Health-Related Product Attributes and Purchasing Behavior in the Ready-to-Eat Cereal Market: An Application with Household-level, Censored Data

Jing Li
Ph.D. Student in Agricultural, Environmental, and Regional Economics
Department of Agricultural Economics, Sociology, and Education
The Pennsylvania State University
Email: jing.li@psu.edu

Edward C. Jaenicke
Associate Professor of Agricultural Economics
Department of Agricultural Economics, Sociology, and Education
The Pennsylvania State University
Email: tjaenicke@psu.edu

Tobenna Anekwe
Research Economist
USDA Economic Research Service
Email: tanekwe@ers.usda.gov

DRAFT: July 2013

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2013 AAEA & CAES Joint Annual Meeting, Washington, DC, August 4-6, 2013.

We thank USDA and Tobenna Anekwe of the USDA's Economic Research Service for collaborating with us on this research under a Third Party Agreement with Nielsen.

Introduction

Consumers are increasingly aware of nutrition information and health claims, and their food purchasing behavior reflects it. During 2009, for example, food products with nutrition claims about fiber, calcium or no salt/sodium rank into the leading groups for sales growth, with an average growth rate from 10 percent for no salt/sodium and 13 percent for fiber (Pirovano, 2010). According to the Dietary Guidelines for Americans in 2010, the USDA and FDA suggest that Americans should increase their intake of whole grain, dietary fiber, and decrease consumption of refined grains, added sugars, solid fats, and sodium (USDA/FDA, 2010). By following these guidelines, Americans could reduce the risk of having heart disease, diabetes, or other chronic health problems. As a result, food products that fit these guidelines become the focus of people's food choice in keeping their healthy-eating lifestyle.

Ready-to-eat breakfast cereals are a good example of products that can fit the dietary guidelines and, sure enough, they have experienced a significant growth of sales from 2002 to 2010. According to the annual reports from the top two cereal companies, General Mills and Kellogg's, total sales of ready-to-eat breakfast cereals have increased by about 36% from \$3.9 billion to \$5.3 billion. During the same time period, product reformulations have improved the nutrition quality of many cereal products through a reduction in sugar and sodium, and an increase in fiber content per serving. From 2006 to 2009, over half of the cereal products offered by the four largest cereal companies have been reformulated to catch up with consumers' increasing awareness of healthy eating. Nutrition quality for existing cereal brands has improved by 14 percent for kids cereal, 12 percent for family cereals and 5 percent for adult cereals during the year 2006 to 2012 (Harris, et al., 2012). Among these quality improvements, large cereal companies contribute most of the positive change.

Existing research, including studies that examine ready-to-eat cereals, has investigated the relationship between health-related product attributes and consumers' purchasing behaviors. However, few studies use detailed nutrition information explicitly as factors in analyzing consumers' food demand. In most cases, only a few nutrition indicators or claims are used (Muth et al., 2009; Thunström, 2010). One of the exceptions is Golub and Binkley (2005), which uses both individual level purchase data and a USDA nutrition dataset. However, they use the nutrition information to generate the healthiness indicators of food product used in the estimation, but do not include them as explanatory factors. On the other hand, our paper incorporates rich nutrition information directly and explores its impact on consumers' choice of breakfast cereal, including fiber, sugar, sodium, calcium, fat, and other information.

One of the most challenging aspects of brand-level demand estimation in differentiated market is the dimensionality problem, which arises when consumers face a vast array of choices among closely related but differentiated products. Brand-level demand estimation in these markets needs to handle large number of parameters in estimating the substitutions among numerous differentiated brands.

Empirically, two types of models attempt to deal with this dimensionality issue. The most commonly used one is discrete choice modeling, including the logit, nested logit, or random coefficient logit models. They map consumers' purchasing decision from product space to characteristic space, which significantly helps to reduce the number of model parameters. However, while these models overcome the dimensionality problem, they still have several limitations. Both the logit and nested logit models suffer from the restrictive IIA assumptions and limit the flexibility of reflecting the substitution patterns among differentiated products. Although random coefficient logit model can solve IIA problems, its ability to model a large number of brands is limited due to significant computational burden of estimating the non-closed form integration of market shares. Moreover, since discrete choice models assume that consumer choose only one unit of product that brings the highest utility, this type of model cannot fit cases where consumers usually make multiple purchases or multiple choices. For example, a discrete choice model may not be a good fit for the ready-to-eat cereal case because many consumers purchase multiple cereal products in each shopping trip. The second type of modeling framework is the neoclassical demand model. Compared to discrete choice models, neoclassical demand models do not require single-unit choice assumption. A good example is Hausman, Leonard and Zona (1994) who adopt the multi-stage budgeting approach (Gorman, 1971) and use the AIDS model to help reduce the unknown parameters in the lowest stage. However, the dimensionality still exists because their model is still product-space based and there are still a large number of brands.

Pinkse, Slade and Brett (2002) propose the Distance Metric (DM) method to handle these issues. In their model, the relationship between differentiated products is captured by the relative distance between products in characteristics space along multiple dimensions. The closer the two products are, the more likely that these two will become substitutes as price changes. As a result, the number of unknown parameters can be significantly reduced because the cross price elasticity is transferred to be a function of distances of key characteristics. Compared to previous modeling efforts, the DM method allows more flexible substitutions between products since it does not impose any prior assumptions of groupings or decision stages as in nested logit model or the multi-stage budgeting approach. It also has fewer computational issues due to its simple functional form compared to the random coefficient logit model.

Because an individual consumer or household often purchases only a very small number of brands, consumer-level scanner data can exacerbate the dimensionality problem associated with differentiated products by recording so-called zero purchases. While this second type of empirical problem may not be an issue for discrete choice demand models, it is for neoclassical demand models, where the likelihood function for a censored demand system includes a large number of integrals. While this problem can sometimes seem insurmountable, several methods, such as Quasi-Maximum Likelihood and Bayesian econometrics, are able to handle the estimation of large number of integrals with large computation burden (Yen and Lin, 2008; Kasteridis, Yen, and Fang, 2011).

In this paper, we suggest that the DM method provides a second benefit – in addition to easily handling the dimensionality problem – by handling with relative ease a censored demand system often encountered with consumer-level purchase data. Because the Pinkse, Slade and Brett (2002) DM method converts demand estimation into characteristic space, and because simplifying symmetry assumptions are made about the cross-price parameters, the DM method reduces estimation to just a single demand equation. A censored demand system can therefore be estimated with a single Tobit model, which requires the solution to a single integral in the likelihood function. To our knowledge, all previous applications of the DM method (Pinkse and Slade, 2004; Slade, 2004; Rojas, 2008, Rojas and Peterson, 2008; Bonanno, 2012) rely on market-level data, and this paper therefore represents the first application of the DM method in the demand estimation to micro-level, censored purchase data. .

This paper evaluates the impact and importance of product attributes, including health-related product attributes on consumer purchasing behavior in the RTE breakfast cereal industry. Cereal purchase data come from the Nielsen Homescan dataset that includes product characteristics and household demographics. Normally, this micro-level data would present the empirical issues associated with both dimensionality and censored data. However, by following Pinkse, Slade and Brett (2002) and Pinkse and Slade (2004), we use the DM method to measure the relative distance between each pair of cereal brands in product attribute space and approximate the closeness of competition among different brands. We incorporate the DM method into a censored regression model and also account for consumers' heterogeneity of preferences for health-related product attributes. Finally, we estimate conditional and unconditional own- and cross-price elasticities for a wide range of product attributes.

In section two of this paper, we will summarize the existing literature on health and consumer purchasing behavior, the Distance Metric method, as well as previous studies about cereal. Section three and four will describe the modeling framework and estimation methods. Section five will introduce the Nielsen Homescan data and nutrition data for cereal products. The last two sections discuss the empirical results and concluding remarks.

Background Information

Research on the relationship between health and consumer demand is not new. In the last two decades, there has been an increasing body of studies in this field and the topics vary widely. Generally they can be classified into three sub areas: (1) How consumer awareness of health can have a measureable effect on consumers food purchasing behaviors (Brown and Schrader, 1990; Chern et. al., 1995; Kinnucan et al., 1997; Blaylock et. al., 1999); (2) How a hedonic pricing model can be used to show how consumers' willingness to pay is affected by some nutrients found in the food product (Lenz, Mittelhammer and Shi, 1994; Kim and Chern, 1995; Shi and Price, 1998; Huffman and Jensen, 2004; Li, McCluskey, and Wahl, 2004; Muth et al., 2009; Barreiro-Hurle et al., 2010); and (3) How nutrient information reflected through nutrition labels or health claims are found to facilitate consumers' food choices, either positively or negatively(National Institutes of Health, 2004; Wansink and Chandon, 2006, Shepherd et al., 1992; Wardle and Huon, 2000; Wansink and Park, 2002; Berning, et al., 2011), depending on consumers' characteristics (Wang et al., 1995; Nayga, 1996; Govindasamy and Italia, 2000; Coulson, 2000; McClean-Meyinsse, 2001; Kim et al., 2001a, b; Drichoutis et al., 2005), and the format of nutrient information (Levy et al., 1996; Wansink, 2003; Barreiro-Hurle et al., 2010).

