiency, which was expected to result in increased profits. Farmers were concerned, however, about a consequent decline in milk prices resulting from increased surpluses of milk, a problem that had already plagued the industry (Butler, 1999). Although there were no benefits to consumers in terms of improved nutrition or taste, consumers would benefit from lower milk prices as a result of increased milk supply. Consumer concerns included the impact of rBST on small farms, animal welfare, environmental impact, and human food safety.

This appearance of rBST seems to be associated with a rise in organic milk sales, with sales of organic milk and dairy products increasing dramatically after its approval (DuPuis, 2000). While organic foods had been available since the natural foods social movement of the 1960s, the organic milk industry did not exist a decade ago (DuPuis, 2000). In addition, another alternative appeared on supermarket shelves in the form of "rBST-free" milk.

The FDA does not, however, require labeling of milk products from cows treated with rBST. In response to requests the FDA established guidelines in February 1994 for the voluntary labeling of milk and milk products by companies that do not use milk from cows supplemented with rBST. Specifically, the FDA recommended use of the following label: "from cows not treated with rbST" or a similar label. The following disclaimer should also be included: "No significant difference has been shown between milk derived from rbST-treated and non-rbST-treated cows."

The overall objective of this study was to determine within the milk market the factors that affect consumer choice of products with controversial, or misunderstood, attributes and the implications for the development of niche markets based on specific product attributes. In order to accomplish this, it was first necessary to understand what product attributes affected demand for milk and then secondly determine how much consumers were willing to pay, in this case to avoid, milk produced from cows treated with rBST.

In order to determine the factors that affected consumer purchases, the multinomial logit framework was used to derive the log partial odds ratio, which was then estimated as the log of the ratio of units sold. In this model, the ratio of units sold was modeled as a function of product attributes and average store demographic variables. The basic model was expanded to include stockpiling effects and store and time fixed effects. In addition, the models were estimated using alternative items as the reference brand to examine the robustness of the results. For the second goal, the hedonic price model was employed to infer WTP from the data. In this approach, price was modeled as a function of product attributes and the coefficients were interpreted as a measure of the implicit market value of those attributes to consumers.

Background

Recombinant bovine somatotropin (rBST), also referred to as recombinant bovine growth hormone (rBGH), is, in simple terms, a milk production enhancer. Bovine somatotropin (bST) is a protein produced naturally in the anterior pituitary gland of cattle (Blayney and Fallert, 1990). Commercially produced bST is a synthetic hormone based on recombinant DNA (rDNA) technology. Marketed under the trade name of Posilac, Monsanto received approval from the Food and Drug Administration (FDA) for rBST as a new animal drug on November 5, 1993 (Federal Register, 1993). Posilac is a prolonged release product, normally injected into dairy cows every two weeks during the last two-thirds to three-fourths of the cow's lactation period (Dobson, 1997).

The ten-year review process for rBST was more lengthy and extensive than most approval processes. During the approval process studies were conducted to substantiate the claimed effects on milk production (Bauman, 1992). Safety studies were conducted to assess the possible health risks to humans as well as dairy cows (Bauman, 1992; Juskevich and Guyer, 1990). The concern for dairy cows focused on the increased risk of mastitis. The FDA finally concluded that approval would "not have a significant impact on the human environment" (Federal Register, 1993).

Unlike organic foods, which have long existed, the organic milk market did not begin until the 1990s.[1] Organic milk first appeared in conventional supermarkets in 1993 (Dimitri and Greene, 2002) and became the fastest growing organic food segment in the United States. Organic milk sales grew from $16 million in 1996 to $31 million in 1997 (Organic Consumers Association, 2003). Organic dairy sales in supermarkets were increasing 36 percent annually, and in 2000 dairy sales accounted for 11 percent of all organic retail sales (Economic Research Service, 2003). For the 12-month period ending June 2001, milk, half and half, and cream was the second leading category of organic food sales in supermarkets with sales equal to $119,315,772 (Dimitri and Greene, 2002).

What can explain the explosive growth from virtually nothing a little more than a decade ago? Articles in trade journals and mainstream media cite rBST as the one reason for the emergence of the organic milk market (DuPuis, 2000). Consumers have changed their consumption practices because of rBST and are purchasing organic milk (DuPuis, 2000). The author argued that organic milk consumption challenges rBST from a "Not-in-My-Body" or "NIMB" politics of refusal. This study will help determine which of these attributes means the most to consumers.

[1] At the time of this study, organic products were labeled based on the Oregon Tilth's Standards and the California Organic Foods Act of 1990.

Literature Review

Prior to the approval of rBST, much of the research conducted focused on supply side issues including, for example, farmer adoption, effects on herd management techniques and costs of production, effects on dairy industry structure with emphasis on small farms, and effects on total output and market prices for milk. It had been assumed that changes in consumption (quantity demanded) would result solely from price changes induced by outward shifts in the market supply function following the adoption of rBST (McGuirk, Preston and Jones, 1992). Potential shifts in demand were largely ignored.

Consumer surveys conducted in the years shortly before rBST approval documented consumer concern about the use of rBST and consumers' expressed intentions to reduce milk consumption if approved (Aldrich and Blisard, 1998). Two studies conducted then focused on this issue of the possible negative consumption effects due to the approval of rBST: McGuirk, Preston and Jones (1992); and Kaiser, Scherer, and Barbano (1992). Both predicted that reductions in fluid milk purchases as a result of consumption changes if rBST were approved.

A later study by Grobe and Douthitt (1995) examined whether consumers overestimate risk from rBST and what factors contribute to that assessment. In the case of rBST, consumers are presented with potential exposure to an artificial, unfamiliar product controlled by someone else that they fear might have possible delayed health effects, which was considered an involuntary risk. They claim that if consumers hold strong beliefs about untreated milk and think human health risks were not given sufficient weight in the regulatory approval process it will be difficult for manufacturers to reverse the negative image of rBST.

