Intern. J. of Research in Marketing 25 (2008) 301–309

Contents lists available at ScienceDirect

Intern. J. of Research in Marketing

journal homepage: www.elsevier.com/locate/ijresmar

What can grocery basket data tell us about health consciousness?☆

Ashutosh Prasad, Andrei Strijnev, Qin Zhang *

School of Management, University of Texas at Dallas SM 32 Box 830688 Richardson, TX 75083-0688, United States

ARTICLE INFO

Article history:
First received in 3 September 2007
and was under review for 3 months

Area Editor: Harald Van Heerde

Keywords:
Health conscious
Marketing
Choice models
Bayesian estimation

ABSTRACT

Health-conscious consumers are a sought-after market segment by manufacturers and retailers alike. But how large is the health-conscious segment? How price sensitive is it? In addition, what are the influences of consumer demographic characteristics on a consumer's health consciousness? To answer these questions, we control for covariates such as price, distinguish health consciousness from intrinsic preferences, and assess purchases over multiple categories with multiple nutritional attributes. We estimate a multi-category brand choice model using purchase history of a large sample of households (1062) in ten commonly purchased grocery categories. We find that health-conscious households constitute 18% of the market and that the more health conscious a household is, the less price sensitive it is. We also show that the following household demographic characteristics have strong impacts on a household's health consciousness: household income, house ownership, employment status of male household head, education level of male household head, presence of young children in the household, and the ethnicity of the household.

1. Introduction

Marketers of food items, both retailers and manufacturers, have been eager to position themselves as health friendly in response to a perception of growing health consciousness among consumers. Leeflang and van Raaij (1995) note a clear trend about concern for health and an increase in the demand for, and supply of, health products. They state that "European consumers are very concerned about the nutritional content of food. This is in line with the tendency towards increased health consciousness." In the US, a survey showed that 74% of consumers claimed to have changed their eating habits due to health and nutrition concerns, 87% considered nutrition to be an important factor in purchasing groceries and 64% stated a willingness to pay more for healthier versions of food items (USB, 2005).

In response, retailers have been stocking (and manufacturers have been producing) diet, low-carb, low-calorie, low-sodium, low-fat, low-cholesterol, caffeine-free, no trans-fat, vitamin-enhanced, calcium-added, high-fiber, natural and organic products as never before. For example, PepsiCo acquired the healthier Quaker and Tropicana brand products and then announced plans to target half of its new foods and beverages to health-conscious consumers (Horovitz, 2003). Its line of "Smart Snack" Lay's products is low-fat and low-calorie and has zero trans-fats. Competitor Coca-Cola introduced Enviga, a health drink containing green-tea extracts, caffeine and 'plant micronutrients', for weight-conscious consumers (Economist, 2007). Kraft Foods launched a Fun Fuel Lunchables line where yogurt and fruit juice replaced candy and soda (Barboza, 2003). Retailers such as Wal-Mart have asked manufacturers to supply low-calorie products exclusively for their shelves (Thompson, 2003). Starbucks, Wendy's and even McDonalds are transitioning to zero trans-fat offerings. These are just a few examples of the response of marketers to offering healthier alternatives that appeal to consumer demand.

With billions of food dollars at stake, it is important to cater to health-conscious consumers. However, the power of the consumer should not be overestimated. The food industry learned a lesson from relying on fads such as the low-carb diet in the early 2000s. Demand for low-carb products in 2005 dropped 20% below the preceding year's level, hurting sellers of exclusively low-carb products; e.g., Atkins Nutritionals Inc. filed for bankruptcy that year (Condie, 2005; Wolk, 2005). Despite the hype given to dieting, the proportion of overweight adults keeps increasing, indicating that many are not health conscious in their consumption habits (Cutler, Glaeser & Shapiro, 2003).

Although marketers require accurate measurements of the size and characteristics of the health-conscious segment for making product launch and targeting decisions, little academic attention has been devoted to this topic. This paper contributes by showing that readily available scanner panel data on household purchases in food categories can provide information for identifying health-conscious

☆ Authors are faculty of Marketing at the University of Texas at Dallas. The paper has benefited from comments from the editor Stefan Stremersch, the AE and two anonymous reviewers, and seminar participants at the Marketing Science Conference 2005 in Atlanta and the Informs International Conference 2006 in Hong Kong. Authors are listed in alphabetical order.
* Corresponding author. Tel.: +1 972 883 6525; fax: +1 972 883 6727.
E-mail address: zhangq@utdallas.edu (Q. Zhang).

0167-8116/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijresmar.2008.05.001

households. The research questions are as follows: How large is the health-conscious segment? How price sensitive is it? And what are the influences of consumer demographic characteristics on their health consciousness?[1]

Products have multiple nutritional attributes. We define *health-conscious* behavior as a consistent preference for healthy versions of products across categories. This preference needs to be distinguished from the consumer's intrinsic preference for a particular nutritional attribute, e.g., a taste preference for fatty food. Conditional on other things (e.g., brand and marketing mix) remaining the same, whether a consumer chooses a healthy or unhealthy product in a category is determined by a combination of two effects. The first effect is the health consciousness that is a consumer trait common across nutritional attributes. The second effect is the consumer's intrinsic preference for the individual nutritional attributes (e.g., due to taste preferences). The requirements to control for covariates, such as price, to distinguish health consciousness from taste preferences, and to assess purchases over multiple categories with multiple nutritional attributes by combining the data from many individuals are some of the reasons why a raw data approach does not work and a more in-depth analysis is needed.

In order to identify consumers' health consciousness as a consistent consumer trait across nutritional attributes, one needs to examine consumers' purchase behavior jointly across multiple nutritional attributes (e.g., sugar, fat and fiber attributes) in multiple categories. To accomplish this, we use a multi-category, attribute-based choice model of the household's purchases and calibrate it on a household purchase dataset in 10 categories over a 53-week period. In the choice model, the households' utilities for healthy nutritional attributes are decomposed into two components: one that is treated as a common consumer trait across multiple nutritional attributes, and one that is specific to each of the nutritional attributes. We apply a random effects model to account for household heterogeneity and calibrate it using Bayesian estimation techniques. This yields the size of the health-conscious segment and the relationship between a household's health consciousness and its price sensitivities. Furthermore, we evaluate the effects of household characteristics such as demographics on the health consciousness traits of a household. This research should be of help to firms in the food industry as they develop better marketing strategies to target health-conscious grocery shoppers.

In the next section, we review the literature. In Section 3, the model is presented and Section 4 discusses the empirical analysis. Section 5 provides the conclusion.

2. Literature review and contribution

In an extensive review of the cross-disciplinary literature on preventive health behaviors, Moorman and Matulich (1993) distinguish between *health information acquisition* behaviors and *health maintenance* behaviors. The former refers to the acquisition of health benefits information through various sources such as media and labels, and friends and family. The present study, however, will focus on the latter, i.e., health maintenance behaviors. Specifically, this study focuses on grocery store food purchases. This is because we do not have information about consumers' information sources and because there were no disruptive informational events during the data availability period (such as the lift of the ban on advertising health claims for food products studied in Ippolito and Mathios, 1990). The domain of applicability of our study, therefore, is restricted to mature categories where the health benefits are established and salient, that is, where there is minimum information cost of acquiring and processing nutrition information (this cost having been shown to be an impediment to inducing healthier purchase by, e.g., Andrews, Netemeyer, & Burton, 1998; Brucks, Mitchell, & Staelin, 1984; Russo, Staelin, Nolan, Russell, & Metcalf, 1986). For example, with respect to the carbonated beverage and milk categories, when examining cans of regular cola and diet cola, it is very salient to the consumer which one has less sugar in it even without needing to read the nutrition labels, and likewise the distinction between regular and low-fat milk is obvious from the front packaging.

