New Economic Approaches to Consumer Welfare and Nutrition
A Food & Agricultural Marketing Consortium Conference
Sponsored by the Economic Research Service

January 14-15, 1999

New Economic Approaches to Consumer Welfare and Nutrition

Conference Proceedings

Food and Agricultural Marketing Consortium
Alexandria, Virginia
January 14-15, 1999

Consumer Label Use and Diet Quality: An Endogenous Switching Regression Analysis

Sung-Yong Kim*
Rodolfo M. Nayga, Jr.
Oral Capps, Jr.
Beverly Tepper

November 1998

* Kim, Nayga, and Capps are Graduate Assistant, Assistant Professor, and Professor, respectively, Department of Agricultural Economics, Texas A&M University. Tepper is Associate Professor, Department of Food Science, Rutgers University. Invited Paper presented at the Food and Agricultural Marketing Consortium Conference on *New Economic Approaches to Consumer Welfare and Nutrition*, Alexandria, Virginia, January 14-15, 1999. The authors thank Brian Gould and Jay Variyam for helpful comments.

Consumer Label Use and Diet Quality: An Endogenous Switching Regression Analysis

Abstract

This study examines the impact of food label use on diet quality of Americans. Using an endogenous switching regression model, results indicate that label use improves diet quality by as much 14 points on a 100 point Healthy Eating Index scale. The analysis also was conducted for each type of information contained on the labels: list of ingredients, nutrient content health claims, nutritional panel, serving size, and health benefit statements. Use of health benefit statements on food labels provides the highest level of improvement in diet quality.

Introduction

Many Americans are not meeting dietary recommendations. The U.S. Department of Agriculture revealed that only about 12 percent of Americans are eating healthfully. This statistic is a concern because four of the top ten causes of death in the United States – heart disease, cancer, stroke, and diabetes – are associated with poor diets. Diet-related health conditions cost society an estimated $250 billion annually in medical costs and lost productivity (Food Review).

Concerns about the effect of diet on health have resulted in the legislation of the Nutritional Labeling and Education Act (NLEA) and its implementation in 1994. The NLEA instituted sweeping changes to replace the voluntary system of labeling established by the Food and Drug Administration (FDA) in 1973 (Nayga, Lipinski, and Savur). Hence, the NLEA was implemented to provide consistent, understandable, and usable nutritional labels that can help consumers make healthier food choices and, therefore, improve the quality of their diets.

Previous analyses on the effect of government programs have been focused on the Food Stamp, National School lunch, and Federal Transfer programs (Akin et al.; Bulter, Ohls, and Posner; Morgan; Devancy and Fraker; Long). Scant information, however, is available concerning the effectiveness of the NLEA in improving the diet quality of Americans. To date, no known empirical work has been conducted on this topic. The purpose of this study is to assess the effectiveness of the NLEA in terms of diet quality improvements. Specifically, this study will attempt to determine the characteristics of consumers who use nutritional labels as well as to evaluate the effect of consumer label use on diet quality as measured by the Healthy Eating Index (HEI) developed by the USDA.

An endogenous switching regression model is employed to estimate the effect of consumer label use on diet quality. The model explicitly controls for the likely unobservable heterogeneity of label users versus non label-users. There are five types of nutritional labels: nutritional panel; information about the size of a serving; nutrient content health claim; the list of ingredients; and statement regarding health benefits. These five types of nutritional label information are examined in the study, in addition to general label use, to determine which type of label information provides the most improvement, if any, in diet quality.

This paper is organized as follows. The theoretical framework, based on Grossman's demand for health, is presented in the second section. The structure of the model and the endogenous switching regression equations are presented in the third section. The estimates, based on data from 1994-96 Diet Health Knowledge Survey, are discussed in the fourth section. Policy implications and conclusions are discussed in the fifth section.

The Theoretical Framework

The household production theory developed by Becker (1965) and the theory on health input demand developed by Grossman (1972) provide the theoretical framework for this study. In Grossman's model, health is both demanded and produced by consumers, and diet is regarded as a health input. Consumers combine various inputs to produce his or her health (commodity), so as to maximize a utility function. Maximization of utility subject to the constraint of health production and resource constraints generates individual demand functions for the health inputs.

Assume that a household preference orderings over health, foods consumed, and non-foods can be characterized by the utility function subject to the usual properties[1]:

(1) $$U = U(\boldsymbol{x}, \boldsymbol{y}, h)$$

where $\boldsymbol{x}$ is a vector of foods consumed, $\boldsymbol{y}$ is a vector of non-foods and h reflects an individual's health status. Health is a choice variable because it is a source of utility and because it determines income and wealth levels. The health production function relates an output of health to diet and non-food health inputs such as exercise and cigarette smoking:

(2) $$h = \Gamma(HEI, \boldsymbol{z} ; \boldsymbol{d}, u)$$

where *HEI* is the healthy eating index, a measure of overall diet quality of the individual, ***z*** is a vector of non-food health inputs which do not augment utility other than their effects on *h* (e.g., exercise)[2]. The production function is affected by the efficiency or productivity of a given consumer, that is the amount of health obtained from a given amount of health inputs-as reflected by ***d***, his or her personal characteristics, and u, exogenous health endowments known to the individual or household but not controlled by them. Expenditures are constrained to equal household income:

(3) $$\boldsymbol{x}\, p_x + \boldsymbol{y}\, p_y = M$$

where ***p*** indicates a vector of prices and *M* is household income.

Given household income and market prices, maximizing utility function (1) subject to the constraints results in an *HEI* equation that is a function of prices, income, personal and household characteristics, and *u*.

Introducing nutritional label use explicitly into the model reflects its role as a factor mediating part of the causality from personal characteristics to his or her health (Variyam et al., 1998, p.3). For example, more educated persons have higher productivity in producing health because they are more informed about the true effects of diet on health. Thus, they have the ability to select a better diet mix through nutritional label use. Making the role of information on the labels explicit, the reduced form *HEI* equation is written as

(4) $$HEI = f\,(\boldsymbol{p}, M ; \boldsymbol{d}, l, u)$$

where ***p*** is a vector of prices, *M* is the household income, and *l* is label use index.