A wide variety of methodologies have been adopted to analyze consumers' responsiveness to health-related product attributes. These methods include continuous choice models, such as the AIDS model (Binkley and Eales, 2000), or the Rotterdam demand system (Capps and Schmitz, 1991), and also discrete choice models, such as logit (Drichoutis et al., 2005), multinomial logit or probit (Coulson, 2000; Chowdhury, 2011), or random coefficient logit models (Chidmi and Lopez, 2007). However, when applied to highly differentiated food markets, these models are subject to certain limitations, such as the independence from irrelevant alternative (IIA) problem. Although the random coefficient logit model can solve the IIA problem, it is computationally intense. Also, for continuous choice models, dimensionality becomes an obstacle as the number of brands increase.

To address these problems, Pinkse, Slade and Brett (2002) propose the Distance Metric (DM) method, which can easily accommodate a large number of differentiated brands, and requires few a priori assumptions and restrictions, thereby allowing one to estimate own- and cross- price elasticities with more flexibility. Finally, it introduces the notion that closer distance in characteristic space increases the competition among products.

Following Pinkse, Slade and Brett (2002), Pinkse and Slade (2004) apply the same spatial model to evaluate the impact of mergers on price in the draft beer market in United Kingdom (UK). This study extends Pinkse, Slade and Brett (2002) to demand estimation of consumers instead of the demand of downstream firms. Slade (2004) compares the estimation from both nested multinominal logit (NML) and the DM method and shows that although these two methods both show no evidence of coordinated effects in estimated price-cost margins, the DM specification tends to show more reasonable (negative cross price elasticity) and significant price elasticity results than the NML specification.

A common feature of Pinkse and Slade (2004) and Slade (2004) is that they both choose a quadratic indirect utility function and use Roy's identity to derive the uncompensated demand functions. However, this method implicitly assumes that consumers have the same constant marginal utility of income, which is too strong and cannot hold in the dataset for long time periods. To solve this problem, Rojas and Peterson (2008) incorporate the DM method into a modified AIDS model which relaxes the constant marginal utility assumption. Bonanno (2012) investigates the functional yogurt market in Italy and finds that functional yogurt is less price sensitive than conventional yogurt. Moreover, he points that brand loyalty can be analyzed by calculating the elasticity within and across brands since the cross-price elasticity is a function of product distance in brand. The results suggest that brand loyalty plays an important role for consumers' switching decisions. Rojas (2008) uses brand-level data to evaluate producers' price reactions to an increased excise tax in U.S. beer market, and gets consistent results of advertising with those in Rojas and Peterson (2008). Finally, the DM method is also applied in analyzing new product introduction and constructing the unobserved price for new products (Pofahl and Richards, 2009).

Although the DM method has shown strong ability in estimating highly differentiated markets, very few works apply it to the food market to analyze the impact of health-related product attributes on consumers' demand. However, all of these works focus on market level demand estimation. No one, as we know, has applied this method into the demand estimation using household-level purchasing data.

Over the past few decades, there have been large numbers of studies on ready-to-eat breakfast, and these either address the demand side estimation, such as consumer brand choice (e.g., Nevo, 2001), or the competition among producers because of the unique oligopoly market structure (e.g., Schmalensee, 1978; Scherer, 1979). Due to the large variety of differentiated brands and clearly defined health-related product attributes, researchers are also interested in analyzing the relationship between healthy attributes of cereal and consumer demand. Ippolito and Mathios (1989) and, Binkley and Eales (2000) find that the fiber content in breakfast cereal plays an important role in affecting consumers purchase decisions. They also find that consumer demographics have an impact on cereal choices: households with children purchase less healthy cereal products with more sugar, while households with teenagers value energy more than other nutrients (Binkley and Golub, 2010; Thunström 2010). Also, households with higher income, education or age could also contribute to healthier cereal choices (Shi and Price, 1998; Golub and Binkley, 2005). However, taste and habit might become contradicting forces that deter consumers to healthy cereal choices. Binkley and Eales (2000) shows that households with children tend to purchase cereal with relatively less fiber because of the unpleasant taste. Thunström (2010) finds that consumers in the breakfast cereal market are highly habit persistent, some of which are unhealthy. This paper will continue to explore the relation between health-related product attributes and consumers purchasing decisions. It incorporates much richer nutrition information and consumer heterogeneity and uses the DM method to capture this relationship in a very flexible way.

Model and Estimation Method

Lancaster (1966, 1979) posits that consumers actually derive utility from the attributes existing in the products they consume. As a result, consumers' optimal choice can be transferred from a framework based on product space into one based on product-attribute space, where consumers maximize utility by choosing the level of product attributes within their budget constraint. Within this theoretical basis, Pinkse, Slade and Brett (2002) introduce the Distance Metric method and claim that it is preferred in demand analysis for markets with large numbers of differentiated products with more flexibility and tractability as compared to previous demand estimations. Tractability stems from the fact that cross-price parameters are defined as a function of product attributes and relative distance in key characteristics.

Demand model

We follow Rojas and Peterson (2008) by incorporating the Distant Metric method into a linear approximation of the Almost Ideal Demand System (LA/AIDS) from Deaton and Muellbauer (1980). Let $i \in (1,..., I)$ denote all the consumers, $j \in (1,..., J)$ the set of brands, and $t \in (1,..., T)$ the index of markets. Here, the market index, *t*, is defined as a market-year pair. So the expenditure share function for consumer, *i*, who purchases cereal brand, *j*, in market, *t*, can be originally represented as

$$w_{ijt} = a_{jt} + \sum_{\mathrm{k}=1}^{J} b_{ijk} \log p_{ikt} + c_{ij} \log \frac{\mathrm{X}_{\mathrm{it}}}{\mathrm{P}_{\mathrm{it}}^{\mathrm{L}}} + e_{ijt} \qquad (1)$$

where $w_{ijt} = q_{ijt} p_{ijt} / x_{it}$ represents the expenditure share for consumer *i*, purchasing brand *j* in market *t*. q_{ijt}, p_{ijt} are the corresponding purchase quantities and price for consumer, *i*, and $x_{it} = \sum_j q_{ijt} p_{ijt}$ is the total expenditure for consumer *i* for all products in market *t*. The term $\log\mathrm{P}_{\mathrm{it}}^{\mathrm{L}}$ is the price index which is an approximated loglinear analogue of the Laspeyeres index (Moschini, 1995), in which $\log\mathrm{P}_{\mathrm{it}}^{\mathrm{L}} \approx \sum_{j=1}^{J} w_{ij}^{0} \log p_{ijt}$ and w_{ij}^{0} denotes brand *j*'s base share for consumer *i* with $w_{ij}^{0} \equiv Y^{-1} \sum_{y=1}^{Y} w_{ijy}$ and

$y \in (1,...,Y)$ represents the year. Then the base share of brand *j* for consumer *i* becomes yearly average of consumer *i*'s purchase share on brand *j*. Finally, a_{jt} b_{ijk} and c_{ij} are parameters to be estimated.

If one does not impose any restrictions on the parameters of equation (1) and does not consider consumer heterogeneity, then (*J*-1) seemingly unrelated equations need to be estimated. However, if there is a large number of differentiated brands, such as one finds in the breakfast cereal market, then it will become difficult or even impractical to estimate large system of equations since the total number of estimated parameters for cross-price terms can be as high as (*J*-1)**J*/2. In order to overcome this dimensionality problem and fit the model to the breakfast cereal market data, the DM method imposes a few restrictions on this share equation to reduce the number of parameters to be estimated.

First, and most importantly, Pinkse, Slade and Brett's (2002) Distance Metric method specifies each cross-price coefficient b_{ijk} as a function of the distance measures between brands *j* and *k* (denoted as δ_{jk}) in product attribute space, which is $b_{ijk} = g(\delta_{jk})$. It reflects a notion that price competition is decided by the relative position of product in characteristics space. The distance between brands *j* and *k* is symmetric by definition, which has $\delta_{jk} = \delta_{kj}$. Product attributes used are either continuous, δ_{jk}^{c} (e.g. dietary fiber, sugar content etc.) or discrete, δ_{jk}^{d} (whole grain, vitamin fortified, kids), b_{ijk} is defined as a linear combination of continuous distance δ_{jk}^{c} or discrete distance δ_{jk}^{d} separately:

$$b_{ijk} = \begin{cases} \sum_{c=1}^{C} \lambda_c \delta_{jk}^{c} & \text{for continuous distance} \\ \sum_{d=1}^{D} \lambda_d \delta_{jk}^{d} & \text{for discrete distance} \end{cases} \qquad (2)$$

To satisfy the Slutsky symmetry requirement for the cross-price coefficient, $b_{ijk} = b_{ikj}$, λ_c and λ_d are assumed to be the same across equations for all brands. Also, they are constant across consumers, which means $\lambda_{1c} = \lambda_{2c} = \cdots = \lambda_{Ic} = \lambda_c$ and $\lambda_{1d} = \lambda_{2d} = \cdots = \lambda_{Id} = \lambda_d$.

