USDA's Healthy Eating Index and Nutrition Information. By Jayachandran N. Variyam, James Blaylock, David Smallwood, Food and Rural Economics Division, Economic Research Service, USDA, and P. Peter Basiotis, Center for Nutrition Policy and Promotion, USDA. Technical Bulletin No. 1866.

Abstract

A comprehensive model is developed to measure the extent that nutrition knowledge and diet-health awareness, among other factors, influence an individual's Healthy Eating Index (HEI), USDA's measure of overall diet quality. This is the first study that rigorously attempts to examine variation in the index across population groups by controlling for personal and household characteristics and nutrition information levels, as well as test for the endogeneity of nutrition information. Results indicate that one's level of nutrition information has an important influence on one's HEI and that nutrition information and the HEI are simultaneously determined. Other factors explaining variations in HEI's across individuals are income and education levels, race, ethnicity, and age. Evidence supports the hypothesis that higher education promotes more healthful food choices through better acquisition and use of health information.

Keywords: Diet quality, Healthy Eating Index, nutrient demand, nutrition knowledge, health inputs, health production.

Acknowledgments

The authors thank Brian Gould and Jug Capps for reviewing the manuscript and Nedra Williams for providing excellent editorial assistance.

Washington, DC 20036-5831 April 1998

Preface

Background on Nutrition and Health

Nutrition is the bridge between agriculture and health. The American diet—high in fat, saturated fat, and sodium, and low in calcium and fiber-containing foods such as fruits, vegetables, and whole grains—is associated with increased risk for several chronic diseases. Diet is a significant factor in the risk of coronary heart disease (CHD), certain types of cancer, and stroke—the three leading causes of death. In fact, poor diets and/or sedentary lifestyles are responsible for 14 percent of all deaths in the United States and at least 20 percent of deaths from CHD and stroke, and 30 percent of cancers may be preventable through diet.

Diet also plays a major role in the development of diabetes and hypertension. At least 30 percent of diabetes could be prevented through diet and/or control of obesity. New research also suggests that increased consumption of fruits and vegetables can lower high blood pressure as effectively as some medications.

Overweight is another major risk factor for coronary heart disease, stroke, some cancers, diabetes, and hypertension and is associated with diet. The prevalence of overweight has increased 10 percentage points in the last 15 years—more than one in three adults are now overweight and there is increased prevalence of overweight among children and teenagers. The increasing overweight problem will increase the prevalence of chronic health problems and at an earlier age.

Diet is also a risk factor for osteoporosis, which accounts for approximately 1.5 million new fractures annually. Improved diet—in particular, calcium intake—might prevent 40-60 percent of osteoporosis-related hip fractures.

Economic Impacts of Poor Diets

The economic impact analysis for the 1993 nutrition labeling regulations estimated that a 1-percent reduction in intake of fat and saturated fat and a 0.1-percent reduction in intake of cholesterol would prevent over 56,000 cases of CHD and cancer, avoid over 18,000 deaths, and save over 117,000 life-years over 20 years. If all Americans restricted their intake of dietary fat by reducing consumption of saturated fat and cholesterol, coronary heart disease and cancer mortality rates would fall by 5-20 percent, depending on age. Overall, 2-percent of adult deaths would be deferred, equivalent to an increase in average life expectancy of 3-4 months, which would accrue mainly among those older than 65 years.

USDA's Economic Research Service (ERS) estimates that improved dietary patterns could save $43 billion in medical care costs and lost productivity resulting from disability associated with CHD, cancer, stroke, and diabetes in the United States each year and prevent over 119,900 premature deaths among individuals 55-74 years of age, valued at $28 billion per year.

As the average age of the U.S. population continues to rise, the adverse effects of current dietary patterns on chronic health conditions are likely to become an increasing problem, with important consequences for health expenditures and quality of life during one's older years. For example, the direct medical costs of osteoporosis, currently $7 billion yearly, will increase six-fold by the year 2000 and twenty-fold by the year 2040.

Why People Eat Different Diets

No two people eat exactly alike, but what accounts for the vast differences in diet quality? Diet quality is really the end result of the foods eaten, preparation techniques used, and other factors influencing the nutrient content of meals. For example, some people eat diets that are rich in fruits and vegetables while others choose a diet high in grains, meats or dairy products. Still others dine out frequently and some people prefer fried foods. In general, four broad categories of factors influence food consumption: consumer incomes; prices of food and other goods; consumers' knowledge of health and nutrition; and tastes and preferences. To change consumption, one of these influences must change. For example, government nutrition education efforts attempt to change knowledge; commercial advertising may try to influence relative prices or change tastes; and preferences. Virtually all prior economic studies of diet quality have focused on income, prices, and consumer personal and household characteristics to explain variation in food demand. Personal and household characteristics are used as proxies for tastes and preferences, which are unobservable, as well as for nutrition knowledge. Of course, knowledge can be measured but few surveys until recently simultaneously collected nutrition knowledge and consumption data.

Consumers' health and nutrition knowledge may differ because of their sociodemographic background. For instance, more educated individuals may acquire more information about the effects of diets on health, and this may induce them to improve the quality of their diets by, for example, consuming more fruits and vegetables. Similarly women may be more aware of diet-health relationships than men, and this increased awareness may be translated to better diets. Conversely, the link between sociodemographics and food and nutrient intake may also reflect consumers' taste differences. For example, Hispanics may choose a different type of diet than non-Hispanics purely due to ethnicity and tradition. Or a person's food tastes may change with age.

Therefore, the influence of sociodemographic variables on food consumption may reflect a combination of an informational effect and a taste effect. Some attributes (such as education) may have a predominantly informational effect, some (such as age) may have a predominantly taste effect, and some (such as race and ethnicity) may have both effects. Moreover, the two effects may reinforce each other, or work in opposing ways.