The Econometric Model

In evaluating the benefits of label use, a model that can be employed is the following [3]:

(5) $$HEI = X\beta + \delta I + \varepsilon$$

where X is a vector of exogenous personal characteristics and I is a dummy variable (I=1 if the individual uses nutritional label when shopping; I=0 otherwise). However, this model is very restrictive, because the label use decision may create interaction effects with observed or unobserved personal characteristics (Maddala). If the label use decision is based on individual self-selection, it is likely that label users have systematically different characteristics from non-users. This sub-sample heterogeneity is econometrically problematic when unobserved characteristics are distributed differently across label users and non-label users. Thus, unobserved variables may influence both label use decision and diet behavior, resulting in inconsistent estimates of the effect of label use on diet quality.

An endogenous switching regression model avoids these problems (Gould and Lin). It consists of diet quality equations for label users and non-label users, and an equation for the label use decision. Define HEI_i as the observed diet quality; HEI_1 and HEI_0 as the diet quality of label user and non-label user, respectively; I_i^* as a latent variable that determines label use decision; I_i as an indicator variable that equals one if consumer uses nutritional labels and equals zero otherwise; X as a vector of observed characteristics that affect diet quality and Z as vector of characteristics that affect label use. The endogenous switching regression model is written as

(6) $$HEI_1 = X'\beta_1 + \varepsilon_1$$

(7) $$HEI_0 = X'\beta_0 + \varepsilon_0$$

(8) $$I^* = Z\gamma + \mu$$

$$I = 1 \text{ if and only if } I^* > 0$$

$$= 0 \text{ if and only if } I^* \leq 0$$

The observed *HEI* is defined as

$HEI_i = HEI_{1i}$ if and only if $I = 1$

$HEI_i = HEI_{0i}$ if and only if $I = 0$

The error terms of the above equations, ε_1, ε_0, and μ are assumed to have a trivariate normal distribution, with mean vector zero and covariance matrix

$$\operatorname{cov}(\varepsilon_1, \varepsilon_0, \mu) = \begin{matrix} \sigma_{11}^2 & \sigma_{10} & \sigma_{1u} \\ \sigma_{10} & \sigma_{00}^2 & \sigma_{0u} \\ \sigma_{1u} & \sigma_{0u} & 1 \end{matrix}$$

Since the choice of using labels or not is endogenous, the error terms in equation (6) and (7), conditional on the sample selection criterion, have a nonzero expected value. Thus OLS estimates of β are biased. Sample selection corrected Healthy Eating Index equations are specified following Lee (1976):

(9) $$HEI_1 = X\beta_1 + \sigma_{1u}\frac{\phi(Z'\gamma)}{\Phi(Z'\gamma)} + \xi_1 \qquad if \quad I = 1$$

(10) $$HEI_0 = X\beta_0 + \sigma_{0u}\frac{-\phi(Z'\gamma)}{1-\Phi(Z'\gamma)} + \xi_0 \qquad if \quad I = 0$$

where the new residuals

$\xi_1 = \varepsilon_1 + \sigma_{1u}$

$\xi_0 = \varepsilon_0 + \sigma_{0u}$

are uncorrelated. The two-stage procedure for estimating (9) and (10) involves first calculating the Mill's ratio, $\phi(Z'\gamma)/\Phi(Z'\gamma)$ and $-\phi(Z'\gamma)/1-\Phi(Z'\gamma)$, using probit estimates of (8). The ordinary least squares method is used next to estimate (9) and (10). Since the variables, $\phi(Z'\gamma)/\Phi(Z'\gamma)$ and $-\phi(Z'\gamma)/1-\Phi(Z'\gamma)$ have already been estimated, however, the residuals ζ_1 and ζ_0 in equation (9) and (10) cannot be used to determine the variances of the two-stage estimates. Thus the variance-covariance matrix is adjusted using the procedure described by Maddala.

The endogenous switching regression model separates the total marginal effect of an exogenous variable on the average diet quality of label users into two parts: a direct effect on the mean of HEI_1, and an indirect effect from the label use or non-label use decision that appears as a result of correlation in the unobservable components of *HEI* and *I* (Poirier and Rudd):

$$\text{(11)} \qquad \frac{\partial \boldsymbol{E}(\boldsymbol{HEI}_1 \mid \boldsymbol{I} = 1)}{\partial \boldsymbol{X}_{1i}} = \beta_{1i} - \gamma_i \sigma_{1u} \frac{\phi(\boldsymbol{Z}'\gamma)}{\Phi(\boldsymbol{Z}'\gamma)} \left[\boldsymbol{Z}'\gamma + \frac{\phi(\boldsymbol{Z}'\gamma)}{\Phi(\boldsymbol{Z}'\gamma)} \right]$$

Healthy Eating Index

The Healthy Eating Index (HEI), developed by U.S. Department of Agriculture (USDA) Center for Nutrition Policy and Promotion, provides an overall picture of the type and quantity of foods people eat, their compliance with specific dietary recommendations, and the variety in their diet (Bowman, p.2). Thus, the HEI is a summary measure of people’s overall diet quality. It serves as a performance measure for the effectiveness of nutrition intervention effort to improve dietary behaviors. The HEI is composed of 10 components, each representing different aspects of a healthful diet (Kennedy et al.). Components 1-5 measure the degree to which a person’s diet conforms to USDA’s Food Guide Pyramid serving recommendations for the five major food groups-grains, vegetables, fruits, milk, and meat. Components 6 and 7 measure total fat and saturated fat consumption as a percentage of total food energy intakes. Components 8 and 9 measure total cholesterol and sodium intake, and component 10 examines variety in a person’s diet. Data used to calculate the HEI are from the USDA’s 1994-96 Continuing Survey of Food Intakes by individuals, a nationally representative survey containing information on people’s consumption of foods and nutrients. Each component has a possible range of 0 to 10. The maximum overall score is 100. High component scores indicate intakes close to recommended ranges or amounts; low component scores indicate less compliance with recommended ranges or amounts. The mean HEI score is 63.6 for 1994, 63.5 for 1995, and 63.8 for 1996. More information about how the HEI was created are available from the authors upon request.

Data

Besides the use of the 1994-96 CSFII data for the HEI variable, the 1994-96 Diet and Health Knowledge Survey (DHKS) data, the companion data of the CSFII, also are used in this study. The DHKS includes detailed information about the individual's socioeconomic background and questions on label usage. The empirical work uses DHKS respondent files which completed the survey of both day1 and day2 intakes. Due to incomplete data for some of the variables, 5405 observations are used in the analysis.

