Estimating the Effect of Health Knowledge in the Consumption of Soy-Based Foods

Abstract
This paper develops empirical models to investigate whether knowledge about the health benefits of soy-based foods plays a significant role in consumers’ decision to consume soy-based foods and how much to consume (consumption intensity). In addition, the impact of FDA’s 1999 official confirmation of the health benefits of soy foods on consumers’ willingness-to-consume soy foods or willingness-to-increase consumption frequency is examined.

Wanki Moon  
Dept. of Agribusiness Economics  
Southern Illinois University  
Carbondale IL 62901  
Fax) 618) 453-1708  
Tel) 618) 453-6741  
Email) wmoon@siu.edu

Siva K. Balasubramanian  
Dept. of Marketing  
Southern Illinois University  
Carbondale IL 62901  
Tel) 618) 4341  
Fax) 618) 453-7747  
Email) siva@cba.siu.edu
Medical research substantiating the effects of diet on chronic diseases has played a key role in promoting consumers’ nutritional and health awareness and in motivating consumers to change their dietary behavior. For example, the established linkage between saturated fats and cholesterol and coronary heart diseases has exerted far-reaching influence on dietary choices of U.S. consumers as evidenced by consumption trends of meat, milk, fats, eggs and oil (Brown and Schrader, 1990; Capps and Schmits, 1991; Putler and Frazao, 1991; Ward and Moon, 1996). Desire for foods that contribute to better health and longer life is now considered a crucial factor driving U.S. food demands (Senauer, 1990; Kinsey, 1992; Chern et al., 1997).

Recent scientific findings concerning the health benefits of soy-based foods present an opportunity to validate the roles of regulatory agency (FDA) and new information in food consumption. Clinical research in the past few years demonstrated that soy-based foods provide various health benefits relative to chronic diseases such as osteoporosis, heart disease, and cancer. For example, Messina and Barnes (1995) found that consuming soy-based foods had anticarcinogenic effects on breast and prostate cancers. More importantly, 25 grams of soy protein daily has been found to have a significant effect of lowering cholesterol level (Anderson, Johnstone, and Cook-Newell, 1995). In 1999, the series of new findings prompted the FDA to permit food companies to use health claims on soy-based foods which contain a minimum of 6.25 grams (i.e., one-fourth of the daily recommended level of 25 grams) of soy protein per serving. The approval was in response to a petition filed by Protein Technologies International (PTI), a Dupont Business, in May 1998.
The FDA undergoes stringent review of the scientific evidence before it allows health claims on food products. To date, the FDA established only seven allowable health claims, including calcium and a reduced risk of osteoporosis, and sodium and an increased risk of hypertension. In response to a petition from the Quaker Oats Company, the FDA approved in 1997 the first food specific health claim under the NLEA. This opened the door for additional product specific claims, particularly leading to the approval of a petition to link soy protein with reduced blood cholesterol. The permission of health claims for soy foods is of considerable significance to the soy industry. The permission officially authenticated the health benefits of soy-based foods as well as enabling the soy food industry to build new marketing strategies based on accepted scientific findings.

The goal of our research is to assess (i) whether knowledge/awareness of the health benefits of soy proteins play any role in consumers’ decisions to consume soy-based food products and/or the consumption intensity, if they consume, and (ii) whether FDA’s decision to allow food manufacturers to use health claims influence consumers’ willingness to participate in soy-based food market or willingness to increase, if they are currently consuming such foods. The role of health knowledge in food consumption can be empirically assessed using aggregate time-series data. Using aggregate time-series data, several previous studies used a time trend or proxy representing health information to test or account for the changes in preference for specific food product (Brown and Schrader, 1990; Chern, Loehman and Yen, 1995; Kinnucan, 1997). Alternatively, disaggregate cross-sectional data can be collected to identify individual or household differences in responding to evolving information and in modifying behavior (e.g., Putler and Frazao, 1994; Variyam et al., 1999). Given that the FDA approved the health benefits
of soy-based foods only in 1999, analysis using aggregate time-series data aimed at comparing potential changes in consumption over time or prior to and after the approval does not seem to be appropriate at this particular timing of the unfolding of the new information. Therefore, this paper uses cross-sectional data to accomplish our objectives.