Healthy Eating Index

USDA is the lead Federal agency for human nutrition and fulfills its health responsibility through support for a healthful and abundant food supply, getting food to people who need it, and promoting healthy dietary choices. USDA introduced the Healthy Eating Index (HEI) in late 1995 to provide an important new tool for meeting our nutrition goals. It made available for the first time a single summary measure to monitor changes in food consumption patterns. This Index measures how well the diets of all Americans conform to the recommendations of the Dietary Guidelines and the Food Guide Pyramid. It has served as a report on the American diet, allowing researchers to analyze how Americans eat, and aids USDA in more effectively promoting proper nutrition. Preliminary analysis indicated that the diets of most Americans need improvement, and some individuals are more likely than others to consume a poor diet.

The HEI and Nutrition Information

This is the first study to examine the influence of socioeconomic characteristics, nutrition knowledge, and awareness of diet-disease relationships on dietary patterns. The report makes a strong case that information and knowledge are the keys that will unlock the door to better diets and in turn better health, longer lives, and children with improved cognitive and learning abilities.

The study found that nutrition information plays a large role in determining the quality of an individual's diet. For two individuals with identical sociodemographic characteristics, the one scoring one unit higher on a nutrition knowledge scale also scored four to five points higher on the HEI scale.

The study shows that the positive effects of income and education on diet quality found in previous studies are really due to the positive effects of these factors on nutrition information. That is, individuals with greater income or education tend to acquire more nutrition information and this, in turn, improves the quality of their diets. If this informational advantage were to disappear, for example through nutrition education targeted to low-income individuals or starting early in childhood, then individuals with greater incomes or education may, in fact, have diets that are no better, or possibly poorer, than individuals with lower incomes or education. This is because individuals with higher incomes or education may have greater preference for less nutritious convenience foods and may eat out more often. Restaurants meals are usually higher in fats, sodium, and cholesterol, and lower in fiber than meals prepared at home.

Informational differences also explain the effects of gender, race, ethnicity, and employment status on diet quality. If everything else is held equal, men and women tend to have the same diet quality. However, on average, women tend to have a higher stock of nutrition information than men, and this shows up in an HEI 5 points higher for women than for men. The study shows that blacks and Hispanics are handicapped by relatively low levels of nutrition information. If their nutrition information levels were brought up to that of whites or non-Hispanics, other factors being equal, then blacks and Hispanics would have significantly higher HEIs than whites and non-Hispanics.

Diet quality tends to improve with age. However, this effect is entirely due to changing tastes since age has no effect on nutrition information once other sociodemographic effects are taken into account. Similarly, smokers are as informed about health and nutrition as non-smokers. Nevertheless, smokers tend to prefer a less healthful diet than non-smokers and, as a result, tend to have a lower HEI than non-smokers. Diet quality deteriorates for the overweight although they are equally informed about health and nutrition as those who are not overweight.

Contents

Page

Introduction 1

Economic Theory, Nutrition Information, and the Healthy Eating Index 3

The Healthy Eating Index 4

Food Group Components of the Healthy Eating Index 4

Other Components of the Healthy Eating Index 4

Information Measures and Explanatory Variables 6

Sample Description 6

Nutrition Information Measures 6

Information and the HEI: Differences Across Demographic and Socioeconomic Groups 7

Explanatory Variables 8

Estimated Models and Empirical Results 11

OLS Models 11

2SLS Models 13

LV Model 16

Conclusions 18

References 19

Appendix 21

USDA's Healthy Eating Index and Nutrition Information

Jayachandran N. Variyam
James Blaylock
David Smallwood
P. Peter Basiotis

Introduction

Deaths from disease associated with dietary excesses—coronary heart disease, some types of cancer, stroke, and noninsulin-dependent diabetes mellitus—account for nearly two-thirds of the deaths each year in the United States (Singh, Kochanek, and MacDorman, 1996). Not only do poor diets extract a heavy toll on individuals, the costs to society are high and continue to rise. The American Heart Association estimates that about 1.5 million heart attacks occur each year, and that coronary heart disease costs Americans an estimated $52 billion in direct health care spending and lost productivity (Frazão, 1995). Associated costs for cancer are even higher, $104 billion. Experts agree that our diets are an important factor in deaths from these diseases but not the only factor (others being genetics, the environment, and aging). A recent study suggests that dietary factors and sedentary activity patterns together account for at least 300,000 U.S. deaths per year (McGinnis and Foege, 1993). The Economic Research Service estimates that illnesses and premature death due to diet-related diseases and conditions costs society about $250 billion per year (Frazão, 1995).

Among the numerous factors affecting dietary choices, nutrition knowledge and beliefs about foods and health—or nutrition information in short—is the most amenable to modification (Thomas, 1991). Presumably, increased knowledge of the nutrient content of foods and heightened awareness of diet-health relationships lead to more healthful food choices. The close association between diet quality and health, the high cost of diet-related illnesses, and the possibility of improving diet quality through better information have been major motivating factors behind the many public and private campaigns that encourage people to eat more healthful diets. Such campaigns include the USDA Food Guide Pyramid, Dietary Guidelines for Americans, and the 5-A-Day campaign.

While these campaigns have contributed to better diets over the years, a considerable gap still remains between actual and healthful diets (Kennedy, Ohls, Carlson, and Fleming, 1995). A major problem faced by nutrition educators and public-health professionals in their efforts to achieve further dietary improvements is a lack of specifics on the use of diet-health information by individuals. For example, to what degree does nutrition information access and use vary across different segments of the population? Likewise, does more nutrition information help individuals to improve their diet quality? Any understanding of factors slowing the adoption of healthful diets requires empirical knowledge of how diet-health information and its effect on dietary choices vary across the population. Such empirical knowledge can be useful for targeting nutrition education programs, for promoting and marketing foods, and for forecasting food consumption trends (Connor, 1994).