The name, definitions, and means for principal variables are exhibited in Table 1. The dependent variables include the Healthy Eating Index (HEI) developed by USDA, and a binary label use variable. The mean of HEI score is 63.72 for label users and 58.93 for non-label users. The analysis also is disaggregated by type of information contained on food labels. The five types of information that are presented on the food label are (1) the list of ingredients; (2) the short phrases on the label like "low-fat" or "light" or "good source of fiber"(health claims); (3) the nutrition panel that tells the amount of calories, protein, fat, and such in a serving of the food; (4) the information about the size of serving; and (5) the statement on the label that describes health benefits of nutrients or foods. About 76.4% of the sample used the list of ingredients, 74.2% used the health claims, 75.4% used the nutrition panel, 67.6% used the serving size, and 67.9% used the health benefit statements. About 82.9 % of the sample used at least one of these information on the label, defined as the general use of label. Binary variables (1=use; 0=not use) are used to capture the decision to use each type of information on the food labels.

Independent variables consist of personal or household characteristics, demographic factors, participation in government programs such as the Food Stamp Program, and knowledge about the food guide pyramid. Personal or household characteristics include body mass index, age, gender, level of education, ethnicity, race, exercise status, smoking status, employment status, and special diet status[4]. Other demographic factors include region, urbanization, household size, and income.

Consumer's knowledge about the Food Pyramid Guidelines (PYRAMID) is constructed as a measure of diet-health knowledge. The variable measures how much consumers know the servings recommended for the five primary food groups (grains, fruits, vegetables, dairy and meat) in the Food Guide Pyramid. Since PYRAMID reflects the answers to 5 questions, the variable has values ranging from 0 to 5.

The other variables in the label use probit equation are a dummy variable indicating whether the individual is a major shopper or not and a variable reflecting consumer's awareness about the linkage between diet and health (DHA). The variable, DHA is constructed following Variyam et al. The seven questions in the DHKS used to construct the DHA variable take the general form: "Have you heard about any health problems that might be related to being overweight and how much of a particular nutrient (such as fat, fiber, salt, calcium, cholesterol, and sugar) a person eats? Each answer of "Yes" is given a value of one while each answer of "No" is given a value of zero. Since DHA reflects the answers to 7 questions, the variable has a lower limit of zero and an upper limit of seven.

Empirical Results

First Stage Probit Label Use Model

Estimates of the first stage probit model for the general label use and for each of the five types of information on labels are presented in Table 2. The table also contains goodness-of-fit measures. For the sake of brevity, only the results of the general use probit model are discussed. This model correctly predicts the label use behavior for 84% of the sample. Based on the results, there is a nonlinear relationship between income and label use. The probability of label use increases with income until an income level of about $59,800 before it declines with subsequent increases in income. Also, a nonlinear relationship exists between age and general label use. The probability of label use increases until age 44 before declining with subsequent increases in age.

Consistent with Nayga's finding, males are less likely to use labels than females. Results also indicate that education is significantly and positively related to label use. This finding is consistent with those of Guthrie et al. Urbanization and regional differences also are evident in the results. Specifically, individuals who reside in nonmetro areas are less likely to use labels than those who reside in suburban areas. In addition, individuals who reside in the South are less likely to use labels than individuals from other regions.

Non-hispanics are less likely to use labels than others. Individuals who are on a special diet are more likely to use labels than individuals who are not on a special diet. Individuals who are more informed about the link between diet and health also are more likely to use nutritional labels. This result is consistent with the argument that poorly informed consumers tend to underestimate the marginal benefit of label use. Major food shoppers are more likely than others to use information on the label when shopping than others. This finding is comforting since a household's major food shopper can potentially influence the quality of the diet of individual household members just from the types of foods he or she decides to purchase.

Second Stage HEI Models

The second-stage estimates of the endogenous switching-regression model for general label use and for the different types of information on labels are exhibited in Table 3. The results are generally consistent across the equations. The parameter estimates for education, body mass index, exercise, food stamp participation, and knowledge about the food guide pyramid are statistically significant and have the expected signs in the model for label users. In the model for non label-users, these coefficients are insignificant, but the estimates for age, household income, the level of urbanization (i.e., city), and some regions (i.e., midwest) are statistically significant.

Results based on the general label use model indicate that income is positively related to diet quality (i.e., HEI) of label users. Income is not significant in the non-label user equation. Black label users and non-label users have HEIs that are about three points lower than the HEIs of white label users and non-label users, respectively. Label users of other races, however, have higher HEI than white label users.

Employed label users have a lower HEI than unemployed label users. The reason for this result is not clear. However, it is possible that the diet quality of employed label users is lower because they do not have as much time as the unemployed to spend on food shopping to make the more appropriate decisions regarding the quality of foods they need to buy.

Non-label users from central cities have an HEI that is more than two points higher than non-label users from suburban areas. On the other hand, label users from nonmetro areas have an HEI that is about 1.5 points lower than label users from suburban areas. Regionally, label and non-label users from the northeast have higher HEIs than those from south. Label users from the west and non-label users from midwest also have higher HEI than their counterparts from the south.

Nonhispanic label users have a HEI that is almost two points lower than hispanic label users. More importantly, food stamp participants who are label users have a HEI that is almost three points lower than non-food stamp participants who are label users. This result implies that the food stamp program does not improve the diet quality of participants to the level of non-participants, despite the use of the labels. This finding is consistent with that of Butler and Raymond. They observed that, controlling for participation in the food stamp program, nutrition intake is negatively affected by food stamp income for a sample of elderly people.

Body mass index is negatively related to HEI for label users. As expected, those who are on special diet have higher HEIs than those who are not on a special diet. In addition, label users who regularly exercise have a higher HEI than label users who do not exercise. Smokers, whether label or non-label users, have HEI which are more than four points lower than those of non-smokers. Label users with higher knowledge about the food guide pyramid also have higher HEIs.

Self-selection occurs in both label user and non-label user equations because the Mill’s ratios (variable lamda) are all statistically significant. These estimates imply that self-selection bias could have occurred if the endogenous switching model was not employed in the estimation of the equations.