A recent study related to this effort used retail scanner data to measure consumer benefits and valuations of the introduction of rBST-free and organic milk (Dhar and Foltz, 2003). Their analysis was based on weekly data of fluid milk purchases in 12 key US metropolitan markets for the period week starting March 9, 1997 to the week ending February 24, 2002. Starting with retail price differentials they showed that on average, price differences between organic and unlabeled milk were approximately $3.00 per gallon and between rBST-free and unlabeled approximately $1.50 per gallon. According to the study, this represented more than a 100% mark-up for organic milk and 50% for rBST-free milk. Additionally it was noted that over the five years from 1997 to 2002 prices increased by 24% for organic milk, 25% for rBST-free milk, and 13% for unlabeled milk. The authors further explained that the asymmetric pattern of price inflation pushed the price differential between organic and unlabeled from $2.68 to $3.64 per gallon (123% of the unlabeled price) and between rBST-free and unlabeled from $1.42 to $2.10 per gallon (70% of the unlabeled price).

Green and Park (1998) used supermarket scanner data from a grocery retail chain in New York to investigate the effects of seasonality and promotion on fluid milk sales. Elasticities were calculated for fluid milk products differentiated by milkfat content (whole, 2%, 1%, and skim). The authors concluded that seasonality and advertising were significant determinants of retail sales of fluid milk. Own-price elasticities were negative and cross-price elasticities were positive for all milk types and were significant and elastic in the case of 2% milk. Lastly, advertising effects were found to be positive and statistically significant with the response to advertising more pronounced for reduced-fat milk types than for whole milk.

Two studies, conducted by Mathios (1998, 2000) were important here as they both applied the multinomial logit model framework to estimate the log partial odds ratio using supermarket scanner data. The first study (1998) examined the consequences of the implementation of the Nutrition Labeling and Education Act (NLEA) on the market for cooking oil in terms of consumers' choice. The second study (2000) examined the impact of moving from a voluntary to a mandatory labeling regime, as a result of implementation of the NLEA, on consumer product choice for salad dressing.

The relevance of both these to the current effort was in the methodology employed. In both, the econometric specification was motivated by the multinomial logit model, which will be outlined in the subsequent chapter. The models estimated were the log partial odds ratio, referred to as the log of the market share of each specific product in both studies. The dependent variable was the log of the market share of product *i*, in store *j*, during week *t*. In other words, this was the log of the ratio of number of units sold of product *i*, relative to the number of units sold of the base brand in the same supermarket at the same time (Mathios, 2000). The independent variables consisted of product specific variables relevant to the research question being explored and the supermarket demographic variables. Lastly, in order to estimate the models, the reference brand for the dependent variable needed to be selected. The technique employed was to first identify the products that were sold in every supermarket in every time period. From these the product with the greatest number of average units sold was selected as the reference brand. This technique was also used in the current study.

Data

Weekly scanner data were collected for twenty stores from a major regional supermarket chain based in upstate New York. The data encompassed the fourth quarter of 2000 and the first quarter of 2001 (week ending October 7, 2000 through week ending March 31, 2001). The company's annual sales $3.02 billion in 2002; the company employs over 32,000 people. Each store ranges in size from 80,000 to 130,000 square feet and offers more than 60,000 products.

Data were collected for the fluid milk – regular and soy category, which consisted of approximately 275 items. The fluid milk category contained the following subcategories as given by the supermarket: whole milk; low fat and skim; flavored and buttermilk; natural foods – milk; eggnog; milk substitutes and lactaid; and kosher. For each item a commodity code and item description were provided. Aggregate data was collected for dollar sales and units, number of items sold, which represented totals for a 7-day period. Two additional measures were collected: percent sales sold on special and percent units sold on special. Also added were dates for secular and non-secular holidays.

The twenty stores were selected using the summary report of store average demographics (household size, income, education, and age), which was based on shoppers' club card memberships. The summary report represented card member information for a 52-week period ending October 23, 2000. No further updates were received during the period the data was collected.[2] Using the demographic summary, selection was done as to maximize the variation in average education.

[2] A store representative advised that most neighborhoods were rather "stable." Hence, it was assumed that the demographic profile changed very little with each update.

The data set was restricted to half-gallon size containers since both rBST-free and organic milk items were only available in such containers. Specifically there were eight rBST-free milk items and ten organic milk items with positive units sold for at least one supermarket during the twenty-six week period. It was assumed that consumers had preferences for half-gallon milk and it was noted, however, that possible substitutes were being eliminated from the analysis. The final data set therefore consisted of forty milk items; of which six were lactose reduced and sixteen were considered conventional.

The last task was the determination of the reference brand for analysis. The criterion was to select product with the highest average number of units sold for which there were positive sales for each of the twenty stores for all weeks of the data period. This product was Store Brand Fat Free milk in a carton container.

Variable Description

Variables names and definitions were presented in Table 1 with simple statistics reported in Table 2. The subscripts indicated how the variable varied: over product *j*, over supermarket *s*, over time *t*, or a combination. Restricting the data set to include only items with positive units sold resulted in a sample of 8,571 observations where units sold ranged from 1 to 1,190.[3] Retail price, $Price_{jst}$, was not provided but was calculated by dividing dollar sales by units sold and ranged from \$0.69 to \$4.99.

The next three variables, $Sbrand_j$, $Region_j$, and $Nation_j$, were created based on both the item description and company websites, when available, to capture the effect of brand name. The first of these indicated whether or not the item was the store brand. A distinction was then made as to whether the brand was regionally distributed or nationally

[3] A zero could mean that no units were sold that particular week or that the item was not available at the store thus resulting in no units sold.

distributed. The largest portion of the milk sample, 40.36%, was classified as national. Store brand accounted for 34% , and the remaining 29% was classified as regional.