Information plays a key role in the economic theories of health behavior (Grossman, 1975; Ippolito and Mathios, 1990; Kenkel, 1991), and several recent empirical studies have verified that certain forms of nutrition information influence the intake of selected nutrients and foods, such as fat, cholesterol, fiber, eggs, pork, poultry, and fish (Brown and Schrader, 1990; Capps and Schmitz, 1991; Carlson and Gould, 1994; Chern, Loehman, and Yen, 1995; Gould and Lin, 1994; Guthrie and Fulton, 1995; Putler and Frazão, 1994; Variyam, Blaylock, and Smallwood, 1996, 1997; Yen, Jensen, and Wang, 1996). However, one drawback of these studies is that they attempted to examine the link between nutrition information and the intake of selected foods or nutrients rather than overall diet quality. Focusing on the intake and its determinants for a particular nutrient, such as saturated fat, is useful if there is concern about its over- or underconsumption. However, most dietary experts, as well as the *Dietary Guidelines for Americans*, advocate a diet that contains a variety of nutrients and foods but at recommended levels (U.S. Department of Agriculture and U.S. Department of Health and Human Services, 1995). Consequently, a measure of overall diet quality is desirable for monitoring progress in meeting dietary guidelines as well as for assessing the influence of nutrition information on diet quality.

A second drawback of some previous studies is that they treated nutrition information as an exogenous determinant of intake. Key intake determinants such as income and education, as well as unobserved individual heterogeneity, may influence both intake and nutrition information levels simultaneously. Therefore, treating information as an exogenous determinant will lead to simultaneous equations bias.

USDA's Center for Nutrition Policy and Promotion recently developed an instrument to assess overall diet quality—the Healthy Eating Index (HEI). USDA developed this index to provide a single summary measure of dietary quality in America. Most earlier dietary assessment instruments focused on specific components, such as fat and cholesterol, but few assessed overall diet quality. The HEI combines information on the amount and variety of food in the diet and compliance with specific dietary recommendations for food components that should be consumed in limited quantities (U.S. Department of Agriculture, 1995).

The Index was applied to the 1989-90 USDA Continuing Survey of Food Intake by Individuals (CSFII) and examined for variability across income and demographic groups. These simple cross-tabulations revealed that an individual's HEI improved with higher income and more education, that women tended to score higher than men, and that scores were higher than average for children and older people. However, these findings may change in a statistical analysis of the data whereby the influence of one variable on the HEI can be separated from the effects of other variables. In simple cross-tabulations, one cannot be certain if the observed variability in the HEI across income groups, for example, is due solely to income or whether education is also playing a role since income and education are closely related. The purpose of this report is to estimate the effect of nutrition information on overall diet quality as measured by the HEI. We control for an extensive set of personal and household characteristics that simultaneously influence both nutrition information and the HEI, and we use several model specifications and estimation methods to assess the stability of our results.

Economic Theory, Nutrition Information, and the Healthy Eating Index

Many of the economic analyses of the demand for foods and nutrients are based on the conceptual framework provided by the theory of household production developed by Becker and the characteristics model of consumer demand developed by Lancaster (Behrman and Deolalikar, 1988; Pitt and Rosenzweig, 1985; Senauer and Garcia, 1991). In this framework, households combine various inputs to produce "commodities," such as the health of family members, so as to maximize a joint utility function. The inputs (for example, food and medical care) derive their value by supplying characteristics (nutrients and medical services) necessary for the production of the commodities (health). Subject to the constraints of household technology and resources, utility maximization generates individual and household demand functions for the inputs and characteristics.

Assume that a representative household with M members has a joint utility function, U:

$$(1) \qquad \max_{\mathbf{F},\mathbf{z}} U = U(F, \mathbf{z}, h),\ U' > 0,\ U' < 0,$$

where **F** is a matrix of foods consumed, and **z** and **h** are vectors of nonfoods and health status for each family member. Health and food intakes enter directly into the utility function because good health is valued in itself and because foods are consumed for reasons other than their nutritional value, such as taste. We assume the utility function satisfies certain conditions. First, if the level of food, nonfoods, or health status increases, so does a person's utility level, U'>0. Second, these increases in utility occur at a deceasing rate as any component of the utility function reaches ever higher levels, U"<0.

Given household income and market prices, the preference function 1 is maximized, subject to three sets of constraints. First, the health of each family member is constrained by the health production technology:

$$(2) \qquad h_m = h\left(HEI_m, \mathbf{g}_m \middle| \mathbf{x}_m, u_m\right),\ m = 1, \ldots, M,$$

where HEI_m is the Healthy Eating Index value (as a measure of overall diet quality) of the mth household member, and $\mathbf{g}_m$ is a vector of nonfood health inputs such as exercise and medical services. The efficiency of producing health from HEI_m and $\mathbf{g}_m$ is conditional on $\mathbf{x}_m$, a vector of personal and household characteristics, and u_m , an exogenous health endowment beyond the individual's or household's control.

Second, expenditures are constrained to equal household income:

$$(3) \qquad \mathbf{i}'\left(\mathbf{F}\mathbf{p}_F + \mathbf{z}\mathbf{p}_z\right) = I,$$

where **p** denotes prices, I is household income, and **i** is a unit vector.

Third, the HEI input into the health production function is constrained by the production technology:

$$(4) \qquad HEI_m = \mathbf{Q}\mathbf{f}_m,$$

where **Q** is a matrix of fixed weights representing nutrient levels in each food, number of servings, and a unit vector to count the kinds of foods eaten; and **f** is the vector of food consumed by the mth household member.

Under the assumption that the relevant functions have desirable properties to ensure unique interior solutions, the first-order conditions to maximize equation 1 subject to the three constraints give, among other relations, a member-specific HEI demand equation as a function of prices, income, personal and household characteristics, and u_m.

Introducing diet-health information explicitly into the model reflects its role as a factor mediating part of the influence of $\mathbf{x}_m$ on h_m. For example, consider a key component of $\mathbf{x}_m$: education. More educated people are more efficient producers of health because they are more informed about the true effects of inputs on health. That is, they have higher allocative efficiency—the ability to select a better input mix (Grossman and Kaestner, 1995). Education, therefore, affects health through information. Other personal characteristics that influence an individual's acquisition and use of information, such as income, play a similar role in the production of health.[1]

Making the role of information explicit, the reduced-form HEI demand function for the mth household member may be written as:

$$(5) \qquad HEI_m = f\left(\mathbf{p}, I \middle| \mathbf{x}_m, \mathbf{k}_m, u_m\right),$$

where **p** is a vector of prices, I is the household income, and $\mathbf{k}_m$ is a vector of nutrition information variables.