**Literature Review: Health Information, Regulation and Food Consumption**

The growing demand for healthy foods coupled with accumulating scientific knowledge on the link between diet and disease touched off a major change in the marketing strategies of firms in the food sector. Addressing health concerns became a central theme to various new marketing strategies in the food industry (e.g., new product development, market segmentation, advertising or promotion). For example, some food products carry health claims associating certain nutrients to the reduced risks of specific diseases. These new marketing strategies underscoring the health or nutritional attributes of food products sparked a controversy over the truthfulness of such health claims. As a consequence, whether and how to regulate nutritional labeling and health claims on food products have become public policy concerns.

The evolving interrelationship among changes in consumer preferences, marketing strategies of food companies, and regulations has attracted considerable academic interest. Theoretically, the most fundamental condition for the efficient operation of a market is the free flow of truthful information among market participants (Stiglitz, 1996). Information for private goods is provided primarily through labeling, advertising, promotion, or warranty program. For the particular case of nutritional and health information in food markets, the public nature of such information and the potential for deception support a role for government. Some form of regulation or oversight is usually necessary to facilitate the flow of information and to ensure the
truthfulness of the information. While labeling panels for the nutritional contents of food products were required to assist consumers in food choices, the concern over potentially deceptive messages initially led the FDA to ban all health claims on food products. Such a measure of regulation with regard to health claims was seen by the food industry and public interest groups as causing inefficiencies in the flow of information. Hence, the key issue of interest to policy-makers was whether to permit specific health claims on nutritional labeling on food products. Closely linked to this issue was how to best communicate the growing diet-disease information to public. A series of studies illustrate the effectiveness of the private provision of diet-disease information and present evidence that regulation impeded the flow of information (Ippolito and Mathios, 1990, 1991, 1993, 1995, Mathios, 1998). For example, a study about the market for breakfast cereal showed how marketing strategies using health claims contributed to enhancing consumer knowledge of the relationship between diet and disease and affected the demand for breakfast cereal. The accumulation of scientific evidence linking dietary fiber to colon cancer led the breakfast cereal industry to use health claims on their products. Since this strategy was in direct violation of FDA policy which had essentially banned health claims on food products, the industry's initiative provoked a full review of public policy toward health claims in food promotion. Subsequently, the ban against health claims was suspended in 1985. Ippolito and Mathios (1991) demonstrated that the marketing practices of using health claims led to significant increases in consumer knowledge of the fiber-cancer relationship, in the increased consumption of fiber cereal, and in product innovation. In light of the significance of the change in the regulation of health claims, Ippolito and Mathios (1995) also showed that the consumption of fats and saturated fats decreased faster after 1985 as compared to the 1977-1985
Survey Design

Survey Instrument A survey instrument was designed to measure various conceptual variables pertinent to accomplishing the objective of this research. The survey instrument consist principally of two sections. The first section measures health status, motivation and knowledge in general along with perceived knowledge of health benefits specific to soy-based foods. Question items measuring general health knowledge are drawn from Mooreman and Maulitch (1993) while measures of health status and motivation are constructed from previous surveys related to food consumption. There have been no published studies measuring consumer attitudes or perceptions specifically related to soy-based foods. Therefore, knowledge/awareness of the health benefits of soy foods and perceptions about other attributes including taste, price, or convenience are measured using question items generated for this project. In addition, frequency measures of the consumption of various soy-based foods (i.e. Tofu, Soy veggie burgers, Soy milk, Soy Protein bars, Soy supplements, Soy cheese, and Meat Substitutes) are elicited from respondents. Finally, the survey instrument elicits information on demographic characteristics including education, age, gender, income, place of residence, and ethnic background. Such demographic profiles may impact the consumption of soy-based foods directly as well as indirectly via their effects on general health knowledge or knowledge/awareness of the health benefits of soy protein.

Using split-sample technique, the second section evaluates the value of the FDA regulatory action regarding the health benefits of soy-based foods. The survey instrument is designed in such a way that half of the sample is exposed to the following information:
The Food and Drug Administration (FDA) officially confirmed the health benefits of soy-based foods with a 1999 ruling that food manufacturers can claim 25 grams of soy protein a day as part of a diet low in saturated fat and cholesterol may reduce the risk of heart disease. This ruling is based on the scientific finding that this soy protein dosage reduced blood cholesterol levels by 9.3% and the risk of coronary heart disease by 18.6%.