Second, to help reduce the system of equations to one estimated equation, it further assumes that the constant term, a_{jt}, own-price coefficient, b_{ijj}, and the coefficient of the price index. c_{ij}. are linear functions of brand *j*'s characteristics: $a_{jt} = a_0 + \sum_{l=1}^{L} a_l z_{jl}^{a}$, $b_{ijj} = b_0 + \sum_{m=1}^{M} b_m z_{jm}^{b}$ and $c_{ij} = c_0 + \sum_{n=1}^{N} c_n z_{jn}^{c}$, where z_{jl}^{a}, z_{jl}^{b}, and z_{jl}^{c} all represent brand *j*'s characteristics. However, in order to avoid multi-collinearity, each one represents a subset of brand *j*'s characteristics separately. By construction, a_{jt}, b_{ijj} and c_{ij} are equal across consumers, which is similar to the cross-price coefficient, b_{ijk}, as mentioned above.

Therefore, equation (1) can be rewritten by imposing the restrictions above and adding an error term:

$$w_{ijt} = a_{jt} + b_{ijj} \log p_{ijt} + \sum_{c=1}^{C} (\lambda_c \sum_{k \neq j} \delta_{jk}^{c} \log p_{ikt}) + \sum_{d=1}^{D} (\lambda_d \sum_{k \neq j} \delta_{jk}^{d} \log p_{ikt}) + c_{ij} \log \frac{X_{it}}{P_{it}^{L}} + e_{ijt} \qquad (3a)$$

Substituting a_{jt}, b_{ijj} and c_{ij} in equation (3a), the share takes the following form:

$$\begin{aligned} w_{ijt} &= a_0 + \sum_{l=1}^{L} a_l z_{jl}^{a} + (b_0 + \sum_{m=1}^{M} b_m z_{jm}^{b}) \log p_{ijt} \\ &+ \sum_{c=1}^{C} (\lambda_c \sum_{k \neq j} \delta_{jk}^{c} \log p_{ikt}) + \sum_{d=1}^{D} (\lambda_d \sum_{k \neq j} \delta_{jk}^{d} \log p_{ikt}) \qquad (3b) \\ &+ (c_0 + \sum_{n=1}^{N} c_n z_{jn}^{c}) \log \frac{X_{it}}{P_{it}^{L}} + e_{ijt} \end{aligned}$$

3.1 Distance Metric

Based on available data on product attributes, two types of distance measures are utilized in this paper. The first is continuous distance measure, denoted as δ_{jk}^{C}. It is computed by using continuous product attributes (z_j^c) under Euclidean space. Six continuous attributes are considered for this research: fiber, sugar, sodium, fat, and calcium content, and vitamin fortification. Here, vitamin fortification is approximated by the total number of vitamins which are fortified based on a 2,000 calorie diet, in accordance with U.S. Food and Drug Administration (FDA). As in Rojas (2008), a multi-dimensional continuous distance is computed by using the inverse distance between brand *j* and *k* across the six attributes. Therefore, the continuous distance used in this paper, δ_{jk}^{C}, is defined as:

$$\delta_{jk}^{C} = \frac{1}{1 + 2\sqrt{\sum_{c=1}^{C} (z_j^c - z_k^c)^2}} \qquad (4)$$

Where z_j^c represent the *c*-th continuous attribute and *C* equals to the total number of continuous attributes used to generate the multi-dimensional continuous distance. The reason of using the inverse distance between brand *j* and *k* is for easy interpretation. Increased "closeness" in continuous attribute space leads to a larger value of δ^{C}_{jk} and these cereal brands are more likely to be substitutes to each other. Because this paper uses only one continuous distance measure, the summation of continuous distance across different continuous attributes in (3b) can be simplified to only one continuous indicator, δ^{C}_{jk}.

The second type of distance measure (δ^{d}_{jk}) is constructed based on discrete (binary) product attributes (z^{d}_{j}). Four binary attributes are utilized: whole grain, flavor added, manufacturer, and targeted to children. The distance between brand *j* and *k* for each discrete attribute is defined to be one if they belong to the same product type (e.g., both are whole grain, or produced by the same manufacturer), and zero otherwise. Different than the continuous distance measure, the discrete distance measure captures the competition among products within the same type. In other words, discrete distance captures the local competition while continuous distance can capture the global competition. The distance measure for each discrete product characteristic is defined as:

$$\delta^{d}_{jk} = \begin{cases} 1 & \text{if } \left| z^{d}_{j} - z^{d}_{k} \right| = 0 \\ 0 & \text{if } \left| z^{d}_{j} - z^{d}_{k} \right| \neq 0 \end{cases} \qquad (5)$$

Finally, as defined above, for each product characteristic, the distance metric for brand *j* relative to all other brands *k*, $k \neq j$, is symmetric and negative-semidefinite. The diagonal elements are represented as functions of own attributes. The off diagonal elements are the relative distances between brands *j* and *k*, either continuous or discrete. They are symmetric and have non-negative values between zero and one by construction.

3.2 Demographic translating of the LA/AIDS model

As mentioned above, consumers' purchasing decisions are also affected by socio-demographic factors. Particularly in a market with many differentiated products, consumer heterogeneity in preferences can be even more important. As a result, considering consumer heterogeneity of preferences will help to increase the accuracy of demand estimation. Most, if not all, of the current applications of the Distance Metric method in demand estimation do not include consumer heterogeneity.

This paper incorporates impact of consumers' demographics $(h_{1it} \ldots h_{Hit})$ on purchasing behavior, where h_{hit} represents the *h*-th demographic variables for consumer *i* at market-time *t*, with $h = 1 \ldots H$. Demographics are included as demand shifters in the AL/AIDS model to capture the average taste of consumers on cereal products. So, the form of constant term is changed to:

$$a_{jt} = a_0 + \sum_{l=1}^{L} a_l z_{jl}^{a} + \sum_{h=1}^{H} a_h h_{hit}$$

The demographic characteristics include information about the household head, such as age, marriage, race, and education. And it also contains information about household income, household size, and age and presence of children.

3.3 Extension to a censored dependent variable

In general, since ready-to-eat cereal market is full of numerous differentiated products, each household's purchase is limited to a few brands and no purchases (zeros) for the rest. A zero purchase of a specific brand could be caused by empirical or theoretical reasons. First, empirical selection of a time period can cause the unobserved purchase of a brand. However, given multiple years of data are used in this research, this impact can be trivial. On the other hand, some research argues that a zero purchase outcome is the result of a utility maximization process, and is caused by a relatively high price, unattractive product attributes, or different preferences of consumers. Given that the dependent variable is nonlinear and restricted to be non-negative, zeros might lead to an underestimate of the impact of product attributes (or the distance of attributes) on consumer purchasing decisions if simply apply the standard demand model with a flexible functional form, like the LA/ALDS model mentioned above.

Historically, the Tobit model (Tobin, 1958) is widely used to deal with zero purchases in demand estimation. It is assumed that there is a latent variable of purchase share, w_{ijt}^{*} , representing the purchase behavior for consumer *i* choosing cereal brand *j* at market *t*. This outcome cannot be observed if a consumer does not make a purchase of a specific brand. Since it must be taken into consideration for accurate estimation, a latent purchase share w_{ijt}^{*} is used, where w_{ijt}^{*} is a linear function of product attributes or distance as described in equation (3b). The observed share w_{ijt} is assumed to equal to the latent share w_{ijt}^{*} whenever the latent share is above zero. Finally, an independent and normally distributed error $\varepsilon_{ijt} \sim N(0,\sigma^2)$ is used to capture the random relationship between the two shares. Combining the Tobit model with demographic transferred share equation:

$$w_{ijt} = \begin{cases} w_{ijt}^{*} & \text{if } w_{ijt}^{*} > 0 \\ 0 & \text{if } w_{ijt}^{*} \le 0 \end{cases} \qquad (6)$$

where w_{ijt}^{*} represents a latent variable such that

$$
\begin{aligned}
w^*_{ijt} &= a_0 + \sum_{l=1}^{L} a_l z^a_{jl} + \sum_{h=1}^{H} a_h h_{hit} + (b_0 + \sum_{m=1}^{M} b_m z^b_{jm}) \log\ p_{ijt} \\
&+ \lambda_c \sum_{k \neq j} \delta^C_{jk} \log\ p_{ikt} + \sum_{d=1}^{D} (\lambda_d \sum_{k \neq j} \delta^d_{jk} \log\ p_{ikt}) \qquad (7) \\
&+ (c_0 + \sum_{n=1}^{N} c_n z^c_{jn}) \log \frac{X_{it}}{P^L_{it}} + \varepsilon_{ijt}
\end{aligned}
$$

Under this specification, zero purchase outcomes are assumed to be the corner solution of utility maximization when consumers make a purchasing decision based on their budget constraint. Finally, this model can be consistently estimated by maximum likelihood estimation (Amemiya, 1973, 1984). The impact of healthy attributes in cereal products is evaluated by using this empirical model. Own-, cross-price elasticities, which are functions of distance in attribute space, are calculated following the demand estimation.

The final specification given by (6) and (7) thus easily accommodates two qualities – a large number of differentiated products (dimensionality) and zero purchases (a censored system) – that can plague empirical estimation of demand systems.

Data and Variables

The data used in this paper mainly come from three sources. The first one is the Nielsen Homescan data base. All the purchase information and household-level demographics are collected from this data source. The second one is the National Nutrient Database for Standard Reference (SR) from the USDA, which provides nutrition composition of various food products in the United States. Finally, some non-nutrient related product information is collected by using the web pages archived on the world-wide-web.