We consider four nutritional attributes, sugar, caffeine, fat and fiber. It should be explained why, in this paper and generally in the literature (e.g., Russo, et al., 1986; Ippolito & Mathios, 1990), that low-sugar, low-caffeine, low-fat, and high-fiber products are considered to be healthier. The US government recommends for an adult a daily intake of 65 g of fat and 300 g of carbohydrates based on a 2000 calorie diet. As discussed in Garille and Gass (2001), a remarkably simple diet can achieve these dietary requirements. But the normal diet is complex, with processed, convenience and fast foods. Thus, adherence to the recommended daily intake is difficult (e.g., Chandon & Wansink, 2007). While getting either less or more than the recommended fat and sugar (a carbohydrate) consumption can be unhealthy, the problem in developed societies is primarily one of over-consumption and not malnutrition (Cutler, et al., 2003). Over-consumption contributes to obesity, heart disease, diabetes and other illnesses.

Similarly, caffeine in moderate amounts increases alertness, but over-consumption can lead to insomnia, restlessness, hypertension and other problems (e.g., Palmer, Rosenberg, Rao, & Shapiro, 2008). Coffee consumption can become addictive (e.g., Olekalns & Bardsley, 1996), while consumption of decaffeinated coffee is associated with more healthy behaviors (Leviton & Allred, 1994) and therefore has a perception of being healthier. Besides coffee, carbonated beverages are among the largest sources of caffeine in consumers' diets, and a 12 oz. can of cola contains 45 mg of caffeine.

Regarding fiber, Ippolito and Mathios (1990) show that information about the health benefits of fiber in ready-to-eat cereal had an impact on choice. A dietary intake of 20–35 g of fiber is considered healthy, but fiber tends to be under-consumed in modern diets (Marlett, McBurney, and Slavin, 2002). Thus, fiber is a positive nutrient, whereas fat, sugar and caffeine are negative nutrients. Negative nutrients are especially salient to the consumer and likely to impact behavior (Russo, et al., 1986).

Moorman and Matulich (1993) also distinguish between two types of consumer characteristics. The first, denoted health motivation, overlaps with our notion of health consciousness and represents the willingness and interest in performing healthy behaviors. The second, denoted health ability, refers to consumer resources, skills and proficiencies and is captured in our study through the inclusion of demographic variables such as age, income and employment. We shall compare the effects of these variables with those hypothesized in the literature.

Two features distinguish this paper from the existing literature on health consciousness in food choice. The first is the use of scanner panel data on household purchases in multiple categories as the data source for the analysis, and the second is the application of a multi-category, attribute-based, individual level, choice model. We discuss these features next.

To date, the main instrument for measuring consumer attitudes and behaviors on health and nutrition has been surveys. These data can be primary (e.g., Andrews, et al., 1998; Moorman & Matulich, 1993) or secondary, such as through government reports (e.g., Ippolito & Mathios, 1990; Divine & Lepisto, 2005) provided by the

[1] Note that the grocery basket panel data we use are at the household level, so the results pertain to the household rather than the individual consumer level, though we use the term “consumers” for conceptual development in the introduction and literature review sections. The distinction is less important if one household member is primarily responsible for food purchases — their health consciousness then dictates that of the household. Future research should examine this issue.

Bureau of Labor Statistics (www.bls.gov), Center for Disease Control (www.cdc.gov/brfss/), and USDA. Gould (1990), for example, develops a scale to measure a person's health consciousness and relates this scale to a number of health-associated behavioral variables. Despite its widespread use, self-explication often provides a less reliable measure of preference than observing actual purchase behavior. Surveys about nutritional food preferences risk inflating the number of health-conscious consumers because of demand bias — the respondent knows that the most desirable social, societal and medical response is to claim to be health conscious, and the respondent states preferences accordingly (Lichtman et al., 1992). In addition, the respondent might state a truthful opinion but not act on it when making food purchases (Russo et al., 1986). For these reasons, purchase information is preferable for determining the number of health-conscious consumers and the influences of consumer characteristics. In their study of whether nutrition information influences behavior, Russo et al. (1986) conducted shopping basket audits and conducted individual interviews. However, their analysis is at an aggregate level. Consumer purchase data is now easily available to marketers, but there has been little research based on actual consumer purchase data regarding consumers' health concerns with nutrition.

The second novel aspect of our study is that we apply a household level, attribute-based, multi-category choice model; i.e., we simultaneously determine the utility that consumers place on healthy and non-healthy attributes across multiple nutritional attributes. In contrast, extant research has looked at single category analysis (e.g., Ippolito & Mathios, 1990) or considered entire categories as being healthy or non-healthy. While it may be possible that some categories, such as cigarettes and multivitamin supplements, may be characterized as healthy or non-healthy, this is clearly not applicable to all categories. Divine and Lepisto (2005) improve on the assumption by making the healthiness of the category depend on the number of servings/month. Yet this too does not allow for the possibility that some choices within the category are healthier than others.

We use an approach similar to the variance components approach by Ainslie and Rossi (1998), who decompose a consumer's responses to the marketing mix into two components: one specific to the consumer and common across categories; and the second specific to both the consumer and the category. In our approach, we decompose the “healthy” utility component of a product into two parts: one that is consumer specific and common across all nutritional attributes; and the second that is also consumer specific but unique to each nutritional attribute. The main focus of Ainslie and Rossi (1998), and other multi-category brand choice models in the literature, is on the correlations of consumers' sensitivities to marketing mix or on the correlations of consumer preferences for product attributes across categories (e.g., Singh, Hansen, & Gupta, 2005; Hansen, Singh, & Chintagunta, 2006 etc.). For a detailed review on multi-category brand choice models, see Seetharaman et al. (2005). In this study, we are more interested in the component that is common across nutritional attributes. Our study is distinct from Singh et al. (2005), who study consumers' purchase behavior in multiple categories that share the same nutritional attribute (e.g., fat is a common nutritional attribute that is shared by mayonnaise category and cheese category), since we consider multiple nutritional attributes (e.g., sugar, caffeine, fat and fiber) within and across categories to obtain a trait for health-related behavior.

3. Model

Consider a market with C categories and I households. We represent the number of choice alternatives in a category c by J_c. During trip t, a household i decides which product to purchase. We allow for households to not make a purchase in the category, and we treat the no-purchase option as an additional outcome (Chintagunta, 2002). We denote the indirect utility of household i for product j in category c at trip t by U_{itcj}, and the indirect utility for the no-purchase option by U_{itc0}. Since our objective is to study households' health consciousness across multiple nutritional attributes, we adopt an attribute utility model (Fader & Hardie, 1996). The utility is specified as follows:

$$U_{itcj} = \alpha_{ic}^{'}B_{cj} + \beta_{ic}^{'}X_{tcj} + \tilde{\gamma}_{iH}^{'}\text{Healthy}_{cj} + e_{itcj} \quad (1)$$

$$U_{itc0} = \phi_{ic}\text{INV}_{ict} + e_{itc0} \quad (2)$$

We observe that product j in category c is chosen at trip t when $U_{itcj} = \max_k(U_{itck}), k = 0, 1, \ldots, J_c$.[2] We assume that e_{itc} is normally distributed with a mean of zero and a variance matrix equal to an identity matrix with dimension of J_c+1[3]. This generates a standard multinomial probit model with an additional no-purchase outcome for modeling household decisions within category c. As in Ainslie and Rossi (1998), we do not model the potential inter-dependence between purchases in different categories (e.g., the complement and substitute effects as in Manchanda, Ansari and Gupta (1999) and Chib, Seetharaman and Strijnev (2002)), but we do link the separate brand choice models (with no-purchase option) through the coefficients, β_{ic}, $\tilde{\gamma}_{iH}$ (see the discussion later in the section).