Given such information, consumers' behavioral intentions are measured with questions asking (i) willingness-to-try soy-based foods for non-consumers, (ii) willingness-to-include soy-based foods regularly in their diets for infrequent consumers, and (iii) willingness-to-increase soy-based foods for current consumers. In contrast, the remaining half of the sample answers the same questions given their current preferences and knowledge of soy-based foods (i.e., without being exposed to the FDA decision about the health benefits of soy-based foods).

Data Collection

This paper uses data collected from a pilot study (convenience sample) drawn from a Midwest college town. Two questionnaires (i.e., one with information about the FDA decision and the other without it) were given to students taking introductory Marketing courses. They were asked to administer the surveys to adult, non-student subjects and to verify that respondents have answered all survey items. At the conclusion of the survey, the students were instructed to obtain the first name and phone number of the respondent so that we can verify that they actually participated in the survey. Upon verification, the students receive extra credit that contributes to their total course scores. Total 200 students participated in this project (i.e., 400 questionnaires were distributed) and 315 respondents returned completed questionnaires.

Model Specification

Two-stage Model

When modeling the consumption of soy-based foods using disaggregate cross-sectional data, an issue of considerable importance arises concerning how to
deal with respondents who do not consume any soy-based foods. That is, some consumers may report a zero level of consumption over a specific study period. Deleting those zero observations from statistical analysis can lead to sample selection bias in parameter estimates (Heckman, 1979). More importantly, the issue of dealing with zero observations is directly related to what type of behavioral assumptions are maintained regarding food consumption. For example, the decision to consume or not soy-based foods and the intensity of consumption may be ruled by two different probability distributions and different set of explanatory variables. To deal with potential sample selection bias associated with zero observations and to account for respondents who report that they do not consume any soy-based foods, this research will consider two-stage decision model, the double-hurdle framework developed by Cragg (1971). To take into account of the frequency nature of the soy-based food consumption that will be measured in the survey, we use hurdle count data model that incorporates Cragg's framework (Mullahy, 1986).

Specifically, the following likelihood function representing geometric distribution is estimated,

\[
L = \prod_0 \{1/[1 + \exp(Z, \beta_1)]\} \times \prod_1 \{\exp(Z, \beta_1)/[1 + \exp(Z, \beta_1)]\} \times \prod \exp[(y - 1)Z, \beta_2]/\{1 + \exp(Z, \beta_2)\}^{y-1} = \prod_{1}^{Z, \beta_1} \times \prod_{2}^{G, Z, \beta_2}
\]

where \( y \) is a random variable indicating the consumption frequency of soy-based food products; \( Z_t \) is a vector of explanatory variables affecting both participation and consumption decisions; \( \Pi_0 \) and \( \Pi_1 \) indicate multiplications over zero (0) and positive (1) observations of soy food consumption; \( \beta_1 \) and \( \beta_2 \) are parameter vectors associated with participation and consumption equations, respectively.
The two-stage models permit the researcher to empirically estimate the effects of perceived health benefits, demographic profiles and other attitudinal variables on two interrelated decisions: (1) whether to consume (i.e., the consumers desire to participate) and (2) contingent upon the decision to participate, how much to consume soy-based foods (intensity decision). With the flexibility of the two-stage decision model, this research can delve into the factors shaping the market penetration as well as the consumption intensity of soy-based foods.

In a cigarette consumption study, Jones (1989) laid out a behavioral basis for the use of two-stage decision making process that there might be certain characteristics of smoking which relate directly to the qualitative distinction between smokers and nonsmokers and which are independent of the quantity consumed. Since then, quite a few studies employed the double-hurdle or Heckman sample selection models and demonstrated the usefulness of the two-stage decision frameworks in analyzing food consumption behavior. Gould (1992) used the two-stage decision model to examine economic and nutrition-related attitudinal factors affecting cheese consumption. Blisard and Blaylock (1993) analyzed the expenditure on butter consumption using the double-hurdle model. Lin and Milon (1993) used the double-hurdle model to estimate the effect of safety perception on the consumption of seafoods (oyster and shrimp). Ward and Moon (1995) applied the double-hurdle and sample section models to address the roles of generic advertising and health concerns in the U.S. beef consumption.