Label Use and Diet Quality Improvements

To evaluate the benefit of label use, two types of analyses are conducted. One analysis is conducted to evaluate the impact of label use on the HEI of label-users if they had not used the labels (Table 4). Another analysis is done to determine the potential impact of label use on the HEI of non-label users if they are induced to use the labels (Table 5).

We can calculate the total gross benefit of label use on label users by comparing the outcome when using the label and the expected potential outcome when not using the label. For a label user with characteristics *X* and *Z*, the expected value of HEI_1 is

$$(12) \qquad \boldsymbol{E}(\boldsymbol{HEI}_1 \mid \boldsymbol{I}=1) = \boldsymbol{X}'\beta_1 + \sigma_{1\boldsymbol{u}} \frac{\phi(\boldsymbol{Z}'\gamma)}{\Phi(\boldsymbol{Z}'\gamma)}$$

The expected potential value of HEI if the consumer has not used the labels is

$$(13) \qquad \boldsymbol{E}(\boldsymbol{HEI}_0 \mid \boldsymbol{I}=1) = \boldsymbol{X}'\beta_0 + \sigma_{0\boldsymbol{u}} \frac{\phi(\boldsymbol{Z}'\gamma)}{\Phi(\boldsymbol{Z}'\gamma)}$$

Thus, the expected gross benefit in terms of diet quality due to label use, evaluated at the means of *X* and *Z* for label users, is

$$(14) \qquad \boldsymbol{E}(\boldsymbol{HEI}_1 \mid \boldsymbol{I}=1) - \boldsymbol{E}(\boldsymbol{HEI}_0 \mid \boldsymbol{I}=1) = \boldsymbol{X}'(\beta_1 - \beta_0) + (\sigma_{1\boldsymbol{u}} - \sigma_{0\boldsymbol{u}}) \frac{\phi(\boldsymbol{Z}'\gamma)}{\Phi(\boldsymbol{Z}'\gamma)}$$

The observed effects of label use are decomposed into a structural effect (the first term in the above equation) and an effect through the unobservable (the second term). The second term takes into account the differences in the diet quality of label users and non-label users that may not be attributable to label use but rather to the unobserved characteristics that led them to use the label.

The mean differences are reported in Table 4. The effects of consumer label use on diet quality also are estimated for each of the five types of information on the labels. Consumer label use increases the average expected diet quality by a range of 9.06 and 13.96 points, depending on the type of information. Improvement in the diet is highest when consumers use health benefit statements on the labels.

In terms of the distribution, 54 % of the sample get an improvement of between 5 to 10 points in diet quality, while 39 % get an improvement of 10 to 20 points when using labels (see General Label Use column in Table 4). For the types of information, about three-fourths of the sample get an improvement of 10 to 20 points when using ingredient, health claims, and nutrition panel, while 92 % to 98% of the sample get an improvement of 10 to 20 points when using information concerning serving size or health benefit statement in the label.

The expected improvement in diet quality of non-label users if they are induced to use the labels can be measured by the following expression:

$$(15)\quad \boldsymbol{E}(\boldsymbol{HEI}_1 \mid \boldsymbol{I}=1) - \boldsymbol{E}(\boldsymbol{HEI}_0 \mid \boldsymbol{I}=0) = \boldsymbol{X}'(\beta_1 - \beta_0) + \sigma_{1\boldsymbol{u}} \frac{\phi(\boldsymbol{Z}'\gamma)}{\Phi(\boldsymbol{Z}'\gamma)} - \sigma_{0\boldsymbol{u}} \frac{-\phi(\boldsymbol{Z}'\gamma)}{1-\Phi(\boldsymbol{Z}'\gamma)}$$

.

The difference in the expected HEIs is calculated at the means of *X* and *Z* for non-label users. Thus, this expression indicates the potential increase in diet quality due to use the labels among non-label users (i.e. the latent demand for label use among non-label users). Based on the results exhibited in Table 5, the average expected diet quality improvement ranges from 1.60 (general label use) to 3.57 (nutrition panel) points. Therefore, if consumers are induced to use the information on the nutrition panel, their average diet quality is expected to improve from 57.58 to 61.16 points in the HEI scale.

The simulated conditional healthy eating index values for selected independent variables in the model are exhibited in Table 6. These estimates are derived from the general label use HEI equations. Clearly, the HEI of label users in each of the selected individual characteristic examined is higher than the HEI of non-label users. Moreover, the HEIs of smokers, those who do not exercise regularly, those who are employed, those who are less educated, those from the non-urban areas, and those who participate in the Food Stamp Program are generally lower than their counterparts.

Conclusions and Implications

Concerns about the effect of diet on health have resulted in the legislation of the NLEA. To assess the effect of consumer label use on diet quality, endogenous switching regression techniques are employed to control for unobservable heterogeneity in the label use decision. The results show that label use, indeed, has a positive effect in improving diet quality. Improvements in diet quality as measured by the HEI range from 9 points (general label use) to almost 14 points (health benefit statements), depending on the type of information used on the label.

The variables that are statistically significant in the HEI equation are different between label user and non-label user. The coefficients of education, body mass index, exercise, and knowledge about the food guide pyramid are statistically significant in the HEI model for label users but not for non-label users. In the HEI model for non-label users, the coefficients for age, household income, the level of urbanization, and some regions are statistically significant. Of interest in the results as well is the negative relationship between diet quality and food stamp participation because it raises questions about the role of the Food Stamps Program in improving the diets of participants.

The key findings in this study are of great importance in terms of public policy because of the tremendous benefits that improved diets can provide the society in general in terms of lives saved and reduction of health care costs. For instance, McNutt estimated that the health care savings from improved and better diets could amount to $3.6 billion to $21 billion. Zarkin et al. also reported that estimated number of discounted life-years that could be gained nationwide from diet-related cancers avoided during the first 20 years of the implementation of the NLEA ranges from about 40,000 to a high of 1.2 million. USDA also estimates that improved dietary patterns could save $43 billion in medical care costs and lost productivity from disability associated with coronary heart disease, cancer, stroke and diabetes each year, and prevent over 119,900 premature deaths among individuals 55-84 years of age, valued at $28 billion per year (Frazao). On the other hand, the FDA estimated that the NLEA would cost the food industry $1.4 billion to $2.3 billion and the government $163 million over the next 20 years. These estimates, however, are contingent upon the presumption that consumers' diets are improved by their use of food labels. This study is the first to econometrically document such an improvement in diet quality from label use.