The variables, $Hhsz_s$, Inc_s, $Educ_s$, and Age_s, represented the demographic characteristics available from the shoppers' club summary report. All items for a particular supermarket will have the same value for each of these variables.
Other attribute variables included $Whole_j$, $2Percent_j$, $1Percent_j$, and $Skim_j$, pertaining to fat content. The next three variables, $Conventional_j$, $Rbstfree_j$, and $Organic_j$, pertained to the main attributes of interest. A milk product was classified as being rBST-free if this attribute was specifically noted on the package label. $Conventional_j$, represented products that did not have any information on the package label pertaining to either rBST-free or organic. The variable, $Lactose_j$, represented the lactose-reduced milk items and accounted for 21% of the sample.

The next two variables, $Plain_j$ and $Flavor_j$, indicated whether or not the milk was flavored. The majority of the sample, almost 81%, was classified as plain with the remainder chocolate. Two variables, $Plastic_j$ and $Carton_j$, pertained to the container type. Milk was packaged primarily in either plastic or paper cartons. The majority of the sample, 93%, was classified as available in cartons.

The last variable, $Holiday_t$, was constructed based on a list of secular and non-secular holidays. Fifteen of the twenty-six weeks were identified as having at least one holiday. The variable was interacted with the whole milk attribute to capture possible changes in food purchases around the time of holidays for baking needs.

Methods and Hypotheses

Log Partial Odds Ratio Models

The full model for milk can be written as below, where the β's represented the coefficients on the product attributes and the α's represented the coefficients on the store demographic variables. The attributes of the reference brand were used to determine which of the dummy variables would be omitted to avoid perfect collinearity.

$$
\begin{aligned}
\log\left(\frac{\text{Units}_{jst}}{\text{Units}_{1st}}\right) &= \beta_0 + \beta_1\left(\text{Price}_{jst} - \text{Price}_{1st}\right) + \beta_2\left(\text{Price}_{jst} \times \text{Rbstfree}_j - \text{Price}_{1st} \times \text{Rbstfree}_1\right) \\
&+ \beta_3\left(\text{Price}_{jst} \times \text{Organic}_j - \text{Price}_{1st} \times \text{Organic}_1\right) + \beta_4\left(\text{Nation}_j - \text{Nation}_1\right) \\
&+ \beta_5\left(\text{Regional}_j - \text{Regional}_1\right) + \beta_6\left(\text{Whole}_j - \text{Whole}_1\right) \\
&+ \beta_7\left(\text{Holiday}_t \times \text{Whole}_j - \text{Holiday}_t \times \text{Whole}_1\right) + \beta_8\left(\text{2Percent}_j - \text{2Percent}_1\right) \\
&+ \beta_9\left(\text{1Percent}_j - \text{1Percent}_1\right) + \beta_{10}\left(\text{Rbstfree}_j - \text{Rbstfree}_1\right) + \beta_{11}\left(\text{Organic}_j - \text{Organic}_1\right) \\
&+ \beta_{12}\left(\text{Lactose}_j - \text{Lactose}_1\right) + \beta_{13}\left(\text{Flavor}_j - \text{Flavor}_1\right) + \beta_{14}\left(\text{Plastic}_j - \text{Plastic}_1\right) \\
&+ \alpha_1\left(\text{Hhsz}_s \times \text{Rbstfree}_j - \text{Hhsz}_s \times \text{Rbstfree}_1\right) + \alpha_2\left(\text{Hhsz}_s \times \text{Organic}_j - \text{Hhsz}_s \times \text{Organic}_1\right) \\
&+ \alpha_3\left(\text{Inc}_s \times \text{Rbstfree}_j - \text{Inc}_s \times \text{Rbstfree}_1\right) + \alpha_4\left(\text{Inc}_s \times \text{Organic}_j - \text{Inc}_s \times \text{Organic}_1\right) \\
&+ \alpha_5\left(\text{Educ}_s \times \text{Rbstfree}_j - \text{Educ}_s \times \text{Rbstfree}_1\right) + \alpha_6\left(\text{Educ}_s \times \text{Organic}_j - \text{Educ}_s \times \text{Organic}_1\right) \\
&+ \alpha_7\left(\text{Age}_s \times \text{Rbstfree}_j - \text{Age}_s \times \text{Rbstfree}_1\right) + \alpha_8\left(\text{Age}_s \times \text{Organic}_j - \text{Age}_s \times \text{Organic}_1\right) + \varepsilon
\end{aligned}
$$

The hypothesized signs of the coefficients for this model were presented in Table 3. The characteristics of the reference item were the following: store brand, fat free, conventional, plain flavor, and carton container. Consider first the hypothesized coefficients on the two attributes of interest, rBST-free and organic. It was hypothesized that the market share of each would be less compared to conventional milk products and thus a negative coefficient estimate.

Second, these attribute variables were interacted with price. The signs on the interaction terms themselves were not of interest. The complete price effects were $\beta_1 + \beta_2$ for rBST-free milk and $\beta_1 + \beta_3$ for organic milk. The coefficient on the non-interacted price variable represented the price effect for conventional milk. It was hypothesized that the rBST-free price effect would be greater than the conventional price effect ($\beta_1 + \beta_2 > \beta_1$), and that the organic price effect would be less than the conventional price effect ($\beta_1 + \beta_3 < \beta_1$).

Consider next the signs of the coefficients of the variables for each attribute interacted with the store average shopper club card demographic variables. Interacted first with household size, it was hypothesized that larger households, which may represent the presence of children, would be more likely to purchase rBST-free or organic milk. Therefore, for stores with a higher average household size, the sales ratio for both rBST-free and organic milk was expected to be higher than the sales ratio at stores with lower average household size.

Second, it was hypothesized that households with higher income would be more likely to purchase rBST-free or organic milk, given the price premium. Similarly, then, stores with higher average income were expected to have a higher market share for rBST-free and organic milk compared to stores with lower average income. Next, it was hypothesized that older persons would be less likely to purchase rBST-free or organic milk possibly due to lack of knowledge and understanding regarding the attributes. Again, stores with higher average age were expected to have lower sales ratios of rBST-free and organic milk.