An advantage of the two-stage decision models is that, while the same set of explanatory variables may be included in the two equations, they are allowed to have differential effects between the two decisions. For example, several hypotheses can be proposed concerning the potential effects of perceived health benefits of soy-based foods in the two-stage decision
models: (1) perceived health benefits exert a strong and positive influence on both decisions, raising market penetration and promoting the consumption level of existing consumers; (2) perceived health benefits significantly influence only the decision of whether to consume soy-foods, while having little impact on the decision of how much to consume; and (3) perceived health benefits do not increase the probability of becoming a consumer of soy foods but motivate existing soy consumers to increase their level of consumption.

Measurement of Variables  Equation (1) shows that the vector $Z_t$ determines both participation and intensity decisions regarding soy-based food consumption. In our research, the vector includes consumer perceptions about various attributes associated with soy foods. Health benefits, price, taste, availability, and convenience are identified as salient attributes of soy-based foods. Consumer perceptions about such attributes were measured using a five-point scale ranging from Disagree strongly= to Agree strongly=. Table 1 shows variable descriptions and sample statistics. Only 46 percent of the respondents consumed at least one type of soy-based food products (Tofu, Soy veggie burgers, Soy milk, Soy Protein bars, Soy supplements, Soy cheese, and Meat Substitutes) during the one month recall period. The percentages of consuming individual products at least once a month are: Tofu (20 %), Soy veggie burgers (27.6 %), Soy milk (15.6 %), Soy protein bars (14.3 %), Soy supplements (14.3 %), Soy cheese (11.6 %) and Meat substitutes (25 %). Fifty-five percent of the respondents had attended some college and sixty-three percent had an annual household income between $35,000 ~ $ 49,999. Less than 6 percent of the respondents were non-White.

Estimation Results

Hurdle count model  Table 2 shows parameter estimates for the participation and
consumption intensity equations along with other summary statistics. Results indicate that the
decision of whether or not to participate in the soy-based food market is shaped by perceptions
about health benefits, taste and convenience and the level of educational attainment. Consumers
who perceive health benefits are more likely to consume soy-based food products when
compared to those perceiving no benefits. This result indicates that awareness of health benefits
motivates consumers (mostly white) to choose soy-based food products which are not part of
traditional meals in the US. Perceptions about taste was also a significant factor influencing the
likelihood of participation in the soy-based food market. Respondents who perceive soy-based
food products distasteful are more likely to stay away from the market as compared to those who
consider them tasteful. Among demographic profiles included in the empirical model, only
education and age had a statistically significant effect on the probability of consuming soy-based
foods. The higher the educational level is, the more likely to participate in the soy-based food
market. The older the consumers is, the less likely to consume soy-based food.

As expected, perceptions about health benefits had a statistically significant and positive
effect on the consumption frequency of soy-based foods. Hence, in conjunction with the
previous result of the participation equation, the result demonstrates that awareness of health
benefits affects both participation and consumption decisions. In addition to health benefits,
perceptions about cooking convenience was a statistically significant factor constraining the
consumption of soy-based foods. Interestingly, perceived taste did not influence the
consumption frequency decision, indicating that, while taste is a major determinant of
participation, it becomes irrelevant in the consumption intensity decision.

The Role of FDA Ruling  As stated earlier, the survey instrument exposes half of the
sample to the FDA decision allowing food manufacturers to use health claims on soy foods and asks whether they would be willing to increase the consumption of soy-based foods if the respondents are regular consumers.\(^1\) If they are non or infrequent user of soy-based foods, such respondents are asked whether they would be willing to include soy-based foods more regularly in their diets. The rest of the sample is asked the same questions without being exposed to the information about the FDA decision. This split sample technique allows us to determine whether the FDA regulatory action influences behavioral intentions as measured with willingness-to-increase or willingness-to-include. The two measures were coded with a five-point scale ranging from definitely would to definitely would not.