B_{cj} is a $W_c \times 1$ indicator vector for the brand name of product j in category c, and W_c is the total number of brands in the category. X_{tcj} is a $n \times 1$ vector of marketing mix variables at trip t for the product j in category c, which may include its price, display, and feature. α_{ic} and β_{ic} are the vectors of corresponding coefficients for B_{cj} and X_{tcj}.

Healthy_{cj} is an $H \times 1$ indicator vector variable of whether the nutritional attributes that product j in category c possesses are healthy or not, where H is the total number of nutritional attributes across all C categories. Including the Healthy_{cj} variable in the utility of the brands makes the decision to buy in the category dependent on the healthiness of the available brands. The value of each element in Healthy_{cj} is set to 1 if the product j contains the corresponding nutritional attribute and its level is considered healthy, and 0 otherwise. $\tilde{\gamma}_{iH}$ is a $H \times 1$ vector in which each element measures household i's response to the corresponding nutritional attribute. Thus, $\tilde{\gamma}_{iH}^{'}$ Healthy_{cj} can be regarded as the utility contribution of product j in category c to household i due to j's possessing healthy attributes. As our focus is the household's overall health consciousness, we introduce a new covariate Healthy_{cj0} and decompose the utility component, $\tilde{\gamma}_{iH}^{'}$ Healthy_{cj}, into two parts: one that is due to j's being a healthy product and the second, which can be attributed to the households' intrinsic preferences to a nutritional attribute (e.g., taste preference to sugar). Specifically, the decomposition is to replace the term $\tilde{\gamma}_{iH}^{'}$ Healthy_{cj} in the utility Eq. (1) with:

$$\gamma_{i0} \cdot \text{Healthy}_{cj0} + \gamma_{iH}^{'} \cdot \text{Healthy}_{cj} \quad (3)$$

where the variable Healthy_{cj0} takes a value of 1 if any element of Healthy_{cj} is equal to 1 and 0 otherwise. γ_{i0} is the coefficient that represents the household's preference for healthy products. As this preference is household specific and common across different nutritional attributes as well as categories, we can regard γ_{i0} as the household ‘health consciousness’ trait. γ_{iH} is an $H \times 1$ vector in which each element measures household i's intrinsic preference to the corresponding individual nutritional attribute.

As an illustration, consider diet-and-caffeine-free Coke, which is in the carbonated beverages category. In this category, the nutritional attributes

[2] When $j=0$, we observe that the household chooses not to make a purchase in the category at this trip.

[3] Setting off-diagonal elements equal to zero is a common assumption due to the computational complexity of estimating the multinomial probit model. In a robustness check, we found that estimating some of the variances in our application did not change the results presented here.

Table 1
Descriptive statistics of the selected products

Category	Nutritional attributes	Brands	# of SKUs represented	# of purchases represented	% of category sales represented
Carbonated beverages	Sugar, Caffeine	Private Label, Pepsi, Coke	195	10314	66
Coffee	Caffeine	Folgers, Private Label	100	831	57
Bread	Fiber	Private Label, Aunt Hatties, Oroweat	194	7588	61
Chewing Gum	Sugar	Wrigley's, Trident, Ice Breakers	54	553	68
Ice Cream	Fat	Private Label, Dreyers Edys	266	3380	70
Mayonnaise	Fat	Best Foods, Kraft	43	1145	75
Milk	Fat	Private Label, Shamrock Farms	94	12783	98
Salad Dressings	Fat	Private Label, Kraft, Wishbone	132	1868	75
Salty Snacks	Fat	Private Label, Lays	102	3163	40
Yogurt	Fat	Private Label, Yoplait	143	3561	78

are sugar and caffeine. We get $\gamma_{i0}\cdot Healthy_{cj0}+\gamma'_{iH}Healthy_{cj}=\gamma_{i0}+\gamma_{iH_s}+\gamma_{iH_{caf}}$ because we use the dummy variable coding

$$\begin{array}{lcccccc} & & \text{Low/no sugar} & \text{Decaf/low caff.} & \text{Low/no fat} & \text{High fiber} & \\ \text{Healthy}_{cj} & = & (1 & 1 & 0 & 0) & \end{array} \tag{4}$$

and $\gamma_{iH}=\left(\gamma_{iH_s}\gamma_{iH_{caf}}\gamma_{iH_f}\gamma_{iH_{fiber}}\right)$ and $\text{Healthy}_{cj0}=1$ That is, the elements in Healthy_{cj} that correspond to sugar and caffeine are 1′s because it is a sugar-free and caffeine-free product, and those that correspond to fat and fiber are 0 because those attributes do not affect utility in this category. Because it is a healthy product, $\text{Healthy}_{cj0}=1$. On the other hand, if the jth product is regular Coke, we would get $\gamma_{i0}\cdot Healthy_{cj0}+\gamma'_{iH}\cdot Healthy_{cj}=0$.

A household is health conscious if $\gamma_{i0}>0$. This is because γ_{i0}, being independent of the attributes and categories, is an overall measure of household i's health consciousness, whereas γ_{iH_s}, $\gamma_{iH_{caf}}$, $\gamma_{iH_{fat}}$, $\gamma_{iH_{fib}}$ may be thought of as intrinsic preferences (such as for taste) for sugar, caffeine, fat and fiber, respectively. The summed quantities, such as $\gamma_{i0}+\gamma_{iH}$, indeed capture a preference to buy, e.g., diet Coke. But it is possible that a household that likes the taste of regular Coke chooses diet Coke because it is highly health conscious, while another household that is not health conscious also chooses diet Coke because it likes the taste of diet Coke. Thus, to be health conscious overall is not related to how many of $\gamma_{i0}+\gamma_{iH_s}$, $\gamma_{i0}+\gamma_{iH_{caf}}$, $\gamma_{i0}+\gamma_{iH_{fat}}$ or $\gamma_{i0}+\gamma_{iH_{fib}}$ are positive, because even if γ_{i0} is positive, $\gamma_{i0}+\gamma_{iH_s}$ etc. do not have to be positive.

Our decomposition is similar to that of the variance components approach of Ainslie and Rossi (1998), who decompose the household's preferences to the marketing mix into two components: one specific to the household and common across categories; and the second specific to both the household and the category (in our context, the decomposition would be applied to each element of $\tilde{\gamma}_{iH}$). When there is only one nutritional attribute in each category, the two approaches are identical. The difference arises when there is more than one nutritional attribute in a category. For example, for diet-and-caffeine-free Coke, in our decomposition, $\tilde{\gamma}'_{iH}\cdot\text{Healthy}_{cj}=\gamma_{i0}+\gamma_{iH_s}+\gamma_{iH_{caf}}$, while using the variance components approach, the utility outcome would be $\tilde{\gamma}'_{iH}\cdot Healthy_{cj}=2\gamma_{i0}+\gamma_{iH_s}+\gamma_{iH_{caf}}$.The reasons for the proposed specification are twofold. First, the impact of health consciousness would otherwise be linearly proportional to the number of health attributes in a category. Second, in our specification, when at least one category possesses more than one nutritional attribute, we are able to identify γ_{i0} and each element of γ_{iH} separately, revealing the underlying drivers of household choice behavior.

In order to relate the households' health consciousness across multiple nutritional attributes to the household-specific responses to marketing mix, we follow the variance components approach to decompose the coefficients of marketing mix variables, β_{ic} into two components: $\beta_{ic}=\beta_{i0}+\lambda_{ic}$, where β_{i0} is the household-specific responses to marketing mix that are common across categories and λ_{ic} is the category-specific responses.