Lastly, the effect of education on the sales ratio or rBST-free and organic was uncertain. It could be argued that more educated persons would be more aware of the issues surrounding the use of rBST and be more willing to buy either of the alternatives. At the same time, more educated persons may have better evaluated the risks associated with use of this particular technology and concluded that conventional milk was not more risky, hence a hypothesized negative effect. Therefore the hypothesized signs of the effect of education interacted with each of the attribute variables were not specified.

The log partial odds equation which formed the basis for the model specifications was a linear function of the parameters and could be estimated using linear regression methods. Each model was estimated in SAS using ordinary least squares (OLS) methods. Tests were conducted for the presence of multicollinearity and heteroskedasticity.

Several extensions of the model were proposed under the objectives for the current study. The first was to examine the effects of stockpiling on quantity sold for each product. Several key results have been reported from published research on the effects of price promotions. The first was that temporary retail price reductions substantially increased sales in the form of a significant short-term sales spike (Blattberg, Briesch, and Fox, 1995). One reason for the sales spike was acceleration of consumer category purchase, also known as stockpiling. In this case, consumers were either purchasing a larger quantity (more than usual) or shortening the interpurchase time (buying earlier than usual). The second reason was due to consumers switching their choice from other brands. It was hypothesized that the perishable nature and the need for refrigeration make it unlikely that consumers would stockpile large amounts due to a price promotion.

In an attempt to capture the possible effect of stockpiling, a variable was created to indicate that the product was on sale in period *t-1* and not on sale in period *t*. A two-period effect was also considered in which the item was on sale in period *t-2* and not on sale in both period *t-1* and *t*. It was hypothesized that a one- and two-period effect was sufficient given the perishable nature.

The next objective involved examining possible fixed effects when estimating a model using panel data. There were several ways the model could be extended. In the original model specification, the constant term represented a single effect for all twenty supermarkets. Following the discussion in Greene (2000), the basic framework was a regression model of the form:

$$y_{it} = \alpha_i + \beta' x_{it} + \varepsilon_{it}.$$

In this equation there are K regressors in x_{it}, not including the constant term. The individual effect is α_i, assumed to be constant over time *t* and specific to the individual cross-sectional unit *i*. The fixed effects approach takes α_i to be a group specific constant term in the regression model. A common formulation of the model assumes that differences across units can be captured in differences in the constant term.

The first fixed effects model therefore included store dummy variables to allow for supermarket heterogeneity. The second fixed effects model estimated included the addition of a time-specific effect to the model with group effects. Again following the discussion in Greene (2000), this could be written as follows:

$$y_{it} = \alpha_i + \gamma_t + \beta' x_{it} + \varepsilon_{it}.$$

The next variation of the model was conducted to test the robustness of the estimation results to changes in the item designated as the reference brand. It was determined to estimate the model for one additional possible reference item. This one had the following characteristics: store brand, 2%, conventional production, plain flavor, and carton container. It was hypothesized that the estimation results would not be significantly different from the initial model as the reference items had similar average units sold and varied only by the fat content.

The final objective was to examine the sensitivity of the results to changes in the choice set as implied by the specification test developed by Hausman and McFadden (1984) to test the application of the IIA assumption. The test was based on eliminating one or more alternatives from the choice set to see if underlying choice behavior from the restricted choice set obeyed the independence from irrelevant alternatives property. It was assumed that the IIA assumption would not hold for the overall model. However, a more interesting question was whether the individual estimates significantly changed for the main specific product attributes being considered. The test statistic given in Hausman and McFadden (1984) was applied to test for significant differences.

Hedonic Analysis

Measuring an actual consumer willingness to pay (WTP) using scanner data was not feasible. The most common approach for inferring consumer WTP was hedonic analysis. With this method, price was modeled as a function of the product's attributes. The regression coefficients were referred to as 'hedonic' prices and could be interpreted as a measure of the implicit market value of those attributes to consumers. From these, it was natural to infer consumers' WTP (Lee and Hatcher, 2001). Here, price was modeled as a function of brand (store, regional, or national), fat content as per the item description (whole, 2%, 1% or skim), production characteristics (conventional, rBST-free, or organic), flavor, container (plastic or carton), and lactose-reduced. The model was:

$$\begin{aligned}\text{Price} &= \beta_0 + \beta_1\text{Sbrand} + \beta_2\text{Region} + \beta_3\text{Whole} \\ &+ \beta_4\text{1Percent} + \beta_5\text{Skim} + \beta_6\text{Rbstfree} + \beta_7\text{Organic} \\ &+ \beta_8\text{Flavor} + \beta_9\text{Carton} + \beta_{10}\text{Lactose} + \varepsilon\end{aligned}$$

where the variables have been previously defined and the base product was a national brand, 2%, conventional, non-flavored, non-lactose-reduced milk in a plastic container.

The hypotheses for the parameter estimates were as follows. It was hypothesized that consumer WTP would be least for the store brand, followed by regional brands, and finally highest for national brands. For the different fat contents, it was expected that WTP would be highest for two percent milk since it has the highest level of consumption. WTP for flavored milk and for lactose-reduced milk was hypothesized higher. With regard to container type, no a priori hypothesis was made, although casual observation suggested increased popularity in plastic containers, which may imply higher WTP.

Lastly, it was important to note that functional form remains an issue when estimating hedonic models. Most studies have presented the results for multiple forms, and this convention was followed here. The two most common forms were the linear and semilog, both of which were estimated and reported in the current study. For the semilog, the dependent variable was replaced with the natural log of price.

Results

Log Partial Odds Ratio Models

The log partial odds ratio was estimated using 8,571 observations for half-gallon milk with units sold greater than zero and the reference brand. The first model estimated contained the attribute variables, and the interaction terms of price × rBST-free and price × organic. These results were presented in Table 4. The complete price effect for rBST-free was the coefficient on price plus the coefficient on price × rBST-free and the complete price effect for organic was the coefficient on price plus the coefficient on price × organic. Thus the complete price effect was -2.2684 for rBST-free and -0.78227 for organic. The coefficient on price of -0.78488 represented conventional milk. In each case, the complete price effect was negative and more importantly the effect for rBST-free was the greatest while the effect for organic was the least.