Given the categorical and ordered nature of responses, two ordered probit regression models were developed corresponding to willingness-to-increase and willingness-to-include. In addition to the perceived attributes specified in the previous section, the regression models include a binary variable representing one if information about the FDA decision was given and zero if not. The binary variable was statistically significant in the willingness-to-increase equation (\(t\)-value=4.54) but not in the willingness-to-include equation (\(t\)-value=0.78). This result indicates that respondents (regular users of soy-based foods) who were exposed to the FDA decision would be more inclined to increase their consumption of soy-based foods as compared to those who were not exposed to such information. Yet the information about FDA decision did not influence the behavioral intentions of infrequent- or non-consumers.

\(^1\) Prior to the information on the FDA decision is introduced, a screening question was asked to determine whether or not the respondents are regular consumers of soy-based foods.
References


Heckman., J. 1979. Sample Selection Bias. @Econometrica


Senauer B. 1990. *Major Consumer Trends Affecting the U.S. Food System.* @Journal of
Agricultural Economics.


Table 1. Description of variables and summary statistics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean (Std. Dev.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variable</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation Frequency</td>
<td>Percentage of respondents reporting above zero</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>Consumption frequency in a one month recall period</td>
<td>9.30 (12.4)</td>
</tr>
<tr>
<td><strong>Perceived Attributes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taste</td>
<td>I like the taste of soy-based foods</td>
<td>2.52 (1.15)</td>
</tr>
<tr>
<td>Health Benefits</td>
<td>Soy-based foods offer health benefits</td>
<td>3.69 (0.98)</td>
</tr>
<tr>
<td>Price</td>
<td>Soy-based foods are inexpensive</td>
<td>2.58 (1.03)</td>
</tr>
<tr>
<td>Convenience</td>
<td>Soy-based foods are convenient to cook</td>
<td>2.79 (1.03)</td>
</tr>
<tr>
<td>Recipes</td>
<td>Recipes that use soy-based foods are readily available</td>
<td>2.74 (1.07)</td>
</tr>
<tr>
<td><strong>Demographic Profiles</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>1=less than $20,000, 2=$20,000-$34,999, 3=$35,000-$49,999, 4=$50,000-$64,999, 5=$65,000-79,999, 6=more than $80,000</td>
<td>3.12 (1.71)</td>
</tr>
<tr>
<td>Education</td>
<td>1=Grade sch. 2=Some high sch. 3=High sch grad. 4=Some college. 5=College grad. 6=Postgrad.</td>
<td>4.40 (0.97)</td>
</tr>
<tr>
<td>Age</td>
<td>In years</td>
<td>37.5 (13.28)</td>
</tr>
</tbody>
</table>

Note: Perceived attributes were measured with a five-point scale ranging from *Disagree completely* to *Agree Completely*.
Table 2. Parameter estimates from the hurdle count data model of soy-based food consumption.

<table>
<thead>
<tr>
<th></th>
<th>Participation</th>
<th>Consumption frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter Estimates</td>
<td>Asymptotic t-values</td>
</tr>
<tr>
<td>Constant</td>
<td>-4.678</td>
<td>-4.52***</td>
</tr>
<tr>
<td>Taste</td>
<td>0.699</td>
<td>4.71***</td>
</tr>
<tr>
<td>Health Benefits</td>
<td>0.459</td>
<td>2.78**</td>
</tr>
<tr>
<td>Price</td>
<td>0.032</td>
<td>0.46</td>
</tr>
<tr>
<td>Convenience</td>
<td>0.074</td>
<td>0.42</td>
</tr>
<tr>
<td>Recipes</td>
<td>0.170</td>
<td>1.07</td>
</tr>
<tr>
<td>Income</td>
<td>0.056</td>
<td>0.79</td>
</tr>
<tr>
<td>Education</td>
<td>0.259</td>
<td>1.78*</td>
</tr>
<tr>
<td>Age</td>
<td>-0.025</td>
<td>-2.05*</td>
</tr>
<tr>
<td>Gender</td>
<td>0.234</td>
<td>1.19</td>
</tr>
<tr>
<td>Log-L value</td>
<td>-159</td>
<td></td>
</tr>
<tr>
<td>LR Test</td>
<td>62.3 (p=.000)</td>
<td></td>
</tr>
<tr>
<td>No. of Obs</td>
<td>278</td>
<td></td>
</tr>
</tbody>
</table>

*** $P < 0.01$; ** $P < 0.05$; * $P < 0.1$. 