As seen in Eq. (2), the category inventory of the household at trip t, INV_{ict}, is incorporated into the household's utility for no-purchase in the category.[4] This specification allows the category demand to be affected by both the category attributes at trip t (i.e., the product attributes as well as the marketing activities) and the category inventory levels. INV_{ict} is imputed at a household level assuming a constant consumption rate by the household, which is equal to the total quantity purchase by i over the observation period divided by the time interval. The corresponding coefficient to be estimated is ϕ_{ic}.

Note that, according to Eqs. (2) and (3), the household i's decision of whether to make a purchase in category c is dependent not only on the household's brand preferences to available brand, the marketing mix, and the category inventory but also on the availability of healthy alternatives in the category.

To account for household heterogeneity, we adopt the random coefficient assumption. We denote $\theta_i=\{\alpha_{i1},\ldots,\alpha_{iW_1},\ldots,\alpha_{iW_c};\lambda_{i1},\ldots,\lambda_{iC};\beta_{i0};\gamma_{i0},\gamma_{i1},\ldots,\gamma_{iH}\}$ where W_c is the number of brands in category c, and assume that θ_i follows a multivariate normal distribution with a mean vector, δ, and a covariance matrix, $\sum$, i.e., $\theta_i\sim N(\delta,\sum)$. Since we are interested in the impact of household characteristics (e.g., demographics) on the household's health consciousness trait, γ_{i0}, we model it as a function of household characteristics. Specifically, $\gamma_{i0}=\eta_{i0}+\Delta' Z_i$, where Z_i is a vector that contains the characteristics of household i and Δ is the corresponding coefficient vector, and η_{i0} is the baseline health consciousness when $Z_i=0$. To account for the impact of demographics on health consciousness in the model, we can simply introduce the interactions of health consciousness with demographics as additional covariates and augment the vector θ_i and matrix Σ correspondingly. To keep a flexible specification, we allow the covariance matrix $\sum$ to be unrestricted; i.e., in addition to the diagonal elements of variances, we also estimate the off-diagonal covariance elements in $\sum$.

To estimate the proposed model, we employ hierarchical Bayes methodology by combining the likelihood described in this section with flat non-informative priors, and drawing our inferences about the parameters based on the resulting posterior. Appendix A discusses our estimation approach in more detail.

4. Empirical analysis

In this section we describe the data and the results of the analysis.

4.1. Data

We estimate the proposed model using scanner panel data on household purchases in ten categories in a major grocery store over a 53-week period (Sept. 2002–Sept. 2003) in a metropolitan market in a large city located in the southwest U.S.

[4] An alternative specification is to let the deterministic component of the no-purchase utility be a constant, which is usually set to zero.

The categories are carbonated beverages, coffee, bread, chewing gum, ice cream, mayonnaise, milk, salad dressings, salty snacks and yogurt. Four nutritional attributes — sugar, caffeine, fat and fiber — can be easily identified in these categories. We define "low/no sugar" as "healthy" and regular as "unhealthy" for the sugar attribute, "decaf" or "less caffeine" as "healthy" and caffeinated as "unhealthy" for the caffeine attribute, "low/no fat" as "healthy" and regular fat as "unhealthy" for the fat attribute, and "high fiber" as "healthy" and "unhealthy" otherwise. Table 1 summarizes the nutritional attributes identified in each category. Except for the category of carbonated beverages, which possesses two nutritional attributes, each category possesses one of four nutritional attributes. As mentioned in the model section, since one category (carbonated beverages) has more than one nutritional attribute, we are able to estimate the health consciousness trait that is common across nutritional attributes (i.e., γ_{i0}) as well as the nutritional attribute specific preference (i.e., elements of γ_{iH}).

We describe a product in terms of the category it belongs to, its corresponding brand name and the nutritional attributes it possesses. For each category, we focus on the top selling brands. Except in the categories of coffee and salty snacks,[5] the selected products in each category capture at least 61% of category sales and represent a large number of SKUs. We report these summary statistics also in Table 1.

We include all households in our data that make at least two purchases over the data period.[6] This gives us a total of 1062 households. The number of purchases made by the selected households in each category is also given in Table 1.

The only marketing mix variable that is available in our data is price. We normalize all prices into unit prices and include them in our empirical analysis. Table 2 provides descriptions of the products in our analysis and their average unit prices.

We also have the following demographic characteristics for the households in our data: Household income, home ownership, employment status of male household head, employment status of female household head, education level of male household head, family size, age of female household head, presence of young children (age 0–17), and race of the household. We include these demographic variables in our analysis to profile the households in terms of their health consciousness. Table 3 shows the summary statistics of the household demographics.

4.2. Estimation results and discussions

We calibrated our model on the data described above using the MCMC procedure described in Appendix A. Overall, we have a total of 48 covariates in our model. In addition to the 48 mean effects, we estimate an unrestricted heterogeneity variance–covariance matrix $\sum$, which makes a total of 1224 free parameters to estimate. We discuss the estimates that are closely relevant to our research questions in the text below. The posterior means and variances of several remaining parameters including brand intercepts and category-specific price coefficients are given in Appendix B.

Our first research question is, how large is the health conscious market? With the individual estimates of household health consciousness trait, γ_{i0}, which includes the impact of the household's demographics, we can calculate the proportion of the households that have positive health consciousness (i.e., γ_0 is positive). We find that the size of the health-conscious market in our sample is 18%. Although it may be considered a relatively small market, the respectable size of the market

Table 2
Descriptions and average prices of the products

Product ID	Category	Brand	Low/no sugar	Decaf/low caffeine	Low/no fat	High fiber	Average Unit Price
1	Carbonated beverages	Private label	0	0	0	0	0.011
2	Carbonated beverages	Private label	0	1	0	0	0.013
3	Carbonated beverages	Private label	1	0	0	0	0.011
4	Carbonated beverages	Private label	1	1	0	0	0.011
5	Carbonated beverages	Pepsi	0	0	0	0	0.020
6	Carbonated beverages	Pepsi	0	1	0	0	0.021
7	Carbonated beverages	Pepsi	1	0	0	0	0.020
8	Carbonated beverages	Pepsi	1	1	0	0	0.017
9	Carbonated beverages	Coke	0	0	0	0	0.024
10	Carbonated beverages	Coke	0	1	0	0	0.027
11	Carbonated beverages	Coke	1	0	0	0	0.023
12	Carbonated beverages	Coke	1	1	0	0	0.020
13	Coffee	Folgers	0	0	0	0	0.412
14	Coffee	Folgers	0	1	0	0	1.002
15	Coffee	Private label	0	0	0	0	0.350
16	Coffee	Private label	0	1	0	0	0.516
17	Bread	Private label	0	0	0	0	0.076
18	Bread	Private label	0	0	0	1	0.060
19	Bread	Aunt Hatties	0	0	0	0	0.115
20	Bread	Aunt Hatties	0	0	0	1	0.109
21	Bread	Oroweat	0	0	0	0	0.151
22	Bread	Oroweat	0	0	0	1	0.154
23	Chewing gum	Wrigley's	0	0	0	0	0.052
24	Chewing gum	Wrigley's	1	0	0	0	0.064
25	Chewing gum	Trident	0	0	0	0	0.096
26	Chewing gum	Trident	1	0	0	0	0.091
27	Chewing gum	Ice Breakers	0	0	0	0	0.083
28	Chewing gum	Ice Breakers	1	0	0	0	0.112
29	Ice cream	Private label	0	0	0	0	0.038
30	Ice cream	Private label	0	0	1	0	0.041
31	Ice cream	Dreyers Edys	0	0	0	0	0.097
32	Ice cream	Dreyers Edys	0	0	1	0	0.075
33	Mayonnaise	Best Foods	0	0	0	0	0.113
34	Mayonnaise	Best Foods	0	0	1	0	0.165
35	Mayonnaise	Kraft	0	0	0	0	0.111
36	Mayonnaise	Kraft	0	0	1	0	0.106
37	Milk	Private label	0	0	0	0	0.018
38	Milk	Private label	0	0	1	0	0.024
39	Milk	Shamrock Farms	0	0	0	0	0.040
40	Milk	Shamrock Farms	0	0	1	0	0.041
41	Salad dressings	Private label	0	0	0	0	0.120
42	Salad dressings	Private label	0	0	1	0	0.129
43	Salad dressings	Kraft	0	0	0	0	0.170
44	Salad dressings	Kraft	0	0	1	0	0.187
45	Salad dressings	Wishbone	0	0	0	0	0.159
46	Salad dressings	Wishbone	0	0	1	0	0.169
47	Salty snacks	Private label	0	0	0	0	0.111
48	Salty snacks	Private label	0	0	1	0	0.265
49	Salty snacks	Lays	0	0	0	0	0.189
50	Salty snacks	Lays	0	0	1	0	0.360
51	Yogurt	Private label	0	0	0	0	0.080
52	Yogurt	Private label	0	0	1	0	0.058
53	Yogurt	Yoplait	0	0	0	0	0.118
54	Yogurt	Yoplait	0	0	1	0	0.112