The Nielsen Homescan database used in this paper is a household-based scanner dataset that keeps record of each household’s grocery purchase trips. This dataset includes information on purchase date, store, price, and quantity, as well as some product information. It also contains detailed household-level demographic characteristics which can be attached to each purchase record. All the households in this dataset represent 52 metropolitan market areas, and one area that includes of the rest of the United States.

There are three product modules about cereal products in this Homescan dataset, including granola & natural cereal, hot cereal, and ready-to-eat cereal. This research focuses on ready-to-eat breakfast cereal, and applies to the DM method in a censored framework to analyze the impact of health-related product attributes on consumer purchasing behavior. Filtering for ready-to-eat cereal only yields an initial sample of 723,849, 717,556 and 675,690 purchase records from 36,664, 36,074 and 35,059 households for 2004, 2005, and 2006, respectively. In order to limit the percentage of zeros for the purchase, several selection criteria are applied. Households who have stayed in the panel for all the three years are considered in this research. If some household has less than 15 purchases of RTE breakfast cereal across these three years, this household is dropped. Also, twenty brands are selected since it contains the most observed purchase records. These twenty brands account for about 40% of the total market sale for the RTE breakfast cereal for each year. Private labels are not considered due to the limitation of collecting the product attributes, although, across all stores, private label cereals represent the most purchase records in the data.

Finally, the data sample used consists of 9,650 households with purchase among 20 cereal brands from 2004 to 2006. These observations are further aggregated across purchase trips for each consumer for each brand during each year. This means that each observation used in the estimation represents the yearly purchase for consumer *i* for each brand *j*. To keep the balance of this panel, records for brands that household did not buy are set to a purchase amount of zero, representing the non-purchase. So, the data contain 20 purchasing records for the twenty brands for each household. All of the households selected in this research continue to represent the same markets as the original data.

The price in log form (LNP) used in this research is volume-based. For products that household bought, the price is generated by dividing the total dollars paid by household *i* for brand *j* by the total volume bought by this household for this brand during each year. However, since the Nielsen Homsescan data only contains information for product that is not purchased, we need to construct the missing prices for those products. So, the market average price is used to approximate the missing price. The rest of the variables about purchase are listed in table 1.

Table 1: Summary of purchase and demographics from Homescan dataset (All HH: 578,880; HH w. child: 185,500)

Variable	Descriptions	All HH		HH with child	
		Mean	Std. Dev.	Mean	Std. Dev.
Purchase					
SH	Individual expenditure share for each brand	0.05	0.14	0.05	0.13
LNP	Average purchase price (cents), in log form	2.74	0.27	2.74	0.28
Demographics					
Continuous					
HHSIZE	Number of individuals in the household	2.72	1.33	4.12	1.12
HHINC	Household annual income	19.84	5.59	21.34	5.31
AGEHIGH	Max age category of household head(s)	7.14	1.75	5.71	1.49
Discrete		Frequency of one			
WHITE	Race is white for the household		0.84		0.77
MARRIED	Marital Status of household head(s)		0.75		0.85
CHILD	Presence of children with age under 18		0.32		1.00
HISPANIC	Hispanic origin for the household		0.93		0.89
HIGHSC	Max educational attainment for household head(s): high school		0.21		0.15
SOMECOL	Max educational attainment for household head(s): some college		0.29		0.26

The Nielsen Homescan data also contains detailed demographics for each household. It includes information about both household and household head. The head of household is self-defined by the person participating in the survey, and can be a single person or two persons, regardless of gender, marital or employment statuses. In the full sample, 76.95% of households have both female and male household heads. The rest either only have a female (16.95%) or male (6.09%) household head. In this research, all the information about household head is taken the largest value between the male and female household head if the household has two household heads. All the demographics controlled in the estimation includes household size (HHSIZE), household income (HHINC), married status (MARRIED), highest education of household head (high school, HIGHSC; some college, SOMECOL), highest age category of household head (AGEHIGH), and the binary indicator for ethnicity of the household (WHITE, HISPANIC). Summary statistics for the full sample and households with child sample are provided in table 1 above.

Nutrition composition information is collected from the USDA National Nutrient Database for Standard Reference (SR). All values for the nutrients listed are based on 100 grams of food products. Following previous studies using nutrition information, several important nutrient variables are considered in this research. It includes total per-serving sugar (SUGAR), total lipid fat (FAT), sodium (SODIUM), calcium (CALCIUM), total dietary fiber (FIBER), and the number of vitamin types fortified (NVF). NVF is constructed from the summation of five binary vitamin fortified variables, including vitamin A (IU), vitamin D, vitamin E (total ascorbic acid), vitamin B-6 and vitamin B-12. Based on the Nutrition Labeling and Education Act (NLEA) label declaration of % Daily Value (DV) from the U.S. Food and Drug Administration (FDA), each vitamin level in the food product is identified as fortified or not, and then a binary indicator is used for the record.

Non-nutrient related product attributes are collected from archived web pages for each cereal company. Cereal producers disclose product labels and descriptions on their official website. However, since this information is updated from time to time, it is difficult to observe it for previous years. The "Wayback Machine", a service provided by the Internet Archive, is used to help collect old product information from archived webpages during the year 2004 and 2006 in this research. It includes whole grain type (WG), which is a binary variable defined as having whole grain as the first ingredient. Additional flavor added (ADFLAVOR) is also a binary indicator for products include additional flavor content, besides their original input. Similarly, the variable color added (ADCOLOR) is given value one if this cereal product contains artificial color ingredient which is listed on the nutrition label. Finally, if cereal product includes fruit, the binary indicator (FRUIT) is given a value of one. The classification of cereal product's target is collected from both the official website of cereal producers and the cereal facts report from Yale

University's Rudd Center for Food Policy and Obesity (Harris, et. al., 2009). If this product targets to children, the binary variable KIDS has value of one, otherwise is zero.

Two other non-nutrient attributes are constructed by using the Nielsen Homsescan data. Since the Homescan data disclose the package size associated to each product purchased, this information can be used to recover the average package size (AVESIZE) by taking the average of the package size observed. Also, unique store name (UNISTORE) is counted for each product in each market, and it can be used to roughly represent the coverage of each cereal brand in each market.

Table 2: Summary statistics of health/non-health related product attributes

Variable	Descriptions	Mean	Std. Dev.
Continuous			
FAT	Total lipid (fat)(g/100 g)	2.99	2.45
SUGAR	Total sugars (g/100 g)	27.25	13.58
SODIUM	Sodium (mg/100 g)/100	6.02	1.93
CALCIUM	Calcium (mg/100 g)/100	1.07	1.41
FIBER	Total dietary fiber (g/100 g)	5.47	3.74
NVF	Number of vitamin fortified	3.13	1.93
AVESIZE	average package size (oz.)	18.65	4.57
UNISTORE	Unique names of store selling cereal products	2.88	0.46
Discrete		Frequency of one	
WG	Product has whole grain as first ingredient		0.42
KIDS	Target to children		0.50
ADFLAVOR	Other flavor ingredient added		0.45
ADCOLOR	Other artificial color added		0.40
FRUIT	Fruit added		0.30

Given product attributes and price, both the Distance Metric measures and own price/expenditure interaction terms are constructed. As mentioned in last section, two types of Distance Metric measures are considered, continuous and discrete. The continuous Distance Metric is constructed to be an *n*-dimensional measure (DM_MULTI). It is defined as the inverse of Euclidean distance by using multiple continuous nutrient variables. Six nutrients are considered in this research, including fiber, sugar, sodium, fat, calcium and number of vitamin fortifications. The second type is discrete distance measures. It includes the Distance Metric for whole grain type (DM_WG), targeting at children (DM_KIDS), flavor added (DM_FLAVOR), and producer (DM_MAKER). If two products belong to the same type, for example both are produced by the same producer or both use whole grain as the first ingredient, then the value of element in the distance metric equals to one.

Finally, the own-price interaction terms are constructed by interacting price (LNP) with fat (FAT), sugar (NVF), number of vitamin fortified (NVF), whole grain type (WG), and fruit added (FRUIT) separately. Expenditure interactions are constructed using the same way as the own-price interactions. Market and year fixed effects are controlled for in the estimation.

Empirical Results

The tobit model shown by equation (6) and (7) is estimated by using Maximum Likelihood Estimation method. For comparison purposes, two samples are estimated: one is the full sample with all households and the other is a more limited sample of households with children. In general, for the model performance, the pseudo R-square is 0.0816 for the full sample and is slightly higher for households with children sample, which is 0.1054. Due to the concern of multicollinearity as mentioned above, Variance Inflation Factor (VIF) is calculated. The average VIF for the full sample is 20.25, and it is 19.7 for the sample of households with children. All the estimated parameters are shown in table 3 below.