[1]Personal characteristics affect health production through productive efficiency (that is, amount of health output from given amounts of inputs) and through tastes related to ethnic and cultural factors (Grossman and Kaestner, 1995).

The Healthy Eating Index

The Healthy Eating Index measures how well people's diets conform to recommended healthful eating patterns. The index provides a picture of foods people are eating, the amount of variety in the diet, and compliance with specific *Dietary Guidelines* recommendations. A score on the index represents the sum of 10 different dietary components. Each component has a possible range of 0 to 10. The maximum overall score is 100 points. The following 10 dietary components are included in the index based on different aspects of a healthful diet:

- Components 1-5 measure the extent to which a person's diet conforms to the *Food Guide Pyramid* serving recommendations for the grain, vegetable, fruit, milk, and meat groups.
- Component 6 measures total fat consumption as a percentage of total food energy intake.
- Component 7 measures saturated fat consumption as a percentage of total food energy intake.
- Component 8 measures total cholesterol intake.
- Component 9 measures total sodium intake.
- Component 10 reflects the amount of variety in a person's diet over a 3-day period.

Food Group Components of the Healthy Eating Index

The HEI reflects dietary intake in relation to the five major food groups (grain, meat, milk, vegetable, and fruit) in the *Food Guide Pyramid* (U.S. Department of Agriculture, 1992). The number of recommended servings depends on an individual's caloric requirements (table 1). For example, recommended servings of vegetables range between three for a 1,600-calorie diet to five for a 2,800-calorie diet.

For each of the five food group components of the index, individuals who consumed the recommended number of servings received a maximum score of 10. A score of zero was assigned for any food group where no items from that food group were eaten. Scores between zero and 10 were calculated in proportion to the number of servings consumed. For example, if the recommended number of servings was 8 and an individual consumed 4 servings, the component score for the individual is 5 points (one-half of 10).

USDA, in developing the index, interpolated serving recommendations from the *Food Guide Pyramid* for individuals with food energy requirements different from the 1,600-, 2,200-, and 2,800-calorie levels in the *Guide*. For example, food energy requirements for children between the ages of 1 and 3 are less than 1,600 calories. The recommended number of servings was retained at the minimum (for example, two for fruits), but the serving size was scaled down to be proportionate with their energy requirements. The exception was for the milk group, where adult serving sizes were retained. This approach is consistent with the guidance provided in the *Food Guide Pyramid*. Conversely, the *Food Guide Pyramid* provides no guidance for adjusting serving sizes for adult males between the ages of 15 and 50 whose energy requirements are greater than 2,800 calories. USDA decided that, in lieu of increasing serving sizes for this age group, food portions would be capped at the maximum number of recommended servings for a food group, such as 11 for grains.

Recommended servings from the five food groups by gender and age of an individual are presented in table 2. For example, the recommended daily servings of fruits range from two for children ages 1-3 to four for males age 15-50.

Table 1—Recommended numbers of servings per day at food energy levels discussed in the Food Guide Pyramid Bulletin

Calories	Grains	Vegetables	Fruits	Milk	Meat
		Number of servings			
1,600	6	3	2	2	2.0
2,200	9	4	3	2	2.4
2,800	11	5	4	2	2.8

Source: U.S. Department of Agriculture, 1992.

Other Components of the Healthy Eating Index

The five other components of the Healthy Eating Index are related to the consumption of fat, saturated fat, cholesterol, and sodium; and variety of diet.

Fat and Saturated Fat: The index scores for fat and saturated fat are related to their consumption in proportion to total food energy. Fat intakes less than or equal to 30 percent of total calories are given a score of 10. The score declines to zero when the proportion of fat to total calories is 45 percent or more. Intakes between 30 and 45 percent were scored proportionately. Saturated fat intakes of less than 10 percent of total calories received a score of 10, while zero points were given for saturated fat intakes of 15 percent or more of calories.

Cholesterol and Sodium: Scores for cholesterol and sodium were each given based on milligrams consumed in the diet. A score of 10 was given for cholesterol intakes less than or equal to 300 milligrams daily.

Table 2—Recommended number of servings per day for age/gender categories

Age/gender category	Kilocalories	Grains	Vegetables	Fruits	Milk	Meat
Children 1-3	1,300	6.0[a]	3.0[a]	2.0[a]	2.0[a]	2.0[a]
*	1,600	6.0	3.0	2.0	2.0	2.0
Children 4-6	1,800	7.0	3.3	2.3	2.0	2.1
Females 51+	1,900	7.4	3.5	2.5	2.0	2.2
Children 7-10	2,000	7.8	3.7	2.7	2.0	2.3
Females 11-50	2,200	9.0	4.0	3.0	2.0	2.4
Males 51+	2,300	9.1	4.2	3.2	2.0	2.5
Males 11-14	2,500	9.9	4.5	3.5	3.0	2.6
*	2,800	11.0	5.0	4.0	2.0	2.8
Males 19-50	2,900	11.0	5.0	4.0	2.0[b]	2.8
Males 15-18	3,000	11.0	5.0	4.0	2.0	2.8

[a]Portion sizes are reduced for children age 1-3. [b]Is 3 servings for persons age 11 to 24.
* RDA levels included in the Food Guide Pyramid.
Source: U.S. Department of Agriculture, 1992.

Zero points were given for intakes at or over 450 milligrams. Intermediate scores were given for intakes between the two limits. For sodium, a score of 10 was earned for intakes less than or equal to 2,400 milligrams. A zero score was given for sodium intakes at 4,800 milligrams or higher.

Variety: Dietary variety was assessed by totaling the number of "different" foods eaten by an individual in amounts sufficient to contribute at least one-half of a serving in a particular food group. Food mixtures were broken into their component ingredients and assigned to relevant food groups. Similar types of foods were grouped together and counted only once in measuring the score for variety. A maximum score of 10 was awarded if 16 or more different food items were consumed over a 3-day period. A score of zero was given if six or fewer food items were consumed. Intermediate scores were awarded proportionately for consumption between the cutoffs.