[5] For coffee, only the top two brands offer products that are "decaf" or "less caffeine," and, for salty snacks, only the top two brands offer products that have "low/no" fat.

[6] Since we allow for the no-purchase option, when selecting the households we do not impose the restriction that each household has to make at least one purchase in each of the ten categories as in Ainslie and Rossi (1998) and Singh, Hansen and Gupta (2005).

Table 3
Summary statistics for household demographic variables

Demographics	Level	Distribution (%)
Income	\$0–9999	2
	\$10,000–11,999	2
	\$12,000–14,999	2
	\$15,000–19,999	3
	\$20,000–24,999	6
	\$25,000–34,999	14
	\$35,000–44,999	16
	\$45,000–54,999	16
	\$55,000–64,999	10
	\$65,000–74,999	9
	\$75,000–99,999	12
	\$100,000+	8
	No response	1
Home ownership	Own	85
	Rent	15
Employment status of male household	Employed	64
	Not employed	36
Employment status of female household	Employed	59
	Not employed	41
Education of male household head	Completed grade school	0
	Some high school, graduated high school or technical school	30
	Some college, graduated college, or post grad work	51
	No response or no male head	19
Family size	1	13
	2	39
	3	18
	4	18
	5	7
	6	3
	7	1
Age of female household head	18–29	3
	30–34	7
	35–44	27
	45–54	29
	55–64	20
	65+	12
	No female	3
Have young children (0–17 years old)	Yes	38
	No	62
Race	White	81
	Others	16

seems to justify the marketing effort that food marketers put on introducing healthy food alternatives and targeting health conscious consumers. Having a continuous measure of health consciousness could technically allow us to investigate health consciousness on a finer scale by looking at those households with γ_0 slightly positive and those with γ_0 highly positive. Given that the proportion of those with positive values is small, however, we did not pursue further fragmentation.

We are interested in the correlation, if any, between a household's response to price and its health consciousness. As mentioned in the model section, we decomposed the price response of a household in category c, β_{ic}, into two components — the household-specific price response β_{i0}, which is common across categories; and the category-specific price response λ_{ic}. We report the posterior mean and variance for β_{i0} in Table 4. We find that, on average, households in our sample are sensitive to price. This is because the posterior mean of the household-specific price coefficient is negative and significant. We also find that a great degree of heterogeneity exists in households' price sensitivity, as the posterior mean of the variance is very large.

To understand the relationship between households' price sensitivity and health consciousness, we calculate the correlation between the household-specific price coefficient, β_0, and the baseline household consciousness, η_0. With variances of β_0 and η_0 being 232.302 and 5.071 respectively, and covariance between β_0 and η_0 being 31.454, we obtain the correlation between the two to be 0.92. The result suggests that a household's price response to food purchases is highly correlated with its health consciousness, and that the more health conscious a household is, the less price sensitive it is. Thus, we can suggest that, when targeting healthy food at health-conscious households, marketers may charge a premium. Furthermore, marketers should focus on non-price promotions and advertising to make the healthy attribute of the products more salient.

Households' intrinsic preferences for individual nutritional attributes are measured by γ_{H_s}, $\gamma_{H_{caf}}$, γ_{H_f} and $\gamma_{H_{fiber}}$, where γ_{H_s} measures households' preferences for sugar, $\gamma_{H_{caf}}$ measures households' preferences for caffeine, γ_{H_f} measures households' preferences for fat, and $\gamma_{H_{fiber}}$ is households' preference for fiber. We report the posterior means for these coefficients in Table 4. Since we consider low/no sugar, decaf or less caffeine, and low/no fat as healthy nutritional attributes, a positive coefficient indicates dislike (for sugar, caffeine, fat), and a negative one indicates preference. Since we consider high fiber to be a healthy nutritional attribute, a positive coefficient indicates preference for fiber. As shown in Table 4, we find that, on average, the households in our dataset exhibit an innate liking of caffeine but a dislike of fat and fiber, and we find that they are indifferent to sugar. Also shown in Table 4, all the posterior variances of the heterogeneity distribution for these coefficients are significant and have large values in absolute terms, suggesting that the households are heterogeneous in their intrinsic preferences for these nutritional attributes. As mentioned earlier, the ability of our proposed model to distinguish households' overall health consciousness from their preferences to specific nutritional attributes certainly helps marketers have a better understanding about households' preferences for their products.

We calculate the correlations between $\gamma_{i0}+\gamma_{iH_s}$, $\gamma_{i0}+\gamma_{iH_{caf}}$, $\gamma_{i0}+\gamma_{iH_f}$, and $\gamma_{i0}+\gamma_{iH_{fiber}}$ as shown in Table 5. We find high positive correlations (ranging from 0.897 to 0.921), which implies that households tend to be consistent in their preferences for healthy product alternatives across various health attributes.

We also calculate the correlations between the intrinsic preferences for individual health attributes in Table 6. We find that the correlations in Table 6, in contrast to those in Table 5, are not consistently high. For

Table 4
Estimates of posterior means and variances of household specific price sensitivity and preferences for individual nutritional attributes

Parameters	Mean	Variance
β_{i0} (household specific price coefficient)	−80.897[a] (2.810)	232.302[a] (39.058)
γ_{H_s} (preferences for low/no sugar)	0.209 (0.112)	0.699[a] (0.148)
$\gamma_{H_{caf}}$ (preferences for decaf or less caffeine)	−0.405[a] (0.152)	0.916[a] (0.365)
γ_{H_f} (preferences for low/no fat)	0.356[a] (0.107)	1.086[a] (0.314)
$\gamma_{H_{fiber}}$ (preferences for high fiber)	−0.517[a] (0.115)	0.469[a] (0.181)

Numbers in the parentheses are standard deviations of the estimates.
[a] The 95% credibility interval does not include zero.