Estimated parameters

All the own price and interaction between own price and product attributes are significant at 1% level, but the values are slightly different across these two samples. Households with children show more sensitivity to price change (-0.1611) compared to the average level of full sample (-0.1233). For the interaction terms between own price and product attributes, these two samples show slightly different preference. Both samples show a positive preference for cereals with vitamin fortifications as well as fat, and both show less willingness to pay for the cereals made predominantly of whole grains (-0.1010 and -0.0935 respectively). Similar evidence from previous studies on cereals also showed that fiber and whole grain can have a negative impact on consumer purchase decisions, perhaps because of taste considerations (Binkley and Eales, 2000). For sugar and fruit, households with children show different preferences, on average, than those in the entire sample. Although, on average, households in the entire sample tend to reduce the intake of cereal with higher sugar content as price increase (-0.0024), higher sugar levels has a positive impact on households with children tend (0.0005). This result can possibly be explained by kids' preferences for sweet food. For the fruit, the result is opposite. Households with children show much less interest on cereal with fruit although it is shown with higher willingness to pay for cereal with fruit for the full sample.

Results for the distance metric terms show consumers' response to price changes as cereal products are more competitive in attribute space. The multi-dimensional continuous distance measure (DM_MULTIPLE) has positive and significant parameters across both samples (0.0237 and 0.0179 respectively). This result implies that cereal products that are closer in nutrition profile, including fat, sugar, sodium, calcium, fiber, and number of fortified vitamins, are stronger substitutes. Also, as price increase, consumers are more likely to switch to cereals if they are also made of whole grain (0.0027 and 0.0020 respectively). Cereals that are close in terms of added flavors are also stronger substitutes. For household with children, the estimate for the discrete distance for targeting at kids (DM_KIDS) is positive and significant. This result suggests that households with children tend to switch to another kids cereal if they are motived by a price increase. However, this finding is opposite if we estimate the full sample, which shows a complementary relationship among cereals targeting at kids (-0.0106). Finally, the effect of closeness of manufacturers (DM_MAKER) in the household with child sample suggests that products produced by the same producer tend to complement with each other instead of competing for consumers (-0.0048). This result could be caused by the purpose of diversifying the purchase of cereal for both adults and kids in the household. Additionally, it could suggest that cereal manufacturers have succeeded in positioning their brands so as not to compete with each other. However, this result is not significant in the full sample with negative but insignificant estimates.

Table 3: Estimated parameters for full sample and household with child sample[1]

Sample (N obs)	**All HH (578,880)**			**HH with child (185,500)**		
	Coefficient		**Std. errors**	**Coefficient**		**Std. errors**
HHSIZE	0.0177	***	(0.0007)	0.0053	***	(0.0009)
WHITE	0.0056	**	(0.0023)	0.0112	***	(0.0026)
MARRIED	0.0006		(0.0020)	-0.0014		(0.0031)
HISPANIC	-0.0082	**	(0.0033)	-0.0101	***	(0.0036)
AGEHIGH	-0.0089	***	(0.0005)	-0.0025	***	(0.0007)
HIGHSC	0.0071	***	(0.0021)	0.0039		(0.0030)
SOMECOL	0.0040	**	(0.0018)	0.0062	***	(0.0024)
HHINC	-0.0004	***	(0.0002)	-0.0000		(0.0002)
ADCOLOR	-0.1698	***	(0.0069)	-0.1066	***	(0.0089)
UNISTORE	0.2670	***	(0.0035)	0.1948	***	(0.0045)
AVESIZE	-0.0015	***	(0.0001)	-0.0013	***	(0.0001)
LNP	-0.1233	***	(0.0065)	-0.1611	***	(0.0088)
LNP*FAT	0.0067	***	(0.0008)	0.0135	***	(0.0010)
LNP*SUGAR	-0.0024	***	(0.0001)	0.0005	***	(0.0002)
LNP*NVF	0.0321	***	(0.0010)	0.0187	***	(0.0013)
LNP*WG	-0.1010	***	(0.0040)	-0.0935	***	(0.0054)
LNP*FRUIT	0.0180	***	(0.0015)	-0.0091	***	(0.0019)
DM_MULTIPLE	0.0237	***	(0.0067)	0.0179	**	(0.0084)
DM_WG	0.0027	***	(0.0003)	0.0020	***	(0.0004)
DM_KIDS	-0.0106	***	(0.0014)	0.0100	***	(0.0017)
DM_ADFLAVOR	0.0075	***	(0.0006)	0.0075	***	(0.0007)
DM_MAKER	-0.0010		(0.0015)	-0.0048	**	(0.0019)
LNEP	0.0569	***	(0.0050)	0.1065	***	(0.0068)
LNEP*FAT	-0.0063	***	(0.0007)	-0.0078	***	(0.0008)
LNEP*SUGAR	0.0032	***	(0.0001)	0.0015	***	(0.0002)
LNEP*NVF	-0.0184	***	(0.0009)	-0.0097	***	(0.0012)
LNEP*WG	0.0823	***	(0.0036)	0.0697	***	(0.0048)
COMPANY A	-0.1971	***	(0.0120)	-0.1595	***	(0.0152)
COMPANY B	-0.2404	***	(0.0204)	-0.1657	***	(0.0255)
COMPANY C	-0.3752	***	(0.0382)	-0.1742	***	(0.0479)
D2004	-0.0467	***	(0.0023)	-0.0328	***	(0.0030)
D2005	-0.0267	***	(0.0019)	-0.0168	***	(0.0025)

[1] Estimation result for market fixed effect are excluded due to space limitation

Pseudo R2	0.0816	0.1054
Mean VIF	20.250	19.700

Note: *, **, and *** represent 10, 5 and 1% significance levels, respectively

Among all these estimates for the closeness of products, the nutrition profile plays the largest role of determining the substitution between products. While the closeness of products in whole grain type and flavor has relatively less impact on consumers' switching behaviors, although they have positive estimated coefficients. If we focus on the sample of households with children, closeness of products targeted at kids is shown to be the second most important determinant, in our estimation, on consumers' switching behaviors.

For the expenditure term (LNEP), the estimated results for both two samples consistently show that consumers' cereal purchases will increase as the total expenditure increases (0.0569 and 0.1065 respectively). The purchase of cereal made of whole grain will also increase as the total expenditure of cereal increases (0.0823 and 0.0696). Consumers show a strong interest of sugar, which is reflected by the positive and significant coefficient for the interaction between expenditure and sugar. However, their purchase of cereal with more fat and vitamin fortification will decrease if the total expenditure of cereal increases.

All the three attribute shifters are significant and consistent across samples. On average, added colors and package size have negative impacts on cereal purchases. Larger market coverage, which is approximately represented by larger number of unique retailing stores (UNISTORE), has a positive impact on cereal purchases. Demographic characteristics of households and household heads are specified as demand shifters. Larger families (HHSIZE) tend to have higher expenditure share for those brands they choose. This effect is much larger for the full sample than the households with children, perhaps because the choice of brand is relatively diversified due to the existence of children. Households who are white tend to in purchase more of cereal while Hispanic households purchase less. Education levels of the household head have positive effects on cereal purchases, on average, but this positive effect decreases as education attainment moves from a high school degree to some college. The results for household income and marital status are mixed. All the estimates for married status are not significant across the two samples. Although the estimate for household income is negative and significant for the full sample, it is not significant for the sample of households with children.

Conditional and unconditional elasticity

Given the estimated demand parameters, own- and cross-price elasticities are calculated accordingly. Since households without children tend to mostly buy cereals targeting at family or adults, the estimated cross-price elasticities between kids cereal and family/adult cereals might not be reliable. As a result, we report the elasticities calculated from the sample of households with children[2]. Also, given the censoring feature of consumers' purchase, both the conditional and unconditional elasticities are calculated.

The unconditional price elasticities reflect the overall response of purchase quantity to the changes of price. Following Bonanno (2012), the unconditional own- and cross-price elasticities are calculated as:

$$\eta_{jk}^{U} = \begin{cases} -1 + \dfrac{b_0 + \sum_{m=1}^{M} b_m z_{jm}^{b}}{\overline{w_j}} - (c_0 + \sum_{n=1}^{N} c_n z_{jn}^{c}) & \text{if } \ j = k \\ \dfrac{\lambda_c \delta_{jk}^{C} + \sum_{d=1}^{D} \lambda_d \delta_{jk}^{d}}{\overline{w_j}} - (c_0 + \sum_{n=1}^{N} c_n z_{jn}^{c}) \dfrac{\overline{w_k}}{\overline{w_j}} & \text{if } \ j \neq k \end{cases}$$

where the expenditure share (w), product attributes (z), and distance metric measures (δ) are all taken at the sample average.

The conditional elasticity is used to reflect responsiveness to price changes for consumers who have already purchased those cereal products. Following Yen and Huang (2002), conditional elasticities can be derived by differentiating the conditional expectation of expenditure share. It is shown as:

$$\eta_{jk}^{C} = \begin{cases} -1 + \left[\dfrac{1}{E(w_j / w_j > 0)} - \dfrac{\lambda_j}{\sigma_j} \right] \left[(b_0 + \sum_{m=1}^{M} b_m z_{jm}^{b}) - (c_0 + \sum_{n=1}^{N} c_n z_{jn}^{c}) \overline{w_j} \right] & \text{if } \ j = k \\ \left[\dfrac{1}{E(w_j / w_j > 0)} - \dfrac{\lambda_j}{\sigma_j} \right] \left[(\lambda_c \delta_{jk}^{C} + \sum_{d=1}^{D} \lambda_d \delta_{jk}^{d}) - (c_0 + \sum_{n=1}^{N} c_n z_{jn}^{c}) \overline{w_k} \right] & \text{if } \ j \neq k \end{cases}$$

where $E(w_j / w_j > 0)$ is the conditional expenditure share, σ_j is the standard deviation, and λ_j is inverse Mill's ratio which is calculated as $\lambda_j = [\phi(w_j / \sigma_j)] / [\Phi(w_j / \sigma_j)]$[3], with $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf for standard normal distribution.