Information Measures and Explanatory Variables

Sample Description

Data from USDA's 1989-90 Continuing Survey of Food Intakes by Individuals (CSFII) and the companion Diet and Health Knowledge Survey (DHKS) provided the basis for this report. The 1989-90 CSFII/DHKS surveys were conducted by USDA's Human Nutrition Information Service (HNIS). Two independent samples of households—the "basic" or all-income sample and low-income sample—were selected using a multistage, stratified selection procedure targeted at private households in the 48 contiguous States. In the 1989-90 surveys, 15,801 housing units were selected, which, after screening, resulted in 5,554 eligible households, of which 4,406 (79.3 percent) participated.

The CSFII survey collects information on what, when, and where Americans eat and how much they eat. Each CSFII participant was asked to provide 3 consecutive days of dietary data. The first day's data were collected in an in-home interview using a 1-day dietary recall. The second and third days' data were collected using a self-administered 2-day dietary record. Social, economic, and demographic characteristics of survey participants are also included in the CSFII. There were 11,552 individuals living in the 4,406 participating households; 7,816 (67.7 percent) completed the 3-day record.

In the CSFII survey, each food item eaten was recorded using a coding system that contains about 6,700 food codes. USDA's Agricultural Research Service (ARS) maintains a database with the nutrient composition for each food code. The amount of nutrients in each food was calculated by multiplying the amount of food eaten by its nutritive value.

A DHKS respondent, usually the household's main meal planner, was contacted by telephone about 6 weeks after collection of the dietary data and asked questions about knowledge of and attitudes toward diet, health, and food safety issues. Among the 4,406 participating households, 3,805 (86.4 percent) completed the DHKS.

The surveys have been used to describe food consumption behavior and to assess the nutritional content of diets. Results from the surveys have major implications for policies relating to food production and marketing, food safety, food assistance, and nutrition education. The surveys are a major component of the National Nutrition Monitoring and Related Research Program, a set of related Federal activities intended to provide regular information on the nutritional status of the U.S. population.

Our analysis is restricted to the main meal planner/preparer of the sample households since diet-health knowledge of other household members was not collected. After eliminating cases with missing values, our final sample consisted of 2,442 observations out of 3,805 with complete 3-day intake data.

Nutrition Information Measures

We used responses to two sets of questions in the DHKS to develop measures of meal planners' nutrition information. The first measure represents the "nutrient content knowledge" (NCK) of meal planners. Respondents were asked to choose the correct answer from each of a series of binary-choice questions about sources and occurrence of various nutrients in common food items. We used 21 questions to construct a measure of a respondent's nutrient content knowledge (table 3). The NCK measure represents the number of correct answers given by a respondent. Therefore, the minimum score is zero and the maximum is 21. On average, respondents answered about 15 questions correctly.

The NCK questions probed a respondent's knowledge of the fiber, cholesterol, and fat content of foods. For example, respondents were asked to identify which of two foods has the higher fiber content: fruit or meat, cornflakes or oatmeal, popcorn or pretzels. They were also asked to identify which foods contain more cholesterol: liver or T-bone steak, butter or margarine, skim or whole milk. Other questions probed knowledge about different kinds of fat, the types of foods that contain cholesterol, and the relationship between fat and cholesterol.

Respondents identified the correct answer to some of the comparisons more easily than others. Over 90 percent correctly identified whole-wheat bread as containing more fiber than white bread, but only 56 percent knew that kidney beans contained more fiber than lettuce. Likewise, virtually everyone, 95 percent, knew that skim milk has less cholesterol than whole milk, but only 52 percent correctly identified liver as containing more cholesterol than a T-bone steak. The same held true for the questions concerning fat content. Most knew that ice cream contained more fat than sherbet and that fried chicken was higher in fat than roasted chicken, but far fewer knew that a porterhouse steak contained more fat than a round steak. When respondents were asked what kind of fat (saturated or polyunsaturated) is more likely to be a liquid rather than a solid, only 30 percent could identify polyunsaturated fat as the correct answer. Less than 40 percent of the respondents knew that cholesterol is found only in animal products.

Table 3—Nutrient content knowledge questions and percent responses

Question	Correct	Incorrect
	Percent	
Which has more fiber?		
Fruit or meat	77.7	22.3
Cornflakes or oatmeal	79.5	20.5
Whole-wheat bread or white bread	91.8	8.2
Orange juice or an apple	74.0	26.0
Kidney beans or lettuce	56.3	43.7
Popcorn or pretzels	73.6	26.4
Which has more cholesterol?		
Liver or T-bone steak	52.3	47.7
Butter or margarine	87.2	12.8
Egg whites or yolks	84.6	15.4
Skim milk or whole milk	95.0	5.0
Which has more fat?		
Regular hamburger or ground round	87.8	12.2
Loin pork chops or spare ribs	72.0	28.0
Hot dogs or ham	61.3	38.7
Peanuts or popcorn	90.5	9.5
Yogurt or sour cream	85.9	14.1
Porterhouse steak or round steak	58.8	41.2
Ice cream or sherbet	95.0	5.0
Roast chicken leg or fried chicken leg	94.6	5.4
Which kind of fat (saturated, polyunsaturated) is more likely to be a liquid rather than a solid? Or are they equally likely to be liquids?	29.6	70.4
Is cholesterol found in vegetables and vegetable oils, animal products, or all foods containing fat or oil?	38.7	61.3
If a food is labeled cholesterol-free, is it also low in saturated fat, high in saturated fat, or either?	55.6	44.4

Source: 1989-90 Diet Health Knowledge Survey.

Table 4 lists the questions used to construct a variable measuring the meal planners' awareness of diet-health problems. These questions take the general form: Have you heard about any health problems that might be related to how much of a particular nutrient a person eats? There are eight such questions. About 85 percent of the respondents indicated that they had heard of health problems associated with salt, but less than 50 percent said the same for fiber and iron. We constructed a "diet-health awareness" variable (DHA) by adding together the positive responses for each of the eight questions. Thus, the variable has a lower limit of zero (respondent had heard of no problems associated with any of the nutrients) and an upper limit of eight (respondent had heard of health problems associated with each of the eight nutrients). This is similar to Kenkel's (1991) measure of health knowledge in his study of smoking, alcohol use, and exercise.