An important note was that the coefficient on price × organic was not significant. The main concern was the lack of variability of prices for organic milk in the available data set. Of the 1,157 organic milk observations, 965 (83%) had a price of $3.29. The price ranged from $1.99 to $4.99 for the remaining observations. The majority of prices in this range (37 of 42 or 88%) had less than ten observations. The lack of variability was consistent with the high degree of collinearity found amongst price × organic and the organic attribute alone, which most likely contributed to the insignificance of the coefficient estimate for price × organic.

Other models were built up from there. Next, the model was estimated with the store demographic interaction variables. Due to initial problems with multicollinearity, for each demographic variable a new variable indicating the sample store rank was created. The objective of this approach was to increase the variation in values for each of the demographic variable, and the severe multicollinearity was removed.

Tests for the presence of heteroskedasticity were also conducted as the model progressed, and the problem was found to be present. Given the large sample size, it was plausible that although detected, the effect was negligible. Tests suggested no significant change when using the heteroskedasticity-consistent covariance matrix. Variables that were initially insignificant remain so even after testing with the heteroskedasticity-robust standard errors. Therefore, results presented here were based on the uncorrected covariance matrix.

Considering the results from the added demographics, first note the coefficient estimate of the effect of household size was negative and significant. Stores with higher average household size had lower market shares of rBST-free and organic compared to stores with lower average household size. These results were in contrast to the initial hypothesis that the effect would be positive, possibly reflecting concern for the types of food products purchased for children. It may actually be that larger households have less disposable income and their purchases reflect this constraint. In this case, the price premiums for both rBST-free and organic milk may prevent these consumers from choosing either of these alternatives.

This alternative explanation would be consistent with the second finding for both attributes interacted with income, which were positive and significant. Higher store average incomes were associated with a larger share of both rBST-free and organic milk compared to stores with lower average incomes. This result was expected and consistent with previous studies.

An interesting result was that both variables for the attributes interacted with education were negative and significant. No a priori hypothesis of the sign of the effect of education was made as it could be argued as a positive or negative effect. This result indicated that stores with higher education levels of shoppers had lower sales ratios of both rBST-free and organic milk. More education could be associated with less concern over the use of rBST in the production of milk and therefore these consumers most likely purchased conventional milk.

Lastly, age was found to have a significant negative effect on the sales ratio of both rBST-free and organic milk. Stores with higher average age shoppers had lower sales ratio of both types of milk compared to stores with lower average age shoppers. This was consistent with earlier hypotheses.

Next, the model was extended to capture possible effects of weekly, advertised sales. It was hypothesized that there would be a spike in the units sold of the milk item advertised on sale during a particular week. Sales of other milk products that same week may be lower as consumers switch and purchase the sale item. In addition, some consumers may purchase a larger quantity than usual, due to the sale, for future consumption. In this case, it was hypothesized that sales of the item in the week following a sale would be lower.

To analyze the possible effects of stockpiling, two variables were included in the model. The first variable was a dummy variable to capture the effect of a sale in week $t-1$ on the ratio of units sold in the current period t. The second variable was a dummy variable to capture the effect of a sale in week $t-2$ on the ratio of units sold in the current period t. The variables were created with the condition that the item was not on sale in the following weeks for both the one- and two-period effects.

In the case of milk, it was hypothesized that if the item was on sale in week $t-1$ relative to the reference brand, then in week t, the share of that milk item would be lower. It was further assumed that by week $t+1$, the share of that item relative to the reference brand would return to pre-sale levels. This would be due to the perishable nature of milk, the need for refrigeration, and the fact that milk is a necessity.

While it was hypothesized that a two-period lag sale effect should be included in the model, the results showed that the inclusion of this variable did not significantly improve the model. Specifically, the significance of the one-period lag sale effect was lost when the two-period effect was included. This supported the results of the test for inclusion discussed above.

The last two steps in the development of the final model were the addition of store and time fixed effects. The results of the model with these and the other variables discussed above were displayed in Table 5. Beginning by looking at the store fixed effects, the majority were statistically different from Store01 at the 1% level. The exceptions were Store 03, 06, 09, 14, 15, and 18. Differences across stores were anticipated in part given differences in product availability. The majority of stores in the sample, eighteen, were located in New York and of those, seven were located in the Rochester, New York metropolitan area. Differences in product availability possibly represent regional differences.

Following the fixed effects model analysis in Greene (2000), a test for the significance of the store group effects was conducted. The null hypothesis was that the store constant terms included in the model were jointly equal to zero. The F ratio was calculated as given in Greene (2000) and based on this result the null hypothesis was rejected at the 1% level of significance. Therefore, it was concluded that the store dummy variables should be included.

Looking to the time fixed effects, both February and March were statistically different from the reference month, October. This provided some evidence that differences across time existed and may reflect, in part, differences in product availability over the course of the data collection period. Following the analysis of the fixed effects model in Greene (2000), a test for the significance of the time effects was conducted. The null hypothesis, that the monthly constant terms included in the model were jointly equal to zero, was rejected at the 1% level of significance. Therefore, it was concluded that the time dummy variables should be included.

To examine the robustness of the estimates, the milk item with the next highest average number units sold was used as the reference brand. The only difference of note between these results was with regard to education. Using the initial reference brand, the effects of education were negative and not significant at the 1% level. However, when the model was estimated with the alternative reference brand, these variables were negative and significant. While this suggested some sensitivity due to differences in fat content of the reference brand, the differences in price effects for conventional, rBST-free, and organic were minimal and the ranking was unchanged.

Lastly, tests revealed that the IIA property did not hold. This was consistent with expectations, where concerns were that rBST-free and organic milk were most likely not irrelevant alternatives at the aggregate level. In the end, it was determined that this was not of large concern to the results presented.