[2] Conditional and unconditional price elasticities are also calculated for the full sample. All the own-price elasticities are negative and significant at 1% level. However, most of the cross-price elasticities are also negative. Detailed results are available upon request.
[3] The original equation from Yen and Huang (2002) use estimated expenditure share to calculate the inverse mills ration. However, since the estimated value is not good (with large percent of negative value), we use the sample average of expenditure share to calculate the inverse Mill's ratio.

Table 4: Unconditional Own- and Cross price elasticity for household with child sample

		Kids Cereal (K)										**Family/Adult Cereal (FA)**									
		Company A	**Company B**				**Company C**					**Company A**			**Company B**		**Company C**				**Company D**
		K1	**K2**	**K3**	**K4**	**K5**	**K6**	**K7**	**K8**	**K9**	**K10**	**FA1**	**FA2**	**FA3**	**FA4**	**FA5**	**FA6**	**FA7**	**FA8**	**FA9**	**FA10**
Company A	**K1**	-6.863	0.522	0.421	-0.321	0.859	0.818	0.485	-0.039	-0.261	0.243	-0.168	-0.091	-0.389	-0.843	0.201	-0.162	0.322	-0.047	0.127	0.342
Company B	**K2**	0.315	-0.602	0.123	0.080	0.309	0.307	0.306	0.156	0.324	0.162	0.143	0.143	0.006	0.004	0.065	0.001	-0.003	-0.003	0.005	0.145
	K3	0.420	0.037	-5.293	0.039	0.443	0.399	0.332	0.388	0.460	0.523	-0.156	-0.107	0.195	-1.046	-0.291	0.185	0.066	0.130	0.146	-0.119
	K4	0.081	0.028	0.112	-1.783	0.076	0.072	0.068	0.102	0.098	0.115	-0.004	-0.001	0.047	-0.127	-0.037	0.044	0.035	0.038	0.046	-0.001
	K5	0.794	0.524	0.433	-0.136	-2.159	0.603	0.497	-0.106	0.507	0.023	0.128	0.175	-0.085	-0.506	0.208	-0.305	-0.419	-0.357	-0.343	0.162
Company C	**K6**	0.622	0.394	0.326	-0.147	0.493	-2.478	0.387	-0.109	0.396	-0.002	0.086	0.127	-0.079	-0.456	0.155	-0.253	-0.351	-0.297	-0.285	0.116
	K7	0.485	0.322	0.243	-0.047	0.342	0.330	-1.843	-0.017	0.299	0.026	0.086	0.109	-0.039	-0.261	0.127	-0.168	-0.226	-0.196	-0.180	0.103
	K8	0.126	0.054	0.186	0.063	0.061	0.055	0.053	-1.565	0.047	0.095	-0.047	-0.037	0.067	-0.203	-0.029	0.010	-0.014	-0.002	0.009	-0.040
	K9	1.183	0.980	0.696	0.016	0.794	0.748	0.673	-0.277	-14.901	-0.038	0.438	0.515	-0.216	-0.543	0.566	-0.602	-0.789	-0.677	-0.684	0.513
	K10	0.182	0.110	0.283	0.151	0.096	0.088	0.076	0.139	0.079	-2.283	-0.049	-0.028	0.124	-0.202	-0.021	0.071	0.007	0.050	0.026	-0.035
Company A	**FA1**	0.022	0.053	-0.022	-0.266	0.119	0.102	0.076	-0.238	0.202	-0.159	-5.803	0.259	0.036	-0.009	0.388	0.135	0.090	0.107	0.115	0.367
	FA2	0.028	0.085	-0.025	-0.318	0.161	0.141	0.108	-0.287	0.270	-0.172	0.319	-8.159	0.061	0.068	0.548	0.214	0.111	0.179	0.159	0.503
	FA3	-0.115	-0.121	0.178	-0.085	0.004	-0.012	-0.036	0.040	-0.010	0.123	0.051	0.078	-3.171	-0.149	0.227	0.458	0.393	0.434	0.432	0.229
Company B	**FA4**	0.044	0.018	-0.023	-0.033	0.043	0.043	0.041	-0.012	0.059	-0.005	0.114	0.118	0.058	-1.577	0.090	0.059	0.053	0.057	0.055	0.116
	FA5	0.257	0.050	-0.164	-0.325	0.247	0.236	0.217	-0.167	0.364	-0.090	0.668	0.722	0.332	0.252	-5.194	0.337	0.270	0.320	0.297	0.697
	FA6	0.017	-0.108	0.177	-0.056	-0.121	-0.134	-0.157	-0.068	-0.143	0.054	0.180	0.223	0.460	-0.092	0.233	-3.380	0.265	0.360	0.299	0.206
Company C	**FA7**	0.015	-0.064	0.110	-0.030	-0.071	-0.080	-0.092	-0.040	-0.087	-0.004	0.123	0.121	0.275	-0.074	0.131	0.192	-1.614	0.177	0.187	0.120
	FA8	0.018	-0.071	0.144	-0.018	-0.089	-0.099	-0.114	-0.034	-0.108	0.045	0.152	0.182	0.371	-0.037	0.191	0.309	0.222	-2.310	0.246	0.172
	FA9	0.021	-0.104	0.136	-0.081	-0.100	-0.113	-0.130	-0.076	-0.121	-0.027	0.129	0.145	0.346	-0.162	0.160	0.236	0.192	0.213	-2.486	0.142
CompanyD	**FA10**	0.167	0.096	-0.016	-0.261	0.155	0.138	0.111	-0.241	0.259	-0.151	0.445	0.479	0.225	0.091	0.496	0.193	0.121	0.168	0.162	-6.278

Note: All the own-price elasticities are significant at 1% level.
Among the 380 cross-price elasticities, 297 elasticities are significant at 1% level, 18 are significant at 5% level, and 12 are significant at 10% level.
Due to space limitation, standard errors of elasticity are excluded. Information is available upon request.
Under the agreement for data use, brand name is replaced with artificial name. K represents cereal brands for kids and FA represents brand name for family/adult. Company name is also replaced.

Table 5: Conditional Own- and Cross price elasticity for household with child sample

		Kids Cereal (K)										**Family/Adult Cereal (FA)**									
		Company A	**Company B**			**Company C**						**Company A**			**Company B**		**Company C**				**Company D**
		K1	**K2**	**K3**	**K4**	**K5**	**K6**	**K7**	**K8**	**K9**	**K10**	**FA1**	**FA2**	**FA3**	**FA4**	**FA5**	**FA6**	**FA7**	**FA8**	**FA9**	**FA10**
Company A	**K1**	-1.268	0.024	0.019	-0.015	0.039	0.037	0.014	-0.001	-0.008	0.007	-0.005	-0.004	-0.018	-0.039	0.009	-0.007	0.009	-0.001	0.004	0.010
Company B	**K2**	0.025	-0.968	0.010	0.006	0.025	0.024	0.024	0.012	0.026	0.013	0.011	0.011	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.012
	K3	0.020	0.002	-1.208	0.002	0.021	0.019	0.016	0.019	0.022	0.025	-0.008	-0.005	0.009	-0.051	-0.014	0.009	0.003	0.006	0.007	-0.006
	K4	0.011	0.004	0.015	-1.105	0.010	0.010	0.009	0.014	0.013	0.015	0.000	0.000	0.006	-0.017	-0.005	0.006	0.005	0.005	0.006	0.000
Company C	**K5**	0.030	0.020	0.016	-0.005	-1.043	0.023	0.019	-0.004	0.019	0.001	0.005	0.007	-0.003	-0.019	0.008	-0.011	-0.016	-0.013	-0.013	0.006
	K6	0.028	0.017	0.014	-0.006	0.022	-1.065	0.017	-0.005	0.018	0.000	0.004	0.006	-0.003	-0.020	0.007	-0.011	-0.016	-0.013	-0.013	0.005
	K7	0.014	0.009	0.007	-0.001	0.010	0.010	-1.024	0.000	0.009	0.001	0.002	0.003	-0.001	-0.008	0.004	-0.005	-0.007	-0.006	-0.005	0.003
	K8	0.010	0.004	0.014	0.005	0.005	0.004	0.004	-1.043	0.004	0.007	-0.004	-0.003	0.005	-0.016	-0.002	0.001	-0.001	0.000	0.001	-0.003
	K9	0.111	0.092	0.065	0.002	0.074	0.070	0.063	-0.026	-2.301	-0.004	0.041	0.048	-0.020	-0.051	0.053	-0.056	-0.074	-0.063	-0.064	0.048
	K10	0.013	0.008	0.020	0.011	0.007	0.006	0.005	0.010	0.006	-1.092	-0.004	-0.002	0.009	-0.015	-0.002	0.005	0.000	0.004	0.002	-0.003
Company A	**FA1**	0.004	0.009	-0.004	-0.045	0.020	0.017	0.013	-0.040	0.034	-0.027	-1.813	0.044	0.006	-0.001	0.066	0.023	0.015	0.018	0.020	0.062
	FA2	0.009	0.028	-0.008	-0.105	0.053	0.047	0.036	-0.095	0.089	-0.057	0.105	-3.364	0.020	0.022	0.181	0.070	0.037	0.059	0.052	0.166
	FA3	-0.016	-0.016	0.024	-0.012	0.001	-0.002	-0.005	0.005	-0.001	0.017	0.007	0.011	-1.297	-0.020	0.031	0.063	0.054	0.059	0.059	0.031
Company B	**FA4**	0.009	0.004	-0.005	-0.007	0.009	0.009	0.009	-0.003	0.012	-0.001	0.024	0.024	0.012	-1.119	0.019	0.012	0.011	0.012	0.011	0.024
	FA5	0.073	0.014	-0.047	-0.093	0.071	0.067	0.062	-0.048	0.104	-0.026	0.191	0.207	0.095	0.072	-2.200	0.096	0.077	0.092	0.085	0.199
Company C	**FA6**	0.003	-0.018	0.029	-0.009	-0.020	-0.022	-0.026	-0.011	-0.024	0.009	0.030	0.037	0.076	-0.015	0.038	-1.392	0.044	0.059	0.049	0.034
	FA7	0.002	-0.010	0.017	-0.005	-0.011	-0.012	-0.014	-0.006	-0.013	-0.001	0.019	0.019	0.042	-0.011	0.020	0.029	-1.094	0.027	0.029	0.018
	FA8	0.005	-0.020	0.041	-0.005	-0.025	-0.028	-0.032	-0.009	-0.030	0.013	0.043	0.051	0.104	-0.010	0.054	0.087	0.063	-1.369	0.069	0.048
	FA9	0.007	-0.034	0.045	-0.027	-0.033	-0.038	-0.043	-0.025	-0.040	-0.009	0.043	0.048	0.115	-0.054	0.053	0.078	0.064	0.071	-1.492	0.047
Company D	**FA10**	0.025	0.014	-0.002	-0.039	0.023	0.021	0.017	-0.036	0.039	-0.023	0.067	0.072	0.034	0.014	0.074	0.029	0.018	0.025	0.024	-1.792