Footnotes

1. This discussion principally is from Variyam et al (1998).

2. The quality of diet is used as health input instead of diet, because individual's health is more affected by its quality expressed by variety, moderation, and adequacy rather than the diet itself. An individual's diet can be transformed into the measure of quality by way of HEI = QF, where Q is a matrix of fixed weight's representing nutrient levels in each food, number of servings, and a unit vector to count the kinds of foods eaten.

3. The causality from HEI to label use was tested using a simultaneous equation probit model (Amemiya, 1978). The result of the estimation rejected the causal relationship from HEI to label use.

4. Some of these variables (e.g., exercise, smoking, special diet) may be potential endogenous variables in the HEI equation. However, there are no good instruments in the data set that can be used for these variables. Hence, this potential endogenous problem is a potential limitation of this study.

References

Akin, J.S., D.K. Guilkey, B.M. Popkin, and K.M. Smith. “ The Impact of Federal Transfer Program on the Nutrient Intake of Elderly Individuals.” *The Journal of Human Resource 20* (Summer 1985): 385-404.

Amemiya, T. “ The Estimation of Simultaneous Equation Generalized Probit Model.” *Econometrica* 46(September 1978): 1193-1205.

Becker, G.S. “ A Theory of the Allocation of Time.” *Economic Journal* 75(September 1965): 493-517.

Bowman, S.A., M. Lino, S.A. Gerrior, and P.P Basiotis. *The Healthy Eating Index: 1994-96,* U.S. Department of Agriculture, Center for Nutrition Policy and Promotion, July 1998.

Bulter, J.S., J.C. Ohls, and B. Posner. “ The Effect of the Food Stamp Program on the Nutrient Intakes of the Eligible Elderly.” *Journal of Human Resources* 20 (Summer 1985): 405-20.

Bulter, J.S. and J.E. Raymond. “ The Effect of the Food Stamp Program on Nutrient Intakes.” *Economic Inquiry* 34(October 1996): 781-98.

Devaney, B. and T. Fraker. “ The Dietary Impact of the School Breakfast Program.” *American Journal of Agricultural Economics* 71(1989):932-48.

Food Review, *Moving Toward Healthier Diets*, January-April 1996.

Frazao, E. *The American Diet: Health and Economic Consequences*, U.S. Dept. of Agr., Econ. Res. Serv., AIB-711, 1995.

Gould, B. W., and H. C. Lin, “Nutrition Information and Household Dietary Fat Intake.” *Journal of Agricultural and Resource Economics* 19(1994): 453-59.

Grossman, M. “ On the Concept of Health Capital and Demand for Health.” *Journal of Political Economy* 80(1972): 223-55.

Guthrie, J., J. Fox, L. Cleveland, and S. Welsh. “ Who Uses Nutrition Labeling and What Effects Does Label Use Have on Diet Quality?” *Journal of Nutrition Education* 27(1995): 153-72.

Ippolito, P.M. and A.D. Mathios. “New Food Labeling Regulations and the Flow of Nutrition Information to Consumers.” *Journal of Public Policy & Marketing* 12(1993 Fall): 188-205.

Kennedy, E.T., J. Ohls, S. Carlson, and K. Fleming, "The Healthy Eating Index: Design and Applications", *Journal of the American Dietetic Association*, 95(1995):1103-08.

Lee, L. "Unionism and Wage Rates: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables." *International Economic Review* 19(June 1978): 415-433.

Long, S.K. " Do School Nutrition Program Supplement Household Food Expenditure." *The Journal of Human Resources* 26(1991): 654-78.

Maddala, G.S. *Limited Dependent and Qualitative Variables in Econometrics*, New York: Cambridge University Press, 1983.

McNutt, K. "3.6 to $21 Billion Benefit from New Labeling Regulations." *Nutrition Today* 27 (March/April 1992): 39-43.

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

Nayga Jr., R.M., D. Lipinski, and N. Savur, "Consumers' Use of Nutritional Labels while Food Shopping and At Home", *The Journal of Consumer Affairs*, 32,1(1998):106-120.

Poirier, D.J. and P.A. Rudd. " On the Appropriateness on Endogenous Switching." *Journal of Econometrics* 16(1981): 249-256.

Willis, R.J. and S. Rosen. " Education and Self-Selection." *Journal of Political Economy* 87(1979): s7-s36.

Variyam, J.N., J.Blaylock, and D. Smallwood. " A Probit Latent Variable Model of Nutrient Information and Dietary Fiber Intake." *American Journal of Agricultural Economics* 78(August 1996): 629-39.

________________, *USDA's Healthy Eating Index and Nutrition Information*, U.S. Dept. of Agr., Econ. Res. Serv., TB-1866, April 1998.

Zarkin, G.A., N. Dean, J.A. Mauskopf, and R. Williams, "Potential Health Benefits of Nutrition Label Changes." *American Journal of Public Health* 83(May 1993): 717-724.

Table1 Definition of Variables

	Description	Means	Std.Dev
Dependent Variable			
LBUSE	General use of label(yes=1; no=1)	0.8294	0.3762
INGRNT	Use of the list of ingredient(yes=1; no=1)	0.7637	0.4248
HCLAIM	Use of health claims(yes=1; no=1)	0.7415	0.4378
NPANEL	Use of nutrition panel(yes=1; no=1)	0.7536	0.4310
SERVING	Use of serving size(yes=1; no=1)	0.6762	0.4680
HBENEFIT	Use of health benefit statement(yes=1; no=1)	0.6796	0.4667
HEI	Health Eating Index	62.8991	13.7452
Explanatory Variables			
INCOME	household income(10,000 dollars)	3.5082	2.6401
INCMSQ	Square of household income	19.2765	26.1452
AGE	Age of respondent (in years)	50.9071	17.1429
AGESQ	Square of age of respondent	2885.3582	1808.1358
MALE	Respondent is male (1=yes; 0=no)	0.5034	0.5000
B_RACE	Respondent is black (1=yes; 0=no)	0.1164	0.3207
O_RACE	Respondent is other nonwhite race (1=yes; 0=no)	0.0633	0.2435
EMPLOYED	Respondent is employed (1=yes; 0=no)	0.5819	0.4933
CITY	Respondent resides in the central city (1=yes; 0=no)	0.2949	0.4560
NOMETRO	Respondent resides in the non-metropolitan (1=yes; 0=no)	0.2666	0.4422
EDUCATION	Schooling in years	12.6459	3.0932
NE	Respondent resides in the Northeast (1=yes; 0=no)	0.1919	0.3938
WE	Respondent resides in the West (1=yes; 0=no)	0.2019	0.4014
MW	Respondent resides in the Midwest (1=yes; 0=no)	0.2514	0.4339
FSP	Respondent participates in the food stamps program (1=yes; 0=no)	0.0788	0.2695
EXERCISING	Respondent has regular exercise (1=yes; 0=no)	0.4846	0.4998
BMI_SP	Body-mass ratio of respondent	27.9155	11.3339
SMOKING	Respondent is smoking now (1=yes; 0=no)	0.2564	0.4367
DHA	Diet-health awareness (index)	5.7937	1.6159
PYRAMID	Knowledge on Food Pyramid Guideline(index)	2.3604	1.2312
SHOPPER	Respondent is major food shopper (1=yes; 0=no)	0.6949	0.4605
SPECDIET	Respondent has special diet (1=yes; 0=no)	0.1741	0.3792
NHSP	Respondent is non-Hispanic (1=yes; 0=no)	0.9221	0.268