Hedonic Analysis

Results for the two hedonic price regressions for WTP for milk were presented in Table 6, however this section will focus on the linear model. The results were based on 8,571 observations for half-gallon milk products. The linear model explained slightly more than 91% of the variation in milk prices while the semilog model explained 89% of the variation. Heteroskedasticity, a common problem with hedonic studies, was found.[4] Due to this, the results reported were based on the heteroskedasticity-robust covariance matrix.

---

[4] Multicollinearity, which is also common in hedonic studies, was not found to be a concern.

The hedonic prices, from which WTP was inferred, for nearly all selected milk attributes, were significant at the 1% level. For the most part, the results agreed with hypotheses. The main interests were consumer WTP for rBST-free and organic milk. For both of these, consumers were willing to pay significantly more than for conventionally produced milk ($0.26 more for rBST-free milk and $0.73 more for organic milk). As expected, the premium for organic was larger than that for rBST-free milk. An F-test revealed that this relationship was significant at the 1% level. This showed that consumers were willing to pay an extra $0.47 for the additional attributes associated with organic milk.

The results for the differences in WTP based on brand level were as expected although there was not as large a premium for regional milk compared to store brand as anticipated. WTP was significantly higher for both regional brand ($0.05 more) and national brand ($1.17 more) compared to store brand. The hedonic price for national brand was also significantly more than for regional brands, suggesting a clear pattern of consumer preference. Consumers' willingness to pay a premium for both flavored ($0.53 more) and lactose-reduced ($0.62 more) milk was also expected.

Differences in hedonic prices for fat content were not quite as anticipated. Although 2% was hypothesized to have the highest WTP, results suggested that whole milk had the highest premium of $0.04. Hedonic prices for fat free milk were not significantly different from 2% milk, while there was only weak evidence that WTP for 1% milk was less than that for 2% milk. In all, it appeared that while consumers may prefer different fat contents, these have not translated in the market into different price premiums. Lastly, consumer WTP for milk in cartons was significantly less ($0.15) than for plastic containers. This appeared to support casual observations in the grocery stores where plastic seems to have become the dominant packaging.

Conclusion

The empirical results of the log partial odds ratio models revealed several interesting findings. First for the effects of a sale, it was found that a one-period effect was important. Second it was found that differences in market share existed across stores and over the course of the twenty-six weeks, suggesting that both store and time need to be considered. Third, it was found that the results from the initial models were similar to those estimated using an alternative item as the reference brand. Lastly, as expected for aggregate data, the IIA assumption did not, for the most part, apply.

Summarizing the main results, it was found that the market share for both rBST-free and organic was less than for conventional milk, consistent with initial hypotheses. Calculations of the complete price effect from the full model revealed that the price effect for rBST-free was the largest while the price effect for organic was the smallest. Analysis of the effect of store demographics revealed that both education and age did not have a significant effect on the ratio of units sold when interacted with the organic attribute. Higher store average education levels were associated with a lower sales ratio of rBST-free milk and similarly higher average shopper age was found to be associated with a lower sales ratio of this type of milk. An unexpected result was that household size was found to have a significant negative effect when interacted with both rBST-free and organic. Some of the difficulty in interpreting this result was that the effect captured by household size was ambiguous. As expected given the price premiums for rBST-free and organic milk, it was found that higher levels of income were associated with a higher ratio of units sold for each.

Summarizing the hedonic model, it was found that consumers were willing to pay more for both rBST-free and organic milk compared to conventionally produced milk. As expected, the premium was higher for organic milk compared to rBST-free milk. The difference in premiums for rBST-free and organic was found to be statistically significant. This suggested that consumers were willing to pay more for the additional attributes associated with organic production.

There were two categories of limitations for this study: data concerns and methodology issues. For data, the concern was the length of the sample period. While six months of data were collected, many scanner data studies have used periods of a year or longer. A longer time frame would have allowed a full examination of other seasonal and monthly effects that may affect these markets. In terms of methodology issues, further creation of econometric methods would also alleviate arguably the largest limitation of this study, the difficulty in generating willingness to pay estimates from scanner data. In particular was the lack of a known method of converting the results of the log partial odds models into WTP.

References

Bauman, Dale E., "Bovine Somatotropin: Review of an Emerging Animal Technology," *Journal of Dairy Science*, 75 (12) 1992: 3432-3451.

Blattberg, Robert C., Richard Briesch, and Edward J. Fox, "How Promotions Work," *Marketing Science*, 14 (3) Part 2 of 2: Special Issue on Empirical Generalizations in Marketing, 1995: G122-G132.

Blayney, Don P. and Richard F. Fallert, "Biotechnology and Agriculture: Emergence of Bovine Somatotrophin (bST)," United States Department of Agriculture, Economic Research Service, Commodity Economics Division, June 1990.

Butler, L. J., "rBST: Adoptions and Concerns of California's Dairy Producers," *Small Farm News Winter 1999*, University of California, Small Farm Center, 1999, Online: http://www.sfc.ucdavis.edu/pubs/SFNews/Winter99/rbst.html (Accessed: January 23, 2004).

Dhar, Tirtha and Jeremy D. Foltz, "Market Structure and Consumer Valuation in the rBST-free and Organic Milk Markets," Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Montreal, Canada, July 27-30, 2003.

Dimitri, Carolyn and Catherine Greene, *Recent Growth Patterns in the U.S. Organic Foods Market*, Economic Research Service, U.S. Department of Agriculture, Agriculture Information Bulletin No. AIB-777, September 2002.

Dobson, W. D. "The BST Case" in Government and the Food Industry: Economic and Political Effects of Conflict and Co-operation, L. Tim Wallace and William R. Scroder, Editors, Kluwer Academic Publishers, Boston, 1997: 141-158.

DuPuis, E. Melanie, "Not in my body: rBGH and the rise of organic milk," *Agriculture and Human Values*, 17 (2000): 285-295.

Federal Register, 58 (217) Friday, November 12, 1993 / Rules and Regulations, Department of Health and Human Services, Food and Drug Administration, 21 CFR Part 510 and 522, "Animal Drugs, Feeds, and Related Products; Sterile Sometribove Zinc Suspension; Final Rule," 59946-59947.