Information and the HEI: Differences Across Demographic and Socioeconomic Groups

Before we systematically isolate and analyze the impacts of individual socioeconomic and demographic characteristics on the Healthy Eating Index and the knowledge and awareness variables, it is instructive to examine the average index, knowledge, and awareness scores within different population groups. Examining the average scores within groups gives an indication of which influences are likely to be important in building statistical models. However, the following descriptive analysis should be interpreted cautiously since some characteristics are highly correlated. For example, higher HEI scores that are associated with higher education levels may be partially caused by higher income, since education and income levels are positively correlated.

USDA has developed a grading scale to rate overall diet quality as measured by the HEI. The scale rates

Table 4—Diet-health awareness questions and percent responses

Question	Yes	No
	Percent	
Have you heard about any health problems that might be related to how much:		
Fat a person eats?	71.3	28.7
Saturated fat a person eats?	58.6	41.4
Fiber a person eats?	48.8	51.2
Salt a person eats?	84.7	15.3
Calcium a person eats?	59.3	40.7
Cholesterol a person eats?	81.7	18.3
Sugar a person eats?	79.6	20.4
Iron a person eats?	47.5	52.5

Source: 1989-90 Diet Health Knowledge Survey.

Table 5—Nutrition information and the healthy eating index across selected sociodemographic groups

	Nutrient content knowledge (NCK)	Diet-health awareness (DHA)	Healthy eating index (HEI)
HEI:			
Less than 51	14.41	4.71	44.99
51-80	15.45	5.33	64.79
Greater than 80	16.55	6.04	88.09
Age:			
Less than 30	15.09	4.84	59.28
31-49	15.67	5.64	61.51
50-69	15.68	5.44	67.17
Over 69	14.74	4.84	69.33
Gender:			
Male	14.75	4.95	60.59
Female	15.56	5.39	64.79
Race:			
White	15.74	5.49	64.78
Black	13.76	4.41	59.66
Other	14.12	4.47	63.56
Ethnic origin:			
Non-Hispanic	15.55	5.37	64.04
Hispanic	13.56	4.60	64.11
Income per capita:			
Less than $3,801	14.28	4.72	59.52
$3,801-5,400	14.69	4.74	63.47
$5,401-10,200	15.30	5.18	64.52
$10,201 or above	16.57	6.06	66.83
Education:			
Less than high school	14.10	4.53	62.57
High school	15.56	5.20	62.97
More than high school	16.56	6.21	66.67
Vegetarian:			
Vegetarian	15.61	5.18	67.21
Nonvegetarian	15.41	5.32	63.95
Smoking:			
Smoker	15.04	4.93	58.63
Nonsmoker	15.55	5.45	65.98

Source: 1989-90 CSFII, DHKS.

index scores of greater than 80 as "good," scores of 51-80 as "needs improvement," and scores less than 51 as "poor." Table 5 reports results of tabulating the HEI grades and the nutrition information variables against key socioeconomic groups. We found that higher scores are clearly associated with increased knowledge about the nutrient content of foods as well as about diet-health awareness. For example, individuals with scores rated good answered, on average, two more questions correctly about nutrient content than people with a poor HEI score.

Age appears to be strongly associated with higher HEI scores. On average, people over age 69 scored 10 points higher than individuals under 30 years old. However, there was not a clear association between age and nutrient knowledge or diet-health awareness. On the other hand, women had higher HEI scores than men and higher nutrient knowledge and diet-health awareness levels.

Race and ethnicity appear to have some influence on HEI scores as well as on nutrient knowledge and awareness. Whites had higher HEI scores on average than Blacks, but Hispanics and non-Hispanic scores were virtually identical. Non-Hispanics' nutrient content knowledge and diet-health awareness scores were higher than Hispanics'.

We found that higher education and incomes were systemically related to higher knowledge of the nutrient content of foods, more awareness of diet-health problems, and higher HEI scores. Vegetarians have higher index scores, but their knowledge and awareness levels appear to be about the same as nonvegetarians. Smokers had lower index scores than nonsmokers and slightly lower knowledge and awareness scores.

Explanatory Variables

Table 6 lists the explanatory variables hypothesized to affect nutrition information and/or HEI. The variables fall into three broad categories: household characteris-

Table 6—Description of variables

Variable description	Name	Mean
Dependent variables:		
Healthy eating index	HEI	64.1
Nutrient content knowledge	NCK	15.4
Diet-health awareness	DHA	5.3
Independent variables:		
Household characteristics		
Annual income before taxes	Income	22.8
Household size	Household size	2.6
Children present (less than 20 years old)	Children	37.9
Participate in Women, Infants, and Children Program	WIC	4.2
Participate in Food Stamp Program	FSP	13.0
Head of household status:		
Only female head	Female head	34.6
Only male head	Male head	11.6
Both male and female heads (omitted)	—	53.8
Region:		
Midwest	Midwest	24.2
South	South	37.7
West	West	19.3
Northeast (omitted)	—	18.8
Urbanization:		
Suburb	Suburban	42.8
Nonmetro	Nonmetro	26.8
City (omitted)	—	30.4
Personal characteristics		
Education:		
Completed high school	High school	35.6
Attended but did not complete college	Some college	18.4
Completed college	College	7.4
Completed post-graduate degree	Post-graduate	6.8
Less than high school (omitted)	—	31.8
Age (years)	Age	48.9
Sex-female	Female	82.3
Race:		
Black	Black	13.6
Other[1]	Other	1.4
White (omitted)	—	85.0
Ethnic origin-Hispanic	Hispanic	6.8
Employment status:		
Not employed	Not employed	54.4
Employed part-time	Part-employed	14.0
Employed full-time (omitted)	—	31.6
Smoke cigarettes now	Smoker	26.3
Former, not current, smoker	Quit smoking	17.9
Vegetarian	Vegetarian	3.2
Body Mass Index (BMI)	BMI	25.8
Watch more than 5 hours TV per day	TV5	20.8
Received diet advice from physician or dietitian	Diet advice	10.2
Compare nutrients when shopping:		
Always	Nutri-comp1	14.4
Sometimes	Nutri-comp2	42.3
Rarely/never (omitted)	—	43.3
Survey-related:		
Year of CSFII-DHKS		
1990	1990	49.1
1989 (omitted)	—	50.9
Amount of food eaten (day 1):		
Less than usual	LTU1	17.1
More than usual	MTU1	6.8
Usual (omitted)	—	76.1
Amount of food eaten (day 2):		
Less than usual	LTU2	13.5
More than usual	MTU2	3.6
Usual (omitted)	—	82.9
Amount of food eaten (day 3):		
Less than usual	LTU3	13.0
More than usual	MTU3	4.0
Usual (omitted)	—	83.0