Table 2. Maximum Likelihood Estimates from the First Stage Probit Equations

	General label use	List of Ingredient	Health Claim	Nutritional Panel	Serving Size	Health statement
CONSTANT	-0.7938**	-1.4089**	-0.8837**	-1.1737**	-0.8190**	-1.1108**
	(-3.149)	(-6.182)	(-3.959)	(-5.143)	(-3.832)	(-5.174)
INCOME	0.0789*	0.0543	0.0616*	0.0889**	0.0889**	0.1055**
	(2.411)	(1.837)	(2.166)	(3.023)	(3.263)	(3.897)
INCMSQ	-0.0066*	-0.0039	-0.0045	-0.0069*	-0.0063*	-0.0102**
	(-2.101)	(-1.363)	(-1.640)	(-2.443)	(-2.431)	(-3.977)
HHSIZE	-0.0403*	-0.0377*	-0.0120	-0.2029	-0.0314*	-0.0037
	(-2.301)	(-2.351)	(-0.768)	(-1.807)	(-2.094)	(-0.244)
AGE	0.0176*	0.0259**	0.0131	0.0192**	0.0186**	0.0143*
	(2.312)	(3.718)	(1.910)	(2.748)	(2.827)	(2.169)
AGESQ	-0.0002**	-0.0003**	-0.0001*	-0.0002**	-0.0002*	-0.0002**
	(-3.178)	(-3.816)	(-2.132)	(-3.499)	(-3.932)	(-2.786)
MALE	-0.4667**	-0.3999**	-0.3870**	-0.4235	-0.3223**	-0.2443**
	(-9.191)	(-8.796)	(-8.804)	(-9.326)	(-7.728)	(-5.866)
B_RACE	0.0128	0.0645	0.0412	-0.0593	0.0412	0.0896
	(0.176)	(0.963)	(0.630)	(-0.900)	(0.660)	(1.431)
O_RACE	-0.1476	-0.0112	-0.1596	-0.0881	-0.2183*	0.0298
	(-1.421)	(-0.119)	(-1.765)	(-0.948)	(-2.545)	(0.338)
EMPLOYED	0.0007	0.0373	0.0499	0.0139	-0.0791	-0.0292
	(0.011)	(0.720)	(0.997)	(0.271)	(-1.654)	(-0.614)
CITY	-0.0112	-0.0460	-0.0391	-0.0134	-0.0058	-0.0495
	(-0.196)	(-0.899)	(-0.790)	(-0.263)	(-0.124)	(-1.058)
NOMETRO	-0.1718**	-0.1667**	-0.1977**	-0.2324	-0.1517*	-0.1851
	(-3.140)	(-3.317)	(-4.059)	(-4.676)	(-3.257)	(-3.977)
EDUCATION	0.0682**	0.0664**	0.0486**	0.0561	0.3806**	0.0461**
	(8.211)	(8.603)	(6.509)	(7.330)	(5.288)	(6.406)
NE	0.1988**	0.0809	0.0376	0.1176*	0.0265	0.0943
	(3.065)	(1.390)	(0.668)	(2.016)	(0.493)	(1.756)
WE	0.1844**	0.0816	0.0229	0.0735	-0.0127	0.0081
	(2.813)	(1.388)	(0.404)	(1.256)	(-0.235)	(0.151)
MW	0.1877**	0.1618**	0.1254*	1.1235*	0.0808	0.1162*
	(3.246)	(3.052)	(2.438)	(2.350)	(1.652)	(2.386)
NHSP	-0.2528**	-0.2325**	-0.2511**	-0.0895	-0.1884*	-0.2167**
	(-2.610)	(-2.696)	(-2.963)	(-1.065)	(-2.370)	(-2.691)
SPECDIET	0.2521**	0.2425**	0.2727**	0.3497**	0.2999**	0.2230**
	(3.837)	(4.155)	(4.857)	(5.908)	(5.717)	(4.323)
DHA	0.1519**	0.1587**	0.1393**	0.1509**	0.1186**	0.1360**
	(11.670)	(12.810)	(11.400)	(12.201)	(9.873)	(11.319)
SHOPPER	0.3501**	0.2833**	0.2775**	0.2977**	0.2450**	0.2356**
	(6.687)	(5.828)	(5.838)	(6.127)	(5.346)	(5.120)
Chi-squared	797.394	775.132	662.343	843.561	624.378	569.001
Correct predictions						
Label user	85%	80%	77%	79%	73%	73%
Label non-user	62%	60%	61%	60%	62%	63%
Overall	84%	78%	76%	77%	71%	71%