Green, Geoffrey M. and John L. Park, "New Insights into Supermarket Promotions via Scanner Data Analysis: The Case of Milk," *Journal of Food Distribution Research*, 29 (3) 1998: 44-53.

Greene, William H., *Econometric Analysis*, Fourth Edition, Prentice Hall Inc.: Upper Saddle River, New Jersey, 2000.

Grobe, Deana and Robin Douthitt, “Consumer Acceptance of Recombinant Bovine Growth Hormone: Interplay Between Beliefs and Perceived Risks,” *The Journal of Consumer Affairs*, 29 (1) 1995: 128-123.

Hausman, Jerry and Daniel McFadden, “Specification Tests for the Multinomial Logit Model,” *Econometrica*, 52 (5) 1984:1219-1240.

Juskevich, Judith C. and Greg Guyer, “Bovine Growth Hormone: Human Food Safety Evaluation,” *Science*, 249 (4971) Aug. 24, 1990: 875-884.

Kaiser, Harry M., Clifford W. Scherer, and David M. Barbano, “Consumer Perceptions and Attitudes towards Bovine Somatotropin,” *Northeastern Journal of Agricultural and Resource Economics*, 21 (1) 1992: 10-20.

Lee, Kyung Hee and Charles B. Hatcher, “Willingness to Pay for Information: An Analyst’s Guide,” *The Journal of Consumer Affairs*, 35 (1) 2001: 120-140.

Mathios, Alan D., “The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market,” *The Journal of Law and Economics*, 43 (2) 2000: 651-677.

Mathios, Alan D., “The Importance of Nutrition Labeling and Health Claim Regulation on Product Choice: An Analysis of the Cooking Oils Market,” *Agricultural and Resource Economics Review*, 27 (2) 1998: 159-168.

McGuirk, Anya M., Warren P. Preston, and Gerald M. Jones, “Introducing Foods Produced Using Biotechnology: The Case of Bovine Somatotropin,” *Southern Journal of Agricultural Economics*, July 1992: 209- 223.

Organic Consumers Association, “rBGH-Free Dairy Sources,” Online: http://www.organicconsumers.org/rbgh/rbghist.cfm (Accessed: April 22, 2003).

Suits, Daniel B., “Dummy Variables: Mechanics V. Interpretation,” *The Review of Economics and Statistics*, 66 (1) 1984: 177-180.

Table 1. Variable Names and Definitions

Variable Name	Definition
$Units_{jst}$	Units sold during the course of the week
$Price_{jst}$	Retail price, already reduced to reflect the sale price
$Sbrand_j$	1 if store brand; 0 otherwise
$Region_j$	1 if regional brand; 0 otherwise
$Nation_j$	1 if national brand; 0 otherwise
$Hhsz_s$	Store average shopper club household size
Inc_s	Store average shopper club income
$Educ_s$	Store average shopper club education level
Age_s	Store average shopper club age
$Whole_j$	1 if whole milk; 0 otherwise
$2Percent_j$	1 if 2 percent milk; 0 otherwise
$1Percent_j$	1 if 1 percent milk; 0 otherwise
$Skim_j$	1 if skim milk; 0 otherwise
$Conventional_j$	1 if conventional agricultural techniques; 0 otherwise
$Rbstfree_j$	1 if rBST-free; 0 otherwise
$Organic_j$	1 if organic; 0 otherwise
$Lactose_j$	1 if reduced lactose free; 0 otherwise
$Plain_j$	1 if plain flavored; 0 otherwise
$Flavor_j$	1 if flavored; 0 otherwise
$Plastic_j$	1 if plastic container; 0 otherwise
$Carton_j$	1 if paper carton; 0 otherwise
$Holiday_t$	1 if week contained secular or non-secular holiday; 0 otherwise

Table 2. Descriptive Statistics

Variable Name	Mean	Standard Deviation	Minimum	Maximum
$Units_{jst}$	113.8187	170.3040	1.0000	1,190.0000
$Price_{jst}$	2.2496	0.8528	0.69	4.99
$Sbrand_j$	0.2828	0.4504	0.0000	1.0000
$Region_j$	0.3136	0.4640	0.0000	1.0000
$Nation_j$	0.4036	0.4906	0.0000	1.0000
$Whole_j$	0.2662	0.4420	0.0000	1.0000
$2Percent_j$	0.2429	0.4289	0.0000	1.0000
$1Percent_j$	0.1285	0.3346	0.0000	1.0000
$Skim_j$	0.3624	0.4807	0.0000	1.0000
$Conventional_j$	0.7121	0.4528	0.0000	1.0000
$Rbstfree_j$	0.1530	0.3600	0.0000	1.0000
$Organic_j$	0.1350	0.3417	0.0000	1.0000
$Lactose_j$	0.2126	0.4092	0.0000	1.0000
$Plain_j$	0.8062	0.3953	0.0000	1.0000
$Flavor_j$	0.1938	0.3953	0.0000	1.0000
$Plastic_j$	0.0667	0.2496	0.0000	1.0000
$Carton_j$	0.9333	0.2496	0.0000	1.0000
$Holiday_t$	0.5010	0.5000	0.0000	1.0000