Note: All the own-price elasticities are significant at 1% level.
Among the 380 cross-price elasticities, 298 elasticities are significant at 1% level, 18 are significant at 5% level, and 11 are significant at 10% level.
Due to space limitation, standard errors of elasticity are excluded. Information is available upon request.
Under the agreement for data use, brand name is replaced with artificial name. K represents cereal brands for kids and FA represents brand name for family/adult. Company is also replaced

Table 4 and 5 listed the unconditional and conditional price elasticities separately. All the conditional and unconditional own-price elasticities are negative and significant at 1% level. The unconditional own-price elasticities are larger than the conditional own-price elasticities, with average value of -3.987 and -1.453 respectively. This finding indicates that although consumers are price elastic in general, existing consumers tend to be less sensitive to price change than those new consumers who did not purchase the product before. Among the twenty products, only one product has own-price elasticity less than unity and the rest are all larger than unity.

Both the conditional and unconditional cross-price elasticities show consistent substitution (or complementary) relationships. Among 380 cross-price elasticities, 327 cross-price elasticities are statistical significant under 1%-10% level. The results of cross-price elasticities are mixed, with either positive value showing the substitution or negative values showing complementary relationships between two different products. However, results show that if two cereal products are both targeted for kids (or are both targeted as family/adult cereals), they tend to be stronger substitutes in most of the cases (80 of 90 cross-price elasticities between two kids cereal are positive; 85 of 90 cross-price elasticities between two family/adult cereals are positive). Kids cereals are more likely to be complements with family/adult cereals, with about 50 percent of cross-price elasticities being negative. This finding suggests that households may buy a portfolio of cereals to meet their needs. Both the conditional and unconditional elasticities show consistent patterns, but unconditional cross-price elasticities have larger values than conditional cross-price elasticities on average.

Conclusions

The objective of this paper is to evaluate the impact and importance of health-related product attributes on consumers' purchasing behaviors in ready-to-eat cereal market. As far as we know, it is the first paper to apply the Distance Metric method to the demand estimation using micro-level purchase data, and the first to incorporate it within a censored model.

The micro-level data highlights two major benefits of the DM method: First, it solves the dimensionality problems by converting demand estimation from product space into product attribute space and defining the cross-price relation between products as a function of distance of several key product attributes. In addition, due simplifications employed with the DM method, the system of demand equations is reduced to be only one estimated equation. The second benefit means that the DM method can be easily incorporated into a censored model, thereby helping us avoid the calculation of multiple integrals when dealing with zero purchases for each consumer for some cereal brands. Because of these two benefits, this paper is able to handle the demand estimation for large number of differentiated cereal products and deal with the censored feature of consumer purchase with much less computation burden than precious attempts.

The empirical results show that health-related products do impact consumer purchasing decisions. When price changes, products with similar nutrition profiles, or products with whole grain as the first ingredient tend to be strong substitutes. Also, if we focus on the sub sample for households with children, products with different targeted markets appear to be a strong factor that affects consumers' purchasing decisions. Consumers tend to switch between products that are both kids cereal or both family/adult cereals. On the other hand, two products tend to be complementary if they come from kids cereal and family/adult cereal separately. This result is echoed by cross-price elasticity estimates. Both the conditional and unconditional cross price elasticities are more likely to be positive for products with the same market target, and they are more likely to be negative if they are for two different consumer groups.

We acknowledge that endogeneity issues might exist in our demand estimations. However, since our objective of this paper is to apply the Distance Metric method into the demand estimation using micro-level purchase data and we have rich information of health/unhealthy related product characteristics, for the moment we set aside worries over potential endogeneity. However, further tests and modifications of specification will be considered in the future. Despite this issue, our estimation still shows reasonable outcomes for both estimated parameters and the own- and cross-price elasticities.

Reference

Amemiya, T. 1973. "Multivariate Regression and Simultaneous Equation Models when the Dependent Variables are Truncated Normal." *Econometrica* 42(6): 999-1012.

Amemiya, T. 1984. "Tobit Models: A Survey." *Journal of Econometrics* 24:3-61.

Barreiro-Hurlé, J., A. Gracia, and T. de-Magistris. 2010. "Does Nutrition Information on Food Products Lead to Healthier Food Choices?" *Food Policy* 35:221–229.

Barreiro-Hurlé, J., A. Gracia, and T. de-Magistris. 2010. "The Effects of Multiple Health and Nutrition Labels on Consumer Food Choices." *Journal of Agricultural Economics*, 61(2): 426–443.

Berning, J.P., H.H. Chouinard, and J.J. Mccluskey. 2011. "Do Positive Nutrition Shelf Labels Affect Consumer Behavior? Findings from a Field Experiment with Scanner Data." *American Journal of Agricultural Economics* 93(2): 364–369.

Binkley, J.K., and J. Eales. 2000. "Demand for High Fiber and Low Fiber Cereals." Presented on 2000 Annual Meeting of American Agricultural Economics Association, July 30-August 2, Tampa, FL with number 21826.

Binkley, J.K., and A. Golub. 2010. "Consumer Demand for Nutrition versus Taste in Four Major Food Categories." *Agricultural Economics* 42:65–74.

Bonanno, A. 2012. "Functional Foods as Differentiated Products: the Italian Yogurt Market." *European Review of Agricultural Economics* pp:1–27.

Brown, D.J., and L.F. Schrader. 1990. "Cholesterol Information and Shell Egg Consumption." *American Journal of Agricultural Economics* 72(3):548-555.

Capps, O., and J.D. Schmitz. 1991. "A Recognition of Health and Nutrition Factors in Food Demand Analysis." *Western Journal of Agricultural Economics* 16(1):21-35.

Chern, W.S., E.T. Loehman, and S.T. Yen. 1995. "Information, Health Risk Beliefs, and the Demand for Fats and Oils." *The Review of Economics and Statistics* 77(3):555-564.

Chidmi, B., and R.A. Lopez. 2007. "Brand-Supermarket Demand for Breakfast Cereals and Retail Competition." *American Journal of Agricultural Economics* 89(2): 324–337.

Chowdhury, S., J.V. Meenakshi, K.I. Tomlins, and C. Owori. 2011. "Are Consumers in Developing Countries Willing to Pay More for Micronutrient-Dense Biofortified Foods? Evidence from a Field Experiment in Uganda." *American Journal of Agricultural Economics* 93(1): 83–97.

Coulson, N.S. 2000. "An Application of the Stages of Change Model to Consumer Use of Food Labels." *British Food Journal* 102 (9): 661–668.

Deaton, A., and J. Muellbauer. 1980. "An Almost Ideal Demand System." *The American economic review* 70(3):312-315.

Drichoutis, A.C., P. Lazaridis, and R.M. Nayga. 2005. "Nutrition Knowledge and Consumer Use of Nutritional Food Labels." *European Review of Agricultural Economics* 32 (1):93–118.

Golub, A., and J.J. Binkley. 2005. "Consumer Choice of Breakfast Cereals." RUL: http://www.aae.wisc.edu/fsrg/web/FSRG%20papers/18b%20Golub%20Binkley.pdf accessed 01/15/2012.