* indicates significance at 5 % level ; ** indicates significance at 1% level

Table 3. Parameter Estimates of the Second Stage HEI Equations

	General Label Use		List of Ingredients		Health Claims	
	User	Non-user	User	Non-user	User	Non-user
CONSTANT	59.2580**	51.927*	60.997**	54.587**	60.087**	52.487**
	(22.592)	(10.364)	(20.250)	(13.722)	(20.429)	(12.655)
INCOME	0.5894*	1.1210	0.5458	1.2641*	0.3324	1.5122**
	(1.963)	(1.674)	(1.749)	(2.221)	(1.032)	(2.779)
INCMSQ	-0.0212	-0.0980	-0.0168	-0.1176*	-0.0032	-0.1186*
	(-0.760)	(-1.514)	(-0.583)	(-2.125)	(-0.109)	(-2.288)
HHSIZE	-0.1528	-0.1614	-0.0987	-0.2475	-0.1236	-0.4311
	(-0.927)	(-1.891)	(-0.573)	-0.908)	(-0.715)	(-1.646)
AGE	-0.0399	-0.2659	-0.0896	-0.2517*	-0.0332	-0.2556*
	(-0.525)	(-1.891)	(-1.092)	(-1.984)	(-0.410)	(-2.083)
AGESQ	0.0019*	0.0034*	0.0022*	0.0033**	0.0018*	0.0033**
	(2.449)	(2.533)	(2.709)	(2.730)	(2.190)	(2.774)
MALE	0.2982	-0.9156	0.2185	0.4209	0.5158	0.1526
	(0.530)	(-0.709)	(0.387)	(0.406)	(0.850)	(0.142)
B_RACE	-3.1097**	-3.2132*	-3.1653**	-3.5490**	-3.1186**	-3.7879**
	(-4.639)	(-2.287)	(-4.540)	(-2.916)	(-4.376)	(-3.180)
O_RACE	2.8826**	2.7028	2.8560**	1.4192	2.9659**	3.6408*
	(3.116)	(1.272)	(2.972)	(0.806)	(2.952)	(2.184)
EMPLOYED	-1.6043**	-2.1652	-1.7053**	-2.1279*	-1.7943**	-2.1421*
	(-3.158)	(-1.935)	(-3.230)	(-2.242)	(-3.326)	(-2.313)
CITY	0.1353	2.3789*	0.2110	1.9034*	0.0929	2.1697*
	(0.281)	(2.087)	(0.422)	(1.978)	(0.181)	(2.338)
NOMETRO	-1.4254*	-0.3648	-1.1541*	-1.1791	-1.2022*	-0.3023
	(-2.750)	(-0.346)	(-2.127)	(-1.293)	(-2.114)	(-0.331)
EDUCATION	0.3093**	0.1938	0.2792**	0.1671	0.3224**	0.2133
	(3.205)	(1.078)	(2.697)	(1.049)	(3.169)	(1.426)
NE	1.6808**	3.0121*	2.1069**	2.2952*	2.2455**	2.3469*
	(2.934)	(2.299)	(3.576)	(2.139)	(3.732)	(2.253)
WE	1.7406**	1.4781	2.0014**	1.6997	2.3356**	1.5840
	(3.008)	(1.142)	(3.353)	(1.597)	(3.823)	(1.560)
MW	0.6376	0.3526**	0.5544	2.6296**	0.9005	1.7144
	(1.205)	(2.979)	(1.013)	(2.684)	(1.614)	(1.801)
NHSP	-1.8547*	0.4309	-1.3569	-1.7427	-1.6078	-0.1014
	(-2.174)	(0.218)	(-1.521)	(-1.068)	(-1.752)	(-0.062)
FSP	-2.9135**	-1.7525	-3.1303**	-1.1585	-2.9875**	-1.4378
	(-3.535)	(-1.127)	(-3.570)	(-0.883)	(-3.408)	(-1.093)
BMI_SP	-0.0517**	0.0873*	-0.0488**	0.0412	-0.0463**	0.0340
	(-3.026)	(2.335)	(-2.730)	(1.320)	(-2.597)	(1.072)
SPECDIET	2.8795**	3.9822**	2.8460**	2.7969*	2.4532**	3.3238*
	(5.318)	(2.771)	(5.053)	(2.297)	(4.136)	(2.724)
REGEX	1.4659**	0.3537	1.5297**	0.3568	1.4979**	0.6626
	(3.799)	(0.400)	(3.797)	(0.481)	(3.687)	(0.924)
SMOKING	-4.3461**	-4.4899**	-4.3897**	-4.3091**	-4.3913**	-4.1686**
	(-9.464)	(-4.780)	(-9.090)	(-5.426)	(-8.884)	(-5.453)
PYRAMID	0.5986**	0.0843	0.6758**	-0.1663	0.6732**	-0.1019
	(3.681)	(0.248)	(3.977)	(-0.569)	(3.938)	(-0.355)

(Table 3. Continued)

LAMDA	-7.6420**	-3.6971*	-7.0340**	-5.1975**	-8.0317**	-5.5410**
	(-4.300)	(-2.073)	(-4.142)	(-3.104)	(-4.279)	(-2.948)
N	4483	922	4128	1277	4008	1397
R^2	0.153	0.171	0.153	0.154	0.154	0.154

* indicates significance at 5 % level ; ** indicates significance at 1% level

(Table 3 continued)