Note: Household income, household size, age, and BMI are continuous variables. The standard deviations are 21.4, 1.6, 18.3, and 5.5, respectively. All other independent variables are dummy variables.

[1]Asian/Pacific Islander, Aleut, Eskimo, or American Indian.

tics, personal characteristics, and survey-related controls. Most of the household and sociodemographic variables such as income, household size, age, sex, race, and schooling have been used in previous nutrient intake studies (Behrman and Deolalikar, 1988; Gould and Lin, 1994; Morgan, 1986). The regional, urbanization, and survey-year dummy variables are expected to capture any cross-sectional price variation across households.

Income is represented by gross household income before taxes for the year before the survey. Higher income may provide increased access to dietary information and thus indirectly increase diet quality (Ippolito and Mathios, 1990). On the other hand, intake of meat products and less nutritious convenience foods may rise as income increases, causing a negative direct effect on diet quality. Which of these effects will dominate is uncertain and needs to be determined empirically. Household size, presence of children in the household, household head status, and the employment status of the meal planner all likely influence the household's allocation of resources as well as the time spent in shopping for and preparing food (Gawn et al., 1993; Horton and Campbell, 1991). Hence, these variables are likely to influence both nutrition information and diet quality. The Women, Infants, and Children (WIC) and Food Stamp Program (FSP) variables are included to capture the nutrition effects of program participation (Basiotis, Hirschman, and Kennedy, 1996; Butler and Raymond, 1996).

Education is predicted to have a positive, indirect effect on diet quality by increasing the allocative efficiency of health production (Grossman, 1975; Grossman and Kaestner, 1995; Ippolito and Mathios, 1990; Kenkel, 1991). However, as in the case of income, the direct effect of schooling on intake due to variations in tastes is difficult to predict and remains to be empirically determined.

The role that women have often played in food preparation and shopping leads us to expect they have a higher stock of nutrition information than men. The race, ethnicity, and age variables are expected to capture variations in information, food preferences, and consumption induced by cultural backgrounds, cohort effects, and dietary habits.

Smoking is related to health risk perception and smokers may value health less than nonsmokers (McPhillips, Eaton, and Gans, 1994; Viscusi, 1990). This leads us to expect a negative direct effect of smoking on diet quality. All else equal, vegetarians are expected to have higher HEI's since HEI components emphasize fruit and vegetable consumption. Body mass index (BMI) is a ratio of body weight (in kilograms) divided by the square of height (in meters). Individuals with higher BMI's may receive more of their calories from foods high in fats and fewer calories from foods rich in complex carbohydrates (Dattilo, 1992), and hence we expect BMI to have a negative direct effect on HEI.

Meal planners' use of nutrition information sources is captured by whether the person watches 5 or more hours of television each day (TV5), whether he or she receives dietary advice from a physician or a dietitian (diet advice), and whether the person compares nutrients while shopping (nutri-comp1, nutri-comp2). While some amount of television watching may help a person gain information, an excessive amount (defined as 5 or more hours per day) is likely to hinder information gathering by curtailing alternative activities such as reading (Carlson and Gould, 1994). Both receiving dietary advice and the practice of comparing nutrients while shopping are expected to be positively correlated with a respondent's nutrition information level (Gould and Lin, 1994; Kenkel, 1991; Moorman and Matulich, 1993). Finally, some variation in the intake data is likely to depend on whether the person reported each day's food intake to be less than usual or more than usual. A set of six binary variables is used to control for these survey-related effects. The omitted categories are those reporting "usual" intake.

Estimated Models and Empirical Results

OLS Models

Table 7 reports the main empirical results. Column 1 presents results from a linear model for HEI:

$$\text{(6)} \qquad \text{HEI} = \alpha_0 + \sum_{p=1}^{P} \alpha_p X_p + u,$$

where X_p represents the explanatory variables. The model does not include information variables and was estimated by ordinary least squares (OLS). The standard errors of the coefficient estimates were corrected for heteroskedasticity using White's procedure (Greene, 1995, p. 261). This model is similar to that of Basiotis, Hirschman, and Kennedy (1996). It gives the net effects of explanatory variables on the HEI, without reference to the theoretical framework underlying the inclusion of nutrition information variables, $\mathbf{k}_m$, in the reduced-form equation 5. These estimates provide a useful benchmark against which we can compare our other estimated models.

The sample and independent variables used by Basiotis, Hirschman, and Kennedy (1996) are different from ours. Their sample included all CSFII respondents, while our sample is limited to main meal planners responding to both CSFII and DHKS. This means, for example, that their sample includes children whereas ours does not. The results in column 1, therefore, are different from theirs. To capture nonlinear effects of income and age on diet quality, Basiotis, Hirschman, and Kennedy (1996) include squared and cubic terms of these variables. We experimented with different specifications for income, age, and BMI, and found that a log specification for income and linear specification for age and BMI were appropriate.[2]

According to the linear OLS model, and based on significance at the 10-percent level or lower, the profile of a meal planner with a high HEI value is an older, White, nonsmoking, highly educated female with high household income, low BMI, not employed or employed part-time, and residing in the Northeast. Household variables, except income and geographic region, have no significant effect on HEI. Whether the individual is vegetarian and whether he or she quit smoking also has no influence on HEI. As noted earlier, previous studies have suggested interpreting the effects of education and income in equation 1 as reflecting informational differences (Ippolito and Mathios, 1990; Kushi and others, 1988). Higher income may promote greater access to information and higher education may increase information processing efficiency. Therefore, both are hypothesized to have a positive effect on health behavior, such as diet quality.