Table 5
Correlations among overall preferences

Parameters	$\gamma_0+\gamma_{H_s}$	$\gamma_0+\gamma_{H_{caf}}$	$\gamma_0+\gamma_{H_f}$	$\gamma_0+\gamma_{H_{fiber}}$
$\gamma_0+\gamma_{H_s}$ (overall preference for low/no sugar)	1.0			
$\gamma_0+\gamma_{H_{caf}}$ (overall preference for decaf or less caffeine)	0.916	1.0		
$\gamma_0+\gamma_{H_f}$ (overall preference for low/no fat)	0.921	0.910	1.0	
$\gamma_0+\gamma_{H_{fiber}}$ (overall preference for high fiber)	0.897	0.919	0.898	1.0

Table 6
Correlations among intrinsic nutritional preferences

Parameters	γ_{H_s}	$\gamma_{H_{caf}}$	γ_{H_f}	$\gamma_{H_{fiber}}$
γ_{H_s} (intrinsic preference for low/no sugar)	1.0			
$\gamma_{H_{caf}}$ (intrinsic preference for decaf or less caffeine)	0.36	1.0		
γ_{H_f} (intrinsic preference for low/no fat)	0.45	0.69	1.0	
$\gamma_{H_{fiber}}$ (intrinsic preference for high fiber)	N/S	0.60	0.57	1.0

N/S = non-significant correlations.

example, household-specific intrinsic preference for sugar is not significantly related to preference for fiber; and other correlations range from 0.36 to 0.69. This implies that, though households do not have consistent intrinsic preferences (e.g., taste preferences) to different health attributes, they show consistent overall preferences to healthy product alternatives due to the strong impact of their health consciousness.

When we observe that some households choose healthy options on some attributes, but unhealthy options on other attributes, we infer that the observed choices are determined by the joint impact of different effects in each category. For example, when we see that household *i* chooses decaf coffee but regular ice cream, this could be due to a positive γ_{i0}, a positive $\gamma_{iH_{caf}}$ and a large negative γ_{iH_f}, or, alternatively, it could be due to a negative γ_{i0}, a large positive $\gamma_{iH_{caf}}$ and a negative γ_{iH_f}, as long as $\gamma_{i0}+\gamma_{iH_{caf}}>0$ and $\gamma_{i0}+\gamma_{iH_f}<0$. While the observed choices are the same, the different underlying drivers of consumer behavior would generate different managerial implications. The ability to identify these drivers is an important contribution that can improve marketing strategies. For example, when introducing a new product with an alternative nutritional attribute, e.g., a low-sugar cereal, the knowledge of the level of its target consumers' health consciousness as well as that of their intrinsic preference to sugar will not only help the firm predict consumers' choices, but it can also help the firm tailor different advertising messages (e.g., choose between emphasizing the health benefits of the cereal and its unique low-sugar taste) to promote the product.

Next, we examine the relationship between the health consciousness trait of a household and its demographic characteristics. In Table 7, we report the posterior means as well as the posterior variances of heterogeneity distribution, which are the diagonal elements in $\sum$, for corresponding coefficients. We see that the influences of the demographics on households' health consciousness are heterogeneous among households, with the variation being most salient for the influence of home ownership. Among the nine demographics included in the analysis, six have a strong influence on the household health consciousness. By correlating the purchases in multiple categories, we are able to more accurately identify the relationship between the health consciousness traits of households and their characteristics (Ainslie & Rossi, 1998). We discuss the results below:

1. Income & home ownership: The higher income that a household has, the more health conscious it is. A household that owns its residence is more health conscious than a household that does not own.

Our finding is consistent with that in Moorman and Matulich (1993), who summarize the results on *Income* in the literature as having a positive effect on health behavior. An explanation is that lower income can constrain a household's ability to engage in health-conscious behavior.

2. Employment: If the male household head is employed, the household is less health conscious. Whether the female household head is employed or not has no impact on the household's health consciousness.

Table 7
Estimates of posterior means and variances for household health consciousness

γ_{i0}	Mean	Variance
Intercept η_0	−4.204[a] (0.240)	5.071[a] (1.443)
Income	0.083[a] (0.011)	0.034[a] (0.007)
Home ownership	0.962[a] (0.186)	1.269[a] (0.430)
Male employment	−0.457[a] (0.116)	0.333[a] (0.116)
Female employment	−0.034 (0.069)	168[a] (0.079)
Male education	0.275[a] (0.075)	0.165[a] (0.072)
Family size	0.071 (0.048)	0.091[a] (0.031)
Female age	0.099 (0.057)	0.081[a] (0.027)
Have young children	0.260[a] (0.156)	0.257[a] (0.097)
White	−0.249[a] (0.087)	0.639[a] (0.192)

Numbers in the parentheses are standard deviations of the estimates.
[a] The 95% credibility interval does not include zero.

The *employment* variable has not been considered in previous studies. For reasons similar to the income effect, one might hypothesize employment has a positive relationship with health consciousness. A further force is peer pressure in the workplace, in which case the implication is that advertisements should show the undesirability of being unfit in the workplace as a fear appeal to promote healthy food. But this is not supported.

3. Education: The level of formal education of the male household head increases the household health consciousness.

The posterior mean for the coefficient is 0.27, which is positive and significant at the 5% level. In their literature review, Moorman and Matulich (1993) hypothesized that education encourages most types of health information acquisition behaviors, although their empirical analysis did not find support for an education interaction with health behaviors. With results similar to this study, Divine and Lepisto (2005) conclude that being better educated is a significant predictor of a healthier lifestyle. For marketers, this may indicate that promotional nutritional messages targeted to men should use informative advertising to convince them to purchase healthier foods, as our results suggests that health information transforms into purchase behavior for healthy food.

4. Family size & age & have young children: Having young children in the household has a strong positive impact on the household's health consciousness, while the size of the family and the age of the female household head have no impact.

The posterior mean of the coefficient for the presence of young children, i.e., those under 17 years old, in the household is 0.26, which is positive and significant. This result suggests that marketers of healthy food should target families with young children. The posterior means of the coefficients for the family size and the age of female household head, which is usually used as an indication for the age of the household, are positive but not significant. Our results are not inconsistent with the literature since Moorman and Matulich (1993) summarize the results on *Age* in the literature as mixed. Divine and Lepisto (2005), for example, concluded that being older, being better educated and being female are significant predictors of a healthier lifestyle. However, we do not see this from the purchase behavior.

5. Ethnicity: Ethnicity has an impact on health consciousness of households.

The households in the dataset are classified as white and "others". We denote the dummy variable to be 1 if household is white, and 0 otherwise. The posterior mean of this coefficient is negative and significant at the 5% level, suggesting that, compared with households of other ethnicities, white households are less health conscious. It is difficult to pinpoint what ethnic food preferences might cause this. In contrast, Ippolito and Mathios (1990) found that whites were more likely to eat higher fiber cereals than nonwhites, speculating that the difference must be due to an informational advantage or having a taste preference for it. We do not have information acquisition in this study or study cereal as a category, so a direct comparison is not possible. Also, as defined, the preference for an individual nutritional attribute alone is not an indication of the household's health consciousness. However, recognizing ethnic preference differences can help marketers make better marketing strategies for the healthy food, such as communications and channel designs (e.g., distribute healthy food to the stores where particular ethnical groups are likely to shop).

5. Conclusions

Although consumers may express positive attitudes toward the attributes and benefits associated with food products, this does not necessarily translate into actual purchases. For example, evidence suggests that, even after respondents state their intention to eat non-fattening foods, they do so only about three times out of ten (Jones et al., 2003). The lack of predictive power in self-explicated consumer surveys presents a barrier to understanding consumption of food products and planning food-related research, policies and strategies. Yet there has been limited research based on actual consumer purchase data regarding consumers' health consciousness regarding nutrition. We attempt to fill this gap by studying consumers' preferences for healthy nutrition revealed by their food purchases. To identify health-conscious behavior as a consumer trait that affects consumer's choices across categories with different nutritional attributes, we use a multi-category brand choice model and estimate it on basket data for a large sample of households in ten commonly purchased grocery categories.