Gorman, W. M.: Two Stage Budgeting. Unpublished paper, London School of Economics, 1971.

Govindasamy, R., and J. Italia. 2000. "The Influences of Consumer Demographic Characteristics on Nutritional Label Usage." *Journal of Food Products Marketing* 5(4):55-68.

Harris, J.L., M.B. Schwartz, K.D. Brownell, V. Sarda, M. Weinberg, S. Speers, J. Thompson, A. Ustjanauskas, A. Cheyne, E. Bukofzer, L. Dorfman, and H. Byrnes-Enoch. 2009. "Evaluating the Nutrition Quality and Marketing of Children's Cereals" Rudd Center for Food Policy and Obesity RUL: http://www.cerealfacts.org/media/Cereal_FACTS_Report_2009.pdf accessed 09/01/2012

Harris, J.L., M.B. Schwartz, K.D. Brownell, V. Sarda, C. Dembek, C. Munsell, C. Shin, A. Ustjanauskas, and M. Weinberg. 2012. "Limited Progress in the Nutrition Quality and Marketing of Children's cereals" Rudd Center for Food Policy and Obesity. URL: http://www.cerealfacts.org/media/Cereal_FACTS_Report_2012_7.12.pdf accessed 01/15/2013.

Hausman, J., G. Leonard, and J.D. Zona. 1994. "Competitive Analysis with Differentiated Products." *Annales d'Economie et de Statistique* 34:159–180.

Huffman, S.K., J.H. Helen. 2004. "Demand for Enhanced Foods and the Value of Nutritional Enhancements of Food: The Case of Margarines." 2004 Annual Meeting of American Agricultural Economics Association, Aug 1-4, Denver, CO, with number 20205.

Ippolito, P.M., and A.D. Mathios. 1989. "Health Claims in Advertising and Labeling, A Study of the Cereal Market." URL: http://www.ftc.gov/be/econrpt/232187.pdf accessed 03/15/2012.

Kasteridis, P., S.T. Yen, and C. Fang. 2011. "Bayesian Estimation of A Censored Linear Almost Ideal Demand System: Food Demand in Pakistan." *American Journal of Agricultural Economics* 93(5): 1374–1390

Kim, D.K., and W.S. Chern. 1995. "Health Risk Concern of Households vs. Food Processors: Estimation of Hedonic Prices in Fats and Oils." URL: http://ageconsearch.umn.edu/bitstream/25969/1/bkvp2c8.pdf accessed 04/01/2012.

Kim, S.Y., R. Nayga, and O. Capps. 2001. "Food Label Use, Self-Selectivity, and Diet Quality." *Journal of Consumer Affairs* 35 (2): 346–363.

Kim, S.Y., R. Nayga, and O. Capps. 2001." Health Knowledge and Consumer Use of Nutritional Labels: The Issue Revisited." *Agricultural and Resource Economics Review* 30(1):10-19.

Kinnucan, H.W., H. Xiao, C.J. Hsia, and J.D. Jackson. 1997. "Effects of Health Information and Generic Advertising on U.S. Meat Demand." *American Journal of Agricultural Economics* 79(1):13-23.

Lancaster, K. 1966. "A New Approach to Consumer Theory." *Journal of Political Economy* 74:132-157.

Lancaster, K. 1979. *Variety, Equity, and Efficiency*. New York, NY: Columbia University Press.

Lenz, J.E., R.C. Mittelhammer, and H. Shi. 1994. "Retail-Level Hedonics and Valuation of Milk Components." *American Journal of Agricultural Economics* 76(3):492-503.

Levy, A.S., S.B. Fein, and R.E. Schucker. 1996. "Performance Characteristics of Seven Nutrition Label Formats." *Journal of Public Policy & Marketing* 15(1):1-15.

Li, Q., J.J. McCluskey, and T.I. Wahl. 2004. "Effect of Information on Consumers' Willingness to Pay for GM-Corn-Fed Beef." *Journal of Agricultural and Food Industrial Organization* Vol 2: Article 9.

McLean-Meyinsse, P.E. 2001. “An Analysis of Nutritional Label Use in the Southern United States.” *Journal of Food Distribution Research* 32(1): 110–114.

McGuirk, A., P. Driscoll, J. Alwang, and H. Huang. 1995. “System Misspecification Testing and Structural Change in Demand for Meats.” *Journal of Agricultural and Resource Economics* 20(1): 1–21.

Moschini, G. 1995. “Units of Measurement and the Stone Index in Demand System Estimation.” *American Journal of Agricultural Economics* 77(1): 63-68.

Muth, M.K., C.Zhen, J. Taylor, S. Cates, K.M. Kosa, D. Zorn, and C.J. Choiniere. “The Value to Consumers of Health Labeling Statements on Breakfast Foods and Cereals.” the International Association of Agricultural Economists Conference, Beijing, China, August 16-22, 2009

National Institutes of Health (2004), “Weight Loss and Nutrition Myths: How much Do You Really Know?” NIH Publication No. 04-4561, Bethesda, MD.

Nayga, R.M. 1996. “Determinants of Consumers' Use of Nutritional Information on Food Packages.” *Journal of Agricultural and Applied Economics* 28(2):303–312.

Nevo, A. 2001. “Measuring Market Power in the Ready-to-Eat Cereal Industry.” *Econometrica*, 69(2):307–342.

Pinkse, J., and M.E. Slade. 2004. “Mergers, Brand Competition, and the Price of a Pint” *European Economic Review* 48:617–643.

Pinkse, J., M.E. Slade, and C. Brett. 2002. “Spatial Price Competition: A Semiparametric Approach.” *Econometrica* 70(3):1111-1153.

Pirovano, T. 2010. “U.S. Healthy Eating Trends Part 1 Commitment Trumps the Economic Pinch.” URL: http://www.nielsen.com/us/en/newswire/2010/healthy-eating-trends-pt-1-commitment-trumps-the-economic-pinch.html. Accessed 03/20/2013.

Pofahl, G.M., and T.J. Richards. 2009. “Valuation of New Products in Attribute Space.” *American Journal of Agricultural Economics* 91(2): 402–415.

Rojas, C. 2008. “Price Competition in U.S. Brewing.” *The Journal of Industrial Economics* 56(1):1–31.

Rojas, C., and E.B. Peterson. 2008. “Demand for Differentiated Products: Price and Advertising Evidence from the U.S. Beer Market.” *International Journal of Industrial Organization* 26: 288–307.

Scherer, F.M. 1979. “The Welfare Economics of Product Variety: An Application to the Ready-to-Eat Cereals Industry.” *The Journal of Industrial Economics* 28(2):113-134.

Schmalensee, R. 1978. “Entry Deterrence in the Ready-to-Eat Breakfast Cereal Industry.” *The Bell Journal of Economics* 9(2):305-327.

Shepherd, R., P. Sparks, S. Belier, and M.M. Raats. 1992. "The Effects of Information on Sensory Ratings and Preferences: The Importance of Attitudes." *Food Quality and Preference* 3:147–155.

Shi, H., and D.W. Price. 1998. "Impacts of Sociodemographic Variables on the Implicit Values of Breakfast Cereal Characteristics." *Journal of Agricultural and Resource Economics* 23(1):126-139.

Slade, M.E. 2004. "Market Power and Joint Dominance in U.K. Brewing." *The Journal of Industrial Economics* 52(1):133-163.

Thunström, L. 2010. "Preference Heterogeneity and Habit Persistence: The Case of Breakfast Cereal Consumption." *Journal of Agricultural Economics* 61(1):76–96.

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

United States of Agricultural Department, Center for Nutrition Policy and Promotion. "The Dietary Guidelines for Americans in 2010." URL: http://www.cnpp.usda.gov/dgas2010-policydocument.htm accessed 04/01/2012.

Variyam, J.N., J. Blaylock, B.H. Lin, K. Ralston, and D. Smallwood. 1999. "Mother's Nutrition Knowledge and Children's Dietary Intakes." *American Journal of Agricultural Economics* 81(2):373-384.

Wansink, B. 2003. "How Do Front and Back Package Labels Influence Beliefs about Health Claims?" *Journal of Consumer Affairs* 37(2):305–316.

Wansink, B., and P. Chandon. 2006. "Can "Low-Fat" Nutrition Labels Lead to Obesity?" *Journal of Marketing Research* 43:605–617.

Wansink, B., and S.B. Park. 2002. "Sensory Suggestiveness and Labeling: Do Soy Labels Bias Taste?" *Journal of Sensory Studies* 17(5):483–491.

Wardle, J., and G. Huon. 2000. "An Experimental Investigation of the Influence of Health Information on Children's Taste Preferences." *Health Education Research: Theory and Practice* 15(1): 39–44.

Yen, S.T., and Huang. 2002. "Cross-Sectional Estimation of U.S. Demand for Beef Products: A Censored System Approach" *Journal of Agricultural and Resource Economics* 27(2):320-334.

Yen, S.T., and Lin. 2008. "Quasi-maximum Likelihood Estimation of a Censored Equation System with a Copula Approach: Meat Consumption by U.S. Individuals." *Agricultural Economics* 39 (2008): 207–217