	Nutritional Panel		Serving Size		Health Benefit Statements	
	User	Non-user	User	Non-user	User	Non-user
CONSTANT	60.3680**	54.532**	62.233**	50.092**	63.548**	48.006**
	(20.526)	(13.553)	(18.718)	(13.607)	(19.403)	(13.079)
INCOME	0.5103	0.8387	0.3365	0.7041	0.2666	0.6955
	(1.601)	(1.525)	(0.949)	(1.427)	(0.771)	(1.373)
INCMSQ	-0.0158	-0.0722	-0.0109	-0.0431	0.0104	-0.0342
	(-0.539)	(-1.352)	(-0.337)	(-0.919)	(0.323)	(-0.711)
HHSIZE	-0.1625	-0.2219	0.0605	-0.3045	-0.2259	-0.2969
	(-0.944)	(-0.842)	(0.313)	(-1.257)	(-1.279)	(-1.213)
AGE	-0.0525	-0.3148*	-0.1202	-0.1876	-0.1036	-0.1569*
	(-0.647)	(-2.576)	(-1.305)	(-1.694)	(-1.203)	(-1.387)
AGESQ	0.0021*	0.0039**	0.0029**	0.0031**	0.0025*	0.0027*
	(2.539)	(3.316)	(3.107)	(2.786)	(2.897)	(2.462)
MALE	0.3288	0.6058	0.8534	-0.0541	-0.3017	0.5063
	(0.565)	(0.573)	(1.292)	(-0.058)	(-0.533)	(0.589)
B_RACE	-2.9073**	-2.7679*	-3.0268**	-3.7296**	-3.1697**	-4.0338*
	(-4.087)	(-2.404)	(-3.951)	(-3.520)	(-4.302)	(-3.680)
O_RACE	1.7953	5.3323**	3.3292**	3.9705**	1.5450	4.0409*
	(1.829)	(3.278)	(2.979)	(2.743)	(1.518)	(2.612)
EMPLOYED	-1.7066**	-1.8043	-1.3020**	-1.1645*	-1.5196**	-1.6941*
	(-3.209)	(-1.959)	(-2.201)	(-1.383)	(-2.691)	(-2.022)
CITY	0.0173	1.9653*	-0.3081	2.0969**	0.1396	1.7760*
	(0.034)	(2.105)	(-0.560)	(2.574)	(0.263)	(2.119)
NOMETRO	-1.1263*	-0.2688	-1.3323*	-0.4035	-1.3959*	-0.1608
	(-1.999)	(-0.298)	(-2.173)	(-0.508)	(-2.336)	(-0.195)
EDUCATION	0.3509**	0.1627	0.3489**	0.2239	0.3219**	0.1970
	(3.467)	(1.104)	(3.207)	(1.699)	(2.966)	(1.404)
NE	1.7983**	2.9350**	2.0888**	2.8259**	1.9441**	2.2628*
	(3.032)	(2.766)	(3.234)	(3.081)	(3.110)	(2.347)
WE	1.9336**	2.2013	2.2101**	2.3384**	2.3604**	1.9359*
	(3.229)	(1.944)	(3.376)	(2.565)	(3.728)	(2.082)
MW	0.7501	2.5515**	0.7218	2.1195**	1.0511	1.1434
	(1.370)	(2.749)	(1.208)	(2.550)	(1.808)	(1.339)
NHSP	-2.5280**	-0.3466	-1.7859	-1.1414	-2.2841*	1.0485
	(-2.805)	(-0.235)	(-1.788)	(-0.103)	(-2.409)	(0.721)
FSP	-2.6403**	-2.0760	-2.2593**	-2.7278*	-2.2903**	-2.7650*
	(-2.971)	(-1.642)	(-2.432)	(-2.337)	(-2.479)	(-2.311)
BMI_SP	-0.0534**	0.0391	-0.0614**	0.0354	-0.0488*	0.0131
	(-3.010)	(1.235)	(-3.230)	(1.319)	(-2.525)	(0.506)
SPECDIET	2.3620**	3.2563**	2.0966**	2.2510*	2.6455**	2.6756**
	(4.013)	(2.578)	(3.112)	(2.086)	(4.316)	(2.589)
REGEX	1.3106**	0.9703	1.5641**	0.7124	1.3907**	0.9569
	(3.238)	(1.343)	(3.657)	(1.140)	(3.270)	(1.503)
SMOKING	-4.0100**	-4.9190**	-4.2167**	-4.1744**	-4.1709**	-4.6449**
	(-8.195)	(-6.410)	(-8.129)	(-6.598)	(-8.157)	(-6.646)
PYRAMID	0.6661**	-0.1719	0.7523**	-0.1269	0.5431**	0.2483
	(3.914)	(-0.599)	(4.171)	(-0.503)	(3.033)	(0.963)

(Table 3 continued)

LAMDA	-6.8812**	-4.9280**	-9.3780**	-6.3828**	-7.0237**	-7.0346**
	(-3.991)	(-2.919)	(-4.378)	(-3.092)	(-3.648)	(-3.643)
N	4073	1322	3655	1750	3673	1732
R^2	0.149	0.163	0.154	0.166	0.147	0.169

* indicates significance at 5 % level ; ** indicates significance at 1% level

Table 4. The Effect of Consumer Label Use on Diet Quality

Types of information on the label	General label use	List of Ingredient	Health claims	Nutrition panel	Serving size	Health benefit statement
Difference in the expected HEI	9.06	11.61	11.59	11.51	12.89	13.96
	Distribution of the difference in the expected HEI(%)					
Less than 0	1.43	0.07	0.00	0.12	0.00	0.00
0 to 5	5.44	0.94	0.70	1.33	0.06	0.03
5 to 10	54.34	24.22	21.21	25.02	6.98	1.55
10 to 20	38.79	74.71	78.09	73.51	92.42	98.39
Over 20	0.00	0.05	0.00	0.02	0.00	0.03

Table 5. Potential Improvements in HEI when Non-Label Users are Induced to Use the Labels

Types of information on the label	General label use	Ingredient	Health claims	Nutrition panel	Serving size	Health benefit statement
A. $E[HEI_1 \mid I=1]$	60.594	60.784	61.169	61.155	61.200	62.137
B. $E[HEI_0 \mid I=0]$	58.930	59.181	59.667	57.584	60.228	60.133
C. Difference (=A - B)	1.603	1.603	1.502	3.571	0.972	2.004

Table 6. Simulated Conditional Healthy Eating Index for Selected Characteristics

	Label user	Non-user
Race		
white	64.02	59.44
black	60.91	56.23
others	66.90	62.14
Region		
Northeast	61.41	54.10
West	61.47	52.57
Midwest	60.37	54.44
South	59.73	51.09
Income		
$15,000	62.93	58.61
$40,000	64.35	61.17
$60,000	65.49	63.22
$80,000	66.63	65.26
$100,000	67.76	67.31
Level of Education (years)		
10	62.88	58.71
12	63.50	59.10
14	64.12	59.48
16	64.73	59.87
18	65.35	60.26
Urbanization		
Central city	65.22	59.30
Outside central city	65.08	58.74
Non-urban area	63.66	55.92
Age (years)		
18	59.81	57.26
40	58.98	51.49
65	58.03	44.93
Employment status		
Employed	63.15	57.86
Not-employed	65.40	61.10
Smoking status		
Smoking	60.48	55.98
Non-smoking	68.14	63.42
Exercising status		
Exercising	64.53	59.14
Non-exercising	62.33	58.58
Special diet		
On special diet	66.13	62.48
Not on special diet	60.91	54.95
Food stamps program		
Participation	61.09	57.36
Non-participation	66.70	60.68

Note: Expected HEIs are evaluated at sample means of exogenous variables for each individual characteristic.