Correlating the purchases in multiple categories helps to accurately identify the relationship between the consumer traits and the household characteristics (Ainslie & Rossi, 1998). Our examination on the relationship between household health consciousness and the demographic characteristics showed that six out of nine demographic variables have a strong impact. We find that a household with a higher income is more health conscious. If a household owns its residence, it is more health conscious. Households with a working male household head are less health conscious, but if the male household head is more educated, the household is more health conscious. The presence of young children in the household increases the household's health consciousness.

Our empirical analysis reveals that the size of the health-conscious segment is about 18% of the sample. This finding helps to justify marketing effort in targeting the health-conscious consumers for firms in the food industry. We also find that health-conscious households are less price sensitive, confirming that marketers have the leeway to charge higher prices for healthy version foods. Since our model can distinguish households' overall health consciousness across nutritional attributes from their preferences to specific nutritional attributes, it certainly helps marketers have a better understanding about households' preferences for their products.

A limitation to keep in mind is that the results apply to only the population of consumers who shop at regular supermarkets like the type of store we examined. This does not include health-oriented specialty grocery stores such as Whole Foods. This could create a possible underestimation of the size of the healthy segment.

Future research could consider incorporating more categories and nutritional attributes (e.g., protein, sodium, etc.) where data are available and account for the quantities purchased or consumed per month as indicators of healthy behavior. The present research defined health consciousness as a consistent preference for healthy versions of products across categories. But alternative definitions should also be studied. For example, a health-conscious consumer can make compensatory behaviors across categories. Such a consumer might set a health goal of obtaining a certain amount of nutrients, calories, fat etc, and then shop in the different categories to achieve the goal[7]. Thus, a healthy choice in most categories could compensate less healthy choices in other categories. Scanner data is limited in its ability to uncover precise motivations for purchase and further research could seek to augment it. For example, Pieters, Baumgartner and Allen (1995) study how a number of goals, arranged in a hierarchy, may underlie consumers' weight loss behavior. They suggest that diet and exercise may be sub-goals driven by the focal goal of losing weight, which itself may reflect super-ordinate goals such as leading a long and healthy life, or of being attractive to others and feeling good about oneself. They describe a laddering procedure to uncover the motivations behind consumers' behavior.

Finally, there are also public policy implications of this research that deserve further attention. The same analysis that helps marketing managers also assists health professionals and government agencies to identify healthy and less healthy individuals and plan campaigns to advise them of nutrition habits. They might also try to supplement grocery purchase data with disease incidence data and examine relationships between the two.

Appendix A. Details of the MCMC procedure for the hierarchical Bayesian estimation

We estimate the parameters of this model using Bayesian methodology by combining the likelihood described in this section with flat non-informative priors, and drawing our inferences about the parameters based on the resulting posterior. Specifically, we construct a Gibbs sampler which has four full conditional posterior distributions for the following random variables: augmented data $\{z_{itcj}\}$, random effects $\{\theta_i\}$, mean and variance of the heterogeneity distribution δ and Σ: $\theta_i \sim N_k(\delta,\Sigma)$, where k is the total number of covariates.

The posterior distributions are sampled in the following order.

1. For each household i at time t in category c:

   $z_{itcj}|\delta,\theta_i \sim TN_{(\max_{k\neq j}(z_{itck}),\infty)}(W_{itcj}(\delta+\theta_i),1)$ if option j was chosen $(y_{itcj}=1)$

   $z_{itcj}|\delta,\theta_i \sim TN_{(-\infty,\max(z_{itck}))}(W_{itcj}(\delta+\theta_i),1)$ if option j was not chosen $(y_{itcj}=0)$

2. $\delta|\{z_i\}, \{\theta_i\} \sim N_k(\bar{\beta}, \bar{B})$, where $\bar{B} = (W'W)^{-1}$ and $\bar{\beta} = \bar{B}\left(\sum_{i=1}^{I} W_i'(z_i - W_i\theta_i)\right)$

3. For each household i:

   $\beta_i|\{z_i\},\delta,\Sigma \sim N_k(\bar{\beta}_i,\bar{B}_i)$, where $\bar{B}_i = \left(W_i'W_i + \Sigma^{-1}\right)^{-1}$ and $\bar{\beta}_i = \bar{B}_i\left(W_i'(z_i - W_i\delta)\right)$

4. $D \mid \{\theta_i\} \sim InverseWishart_{I+k+4}(\bar{S})$, where $\bar{S} = I_k + \sum_{i=1}^{I} \theta_i \cdot \theta_i'$, and I_k is an identity matrix of dimension k.

The relationship between the data in our model and the variables needed for the estimation is as follows. At the category-household level, we define, for $j=1,\ldots, J_c$, $W_{itcj}^1=(B_{cj}X_{itcj})$ as the row vector containing the brand intercepts and marketing mix covariates and $W_{itcj}^2=(X_{tcj}\text{Healthy}_{cj}\text{Healthy}_{cj0}\cdot{*}Z_i$ as the row vector containing the price, health dummies and interactions of the health trait with demographics; and for $j=0$, $W_{itc0}^1=(\text{Inv}_{tc})$ as the inventory for category c, and $W_{itc0}^2=()$ as the empty set. Then, we can define $W_{itcj}=(W_{itcj}^1 W_{itcj}^2)$ as the row vector containing W_{itcj}^1 and W_{itcj}^2. Our data y_{itc} and the augmented variable z_{itc} have the following relationship:

$$\begin{cases} z_{itcj} > \max_{k\neq j}(z_{itck}) \text{ if } y_{itcj}=1 \\ z_{itcj} \le \max_{k\neq j}(z_{itck}) \text{ if } y_{itcj}=0 \end{cases}.$$

At the household level, we define z_{it} and W_{it} as

$$\begin{pmatrix} z_{itc0} \\ z_{itc1} \\ \vdots \\ z_{itcJ_c} \end{pmatrix} \text{ and } \begin{pmatrix} W_{itc0}^1 & 0 & \cdots & 0 & 0 \\ 0 & W_{itc1}^1 & \cdots & 0 & W_{itc1}^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & W_{itcJ_c}^1 & W_{itcJ_c}^2 \end{pmatrix}$$

and z_i and W_i as

$$\begin{pmatrix} z_{i1} \\ z_{i2} \\ \vdots \\ z_{iT_i} \end{pmatrix} \text{ and } \begin{pmatrix} W_{i1} \\ W_{i2} \\ \vdots \\ W_{iT_i} \end{pmatrix},$$ where T_i is the total number of observations for household i.

Finally, we define the augmented data vector z and the full covariate matrix W as

$$\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_I \end{pmatrix} \text{ and } \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_I \end{pmatrix}.$$

[7] We thank an anonymous reviewer for pointing this out.