Table 3. Hypothesized Signs

Variable	Effect on Log Partial Odds Ratio
$Price_{jst}$-$Price_{1st}$	(-)
$Price_{jst} \times Rbstfree_j$-$Price_{1st} \times Rbstfree_1$	(-)
$Price_{jst} \times Organic_j$-$Price_{1st} \times Organic_1$	(-)
$Nation_j$-$Nation_1$	(-)
$Regional_j$-$Regional_1$	(-)
$Whole_j$-$Whole_1$	(-)
$Holiday_t \times Whole_j$-$Holiday_t \times Whole_1$	(+)
$2Percent_j$-$2Percent_1$	(+)
$1Percent_j$-$1Percent_1$	(+)
$Rbstfree_j$-$Rbstfree_1$	(-)
$Organic_j$-$Organic_1$	(-)
$Lactose_j$-$Lactose_1$	(-)
$Flavor_j$-$Flavor_1$	(-)
$Plastic_j$-$Plastic_1$	(-)
$Hhsz_s \times Rbstfree_j$-$Hhsz_s \times Rbstfree_1$	(+)
$Hhsz_s \times Organic_j$-$Hhsz_s \times Organic_1$	(+)
$Inc_s \times Rbstfree_j$-$Inc_s \times Rbstfree_1$	(+)
$Inc_s \times Organic_j$-$Inc_s \times Organic_1$	(+)
$Educ_s \times Rbstfree_j$-$Educ_s \times Rbstfree_1$	(+) or (-)
$Educ_s \times Organic_j$-$Educ_s \times Organic_1$	(+) or (-)
$Age_s \times Rbstfree_j$-$Age_s \times Rbstfree_1$	(-)
$Age_s \times Organic_j$-$Age_s \times Organic_1$	(-)

Table 4. Results for X Matrix With Price Interaction Terms

Variable	Parameter Estimate	Standard Error	Pr > |t|
Intercept	-1.29731	0.03056	<.0001
price	**-0.78488**	**0.04866**	**<.0001**
price × rbstfree	**-1.48352**	**0.09528**	**<.0001**
price × organic	**0.00261**	**0.10524**	**0.9802**
nation	-0.75089	0.10229	<.0001
region	-0.54821	0.04012	<.0001
whole	0.25275	0.03730	<.0001
holiday ×whole	0.04372	0.04043	0.2795
twopercent	0.58101	0.03324	<.0001
onepercent	0.47479	0.03729	<.0001
rbstfree	1.91845	0.17024	<.0001
organic	-0.88029	0.35614	0.0135
lactose	0.07582	0.07255	0.2960
flavor	-0.43916	0.04778	<.0001
plastic	-0.52552	0.06565	<.0001

R-Square:	0.5850
Adj. R-Square:	0.5844
F Value:	861.60
Pr > F:	<.0001

Table 5. Full model Results with Store and Monthly Fixed Effects

Variable	Parameter Estimate	Standard Error	Pr > |t|
store01	-1.25044	0.06439	<.0001
store02	-0.89485	0.06890	<.0001
store03	-1.29464	0.06724	<.0001
store04	-0.44296	0.06657	<.0001
store05	-1.46421	0.06529	<.0001
store06	-1.24572	0.06867	<.0001
store07	-1.49165	0.06490	<.0001
store08	-1.09631	0.06792	<.0001
store09	-1.33092	0.07073	<.0001
store10	-0.34135	0.07618	<.0001
store11	-1.55209	0.06778	<.0001
store12	-1.40966	0.06743	<.0001
store13	-1.49783	0.06402	<.0001
store14	-1.25339	0.06301	<.0001
store15	-1.34409	0.06539	<.0001
store16	-0.94225	0.06935	<.0001
store17	-0.96805	0.06114	<.0001
store18	-1.23341	0.06087	<.0001
store19	-1.84872	0.06468	<.0001
store20	-1.05539	0.06474	<.0001
november	0.01723	0.03687	0.6402
december	0.01981	0.03831	0.6050
january	0.04173	0.03711	0.2608
february	0.11017	0.03830	0.0040
march	0.18233	0.03841	<.0001

Table 5. (Continued)

Variable	Parameter Estimate	Standard Error	Pr > |t|
price	-1.01667	0.04836	<.0001
price × rbstfree	-1.13649	0.10009	<.0001
price × organic	0.12946	0.10145	0.2020
nation	-0.43599	0.10083	<.0001
region	-0.64509	0.04391	<.0001
whole	0.17447	0.03704	<.0001
holiday × whole	0.06159	0.03936	0.1177
twopercent	0.55377	0.03234	<.0001
onepercent	0.45026	0.03644	<.0001
rbstfree	2.01360	0.22048	<.0001
organic	-0.93710	0.41183	0.0229
lactose	0.07055	0.07039	0.3162
flavor	-0.42101	0.04722	<.0001
plastic	-0.51407	0.06678	<.0001
prevwksale	-0.15707	0.04032	<.0001
hhsz × rbstfree	-0.03745	0.00782	<.0001
hhsz × organic	-0.04247	0.00966	<.0001
inc × rbstfree	0.08094	0.01400	<.0001
inc × organic	0.07125	0.01562	<.0001
educ × rbstfree	-0.03976	0.01481	0.0073
educ × organic	-0.04568	0.01924	0.0176
age × rbstfree	-0.05529	0.00713	<.0001
age × organic	-0.01318	0.00749	0.0785

Note: The model was estimated without an overall constant term and R-square was no longer an accurate measure of goodness of fit.

Table 6. Hedonic Willingness to Pay Estimates

	Linear Model			Semilog Model		
Variable	**Parameter Estimate**	**Standard Error**	**Pr > ChiSq**	**Parameter Estimate**	**Standard Error**	**Pr > ChiSq**
Intercept	1.52134	0.02276	<.0001	0.44335	0.01216	<.0001
whole	0.03913	0.00766	<.0001	0.02856	0.00342	<.0001
one	-0.01986	0.00710	0.0052	-0.00760	0.00367	0.0382
skim	-0.01251	0.00703	0.0751	0.00640	0.00320	0.0456
rbstfree	0.25947	0.01299	<.0001	0.14300	0.00739	<.0001
organic	0.73059	0.03212	<.0001	0.32453	0.01261	<.0001
flavor	0.52689	0.01974	<.0001	0.24563	0.00957	<.0001
carton	-0.14591	0.02319	<.0001	-0.12994	0.01242	<.0001
region	0.05063	0.01042	<.0001	0.04522	0.00656	<.0001
nation	1.17164	0.02972	<.0001	0.53559	0.01257	<.0001
lactose	0.62093	0.03053	<.0001	0.29307	0.01261	<.0001
R-Squared	0.9126			0.8958		
F-Value	8938.67			7359.45		
Pr > F	<.0001			<.0001		