Appendix B. Estimates of posterior means and variances for brand, price and inventory coefficients

Category	Brands	Brand		Price		Inventory	
		Mean	Variance	Mean	Variance	Mean	Variance
Carbonated beverages	Private label	NA	NA	NA	NA	0.001[a] (0.000)	0.008[a] (0.001)
	Pepsi	0.907[a] (0.180)	1.430[a] (0.196)				
	Coke	1.183[a] (0.098)	1.323[a] (0.167)				
Coffee	Folgers	NA	NA	69.058[a] (2.690)	129.072[a] (20.560)	0.013 (0.019)	0.094[a] (0.024)
	Private label	−1.47[a] (0.100)	0.647[a] (0.136)				
Bread	Private label	NA	NA	52.447[a] (2.602)	2.128[a] (0.366)	0.022[a] (0.007)	0.013[a] (0.002)
	Aunt Hatties	0.087 (0.096)	1.816[a] (0.306)				
	Orowheat	0.751[a] (0.168)	2.694[a] (0.658)				
Chewing gum	Wrigley's	NA	NA	7.284 (4.492)	5.572[a] (1.731)	−0.009 (0.013)	0.045[a] (0.013)
	Trident	−2.816[a] (0.295)	2.882[a] (0.670)				
	Ice Breakers	−2.292[a] (0.231)	2.012[a] (0.345)				
Ice cream	Private label	NA	NA	24.319[a] (3.386)	16.297[a] (7.762)	0.005[a] (0.003)	0.013[a] (0.002)
	Dreyer Edys	1.102[a] (0.100)	0.580[a] (0.102)				
Mayonnaise	Best Foods	NA	NA	35.580[a] (3.188)	1.870[a] (0.570)	−0.052[a] (0.017)	0.021[a] (0.003)
	Kraft	−0.050 (0.339)	0.337[a] (0.240)				
Milk	Private label	NA	NA	35.430[a] (3.169)	4.402[a] (3.050)	0.001[a] (0.001)	0.007[a] (0.001)
	Shamrock Farms	−0.980[a] (0.187)	0.775[a] (0.149)				
Salad dressing	Private label	NA	NA	39.830[a] (2.909)	4.976[a] (2.337)	0.020[a] (0.010)	0.025[a] (0.004)
	Kraft	1.348[a] (0.304)	0.776[a] (0.260)				
	Wishbone	0.954[a] (0.158)	0.321[a] (0.117)				
Salty snacks	Private label	NA	NA	52.786[a] (3.103)	18.518[a] (3.884)	−0.029 (0.023)	0.029[a] (0.008)
	Lays	1.080[a] (0.228)	0.578[a] (0.139)				
Yogurt	Private label	NA	NA	30.801[a] (2.720)	3.556[a] (1.097)	0.012 (0.008)	0.016[a] (0.003)
	Yoplait	1.109[a] (0.209)	1.366[a] (0.236)				

Numbers in the parentheses are standard deviations of the estimates.
[a]The 95% credibility interval does not include zero.

References

Ainslie, A., & Rossi, P. E. (1998). Similarities in choice behavior across multiple categories. *Marketing Science*, *17*(2), 91–106.

Andrews, J. C., Netemeyer, R. G., & Burton, S. (1998). Consumer generalization of nutrient claims in advertising. *Journal of Marketing*, *62*(4), 62–75.

Barboza, D., (2003, July 26). Can Kraft trim the fat in an Oreo world? *New York Times*.

Brucks, M., Mitchell, A., & Staelin, R. (1984). The effect of nutritional information disclosure in advertising: An information processing approach. *Journal of Public Policy and Marketing*, *3*, 1–25.

Chandon, P., & Wansink, B. (2007). Is obesity caused by calorie underestimation? A psychophysical model of meal size estimation. *Journal of Marketing Research*, *44*(1), 84–99.

Chib, S., Seetharaman, P. B., & Strijnev, A. (2002). Analysis of multi-category purchase incidence decisions using IRI market basket data. *Advances in Econometrics*, *16*, 57–92.

Chintagunta, P. K. (2002). Investigating category pricing behavior at a retail chain. *Journal of Marketing Research*, *39*(2), 141–154.

Condie, B. (2005, January 17). Welcome change of diet. *Evening Standard (London)*.

Cutler, D. M., Glaeser, E. L., & Shapiro, J. M. (2003). Why have Americans become more obese? *Journal of Economic Perspectives*, *17*(3), 93–118.

Divine, R. L., & Lepisto, L. (2005). Analysis of the healthy lifestyle consumer. *Journal of Consumer Marketing*, *22*(5), 275–283.

Fader, P. S., & Hardie, B. G. S. (1996). Modeling consumer choice among SKUs. *Journal of Marketing Research*, *33*(4), 442–452.

Garille, S. G., & Gass, S. I. (2001). Stigler's diet problem revisited. *Operations Research*, *49*(1), 1–13.

Gould, S. J. (1990). Health consciousness and health behavior: The application of a new health consciousness scale. *American Journal of Preventive Medicine*, *6*, 228–237.

Hansen, K., Singh, V., & Chintagunta, P. K. (2006). Understanding store-brand purchase behavior across categories. *Marketing Science*, *25*(1), 75–90.

Horovitz, B. (2003, July 1st). Under fire, food giants switch to healthier fare. *USA Today*.

Ippolito, P. M., & Mathios, A. D. (1990). Information, advertising and health choices. *RAND Journal of Economics*, *21*(3), 459–480.

Jones, J., Walker, D., Shim-Prydon, G., Cowan, D., Gatfield, T., Smith, A., et al. (2003, January). *"Drivers of consumer behaviour," Business and Trade Report*. Queensland, Australia: Department of Primary Industries and Fisheries.

Leeflang, P. S. H., & van Raaij, W. F. (1995). The changing consumer in the European union: A "meta-analysis". *International Journal of Research in Marketing*, *12*(5), 373–387.

Leviton, A., & Allred, E. N. (1994). Correlates of decaffeinated coffee choice. *Epidemiology*, *5*, 537–540.

Lichtman, S. W., Pisarska, K., Berman, E. R., Pestone, M., Dowling, H., Offenbacher, E., et al. (1992). Discrepancy between self-reported and actual caloric intake and exercise in obese subjects. *New England Journal of Medicine*, *327*, 1893–1898.

Manchanda, P., Ansari, A., & Gupta, S. (1999). The 'shopping basket': A model for multicategory purchase incidence decisions. *Marketing Science*, *18*(2), 95–114.

Marlett, J. A., McBurney, M. I., & Slavin, J. L. (2002). Position of the American Dietetic Association: Health implications of dietary fiber. *Journal of the American Dietetic Association*, *102*(7), 993–1000.

Moorman, C., & Matulich, E. (1993). A model of consumers' preventive health behaviors: The role of health motivation and health ability. *Journal of Consumer Research*, *20*(2), 208–228.

Olekalns, N., & Bardsley, P. (1996). Rational addiction to caffeine: An analysis of coffee consumption. *Journal of Political Economy*, *104*(5), 1100–1104.

Palmer, J. R., Rosenberg, L., Rao, R. S., & Shapiro, S. (2008). Coffee consumption and myocardial infarction in women. *American Journal of Epidemiology*, *141*(8), 724–731.

Pieters, R., Baumgartner, H., & Allen, D. (1995). A means-end chain approach to consumer goal structures. *International Journal of Research in Marketing*, *12*(3), 227–244.

Russo, J. E., Staelin, R., Nolan, C. A., Russell, G. J., & Metcalf, B. L. (1986). Nutritional information in the supermarket. *Journal of Consumer Research*, *13*(1), 48–70.

Seetharaman, P. B., Chib, S., Ainslie, A., Boatwright, P., Chan, T., Gupta, S., et al. (2005). Models of multi-category choice behavior. *Marketing Letters*, *16*(3), 239–254.

Singh, V. P., Hansen, K., & Gupta, S. (2005). Modeling preferences for common attributes in multi-category brand choice. *Journal of Marketing Research*, *42*(2), 195–209.

Thompson, S. (2003, June 30th). Hawaiian punch launch: Wal-Mart wants more 'lite'. *Advertising Age*.

The Economist (2007, January 6th). *A Magic Potion?*

United Soybean Board (2005). 2005 National Report: Consumer attitudes about nutrition. *12th Annual Survey*.

Wolk, M. (2005, August 2nd). Bankruptcy marks rapid fall of low-carb craze. *MSNBC.COM*.