In columns 2 and 3 of table 7, we report OLS estimates for models that include nutrition information as an explanatory variable:

$$\text{(7)} \qquad \text{HEI} = \alpha_0 + \sum_{p=1}^{P} \alpha_p X_p + \beta \text{INFO} + u,$$

where INFO is either nutrient content knowledge (NCK) or diet-health awareness (DHA). Estimated coefficients for both of these regressors are positive and highly significant, confirming that a higher information level is related to better diet quality as measured by the HEI.[3] Although the coefficients of the information variables are estimated at high significance levels, the estimated elasticities are rather small; 0.155 for NCK and 0.049 for DHA. Holding other explanatory variables constant, a 1-percent rise in the nutrient content knowledge score results in a 0.155-percent increase in the meal planner's HEI, and a 1-percent rise in the diet-health awareness score results in a 0.049-percent increase in the meal planner's HEI. Based on these elasticities, NCK seems to be a better predictor of the HEI than does DHA. This is likely because NCK is measured from a much larger set of components (21) and has a much smaller coefficient of variation (0.19) than DHA, which has only eight components and a coefficient of variation of 0.43 (tables 1 and 2).

Compared with model 1, the estimated effects of most other explanatory variables change when information measures are added in models 2 and 3 of table 7. The effects of income, education, and gender decline, while those of race (Black) and Hispanic ethnicity rise. The effects of age, BMI, and smoking, however, remain

[2]The dependent variable HEI takes on values ranging from 0 to 100. Given this fixed range, a question may arise as to whether a transformation of HEI is necessary to ensure that the predicted values are also bound within this range. This is not strictly applicable here because HEI is not a proportion, a probability, or a percentage for which such a transformation is typical, but rather a score constructed to be between 0 and 100. We did, however, estimate models where the dependent variable was expressed as log(HEI/100-HEI) and obtained quite similar inferences as reported here. See Putler and Frazão (1994) for use of such a transformation.

[3]We estimated an OLS model that included both NCK and DHA. The estimated coefficients (t-values) were 0.534 (5.52) and 0.372 (3.22) respectively. The decrease in the DHA coefficient in this case is due to the relatively high correlation between NCK and DHA; the correlation coefficient is 0.462. To retain compatibility with subsequent models, we do not report these results in table 7; see also footnote 4.

Table 7—Determinants of diet quality (absolute t-values in parentheses)

Explanatory variable	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	LV (6)
Nutrient content knowledge	—	0.643*	—	4.131*	—	—
(NCK)		(7.00)		(5.21)		
Diet-health awareness	—	—	.596*	—	3.803*	—
(DHA)			(5.34)		(3.34)	
Nutrition information	—	—	—	—	—	5.053*
(INFO)						(5.41)
Log income	1.465*	1.099*	1.264*	-.888	.177	-1.392***
	(3.68)	(2.78)	(3.19)	(1.38)	(.30)	(1.91)
High school	1.128***	.672	.951	-1.802***	-.001	-2.151**
	(1.83)	(1.09)	(1.54)	(1.87)	(.00)	(2.12)
Some college	3.891*	3.129*	3.403*	-1.005	.771	-2.312
	(5.17)	(4.13)	(4.52)	(.76)	(.60)	(1.52)
College	5.221*	4.14*	4.386*	-1.714	-.117	-4.102***
	(4.99)	(4.02)	(4.22)	(.95)	(.06)	(1.85)
Postgraduate	5.529*	4.183*	4.466*	-3.121	-1.248	-6.173**
	(5.12)	(3.86)	(4.09)	(1.52)	(.54)	(2.37)
Age	.185*	.183*	.185*	.176*	.187*	.178 *
	(10.23)	(10.26)	(10.32)	(7.52)	(8.67)	(7.31)
Female	4.991*	4.142*	4.73*	-.466	3.320*	-.862
	(4.96)	(4.13)	(4.74)	(.27)	(2.69)	(.50)
Black	-1.613**	-.932	-1.320***	2.765**	.260	3.395**
	(2.14)	(1.22)	(1.73)	(2.06)	(.23)	(2.49)
Other	-.807	.082	-.180	4.905	3.196	6.643**
	(.35)	(.04)	(.08)	(1.58)	(1.41)	(2.22)
Hispanic	1.890**	2.894*	2.225**	8.339*	4.026*	8.901*
	(2.00)	(2.99)	(2.37)	(4.57)	(3.13)	(4.86)
Part-employed	1.737**	1.626**	1.484**	1.026	.122	.154
	(2.35)	(2.21)	(2.02)	(1.10)	(.12)	(.15)
Not employed	1.185***	1.207**	1.092***	1.326***	.586	.959
	(1.92)	(1.98)	(1.78)	(1.70)	(.79)	(1.15)
Children	-1.095	-.996	-.988	-.462	-.416	-.131
	(1.41)	(1.29)	(1.27)	(.46)	(.43)	(.12)
Household size	-.016	.007	-.029	.134	-.102	.067
	(.07)	(.03)	(.13)	(.44)	(.37)	(.20)

See notes at end of table.

—Continued

stable. This indicates that the sociodemographic variables were capturing some of the effects of information in model 1. However, equation 7 does not address the question whether *all* or only *some* of the effects of education, income, and other sociodemographic variables are through information. As noted earlier, the effect, especially of income, is often unclear; higher income may promote increased intake of fat-rich foods such as meats and thus influence HEI negatively— the direct effect. At the same time, higher income may provide greater access to information and thus indirectly influence HEI positively. The relative influence of these different effects can be seen by explicitly modeling the relationship between sociodemographics and nutrition information:

$$\text{INFO} = \gamma_0 + \sum_{q=1}^{Q} \gamma_q X_q + v, \tag{8}$$

where INFO is either NCK or DHA, γ's are unknown parameters, and v is a random error term.

In estimating the parameters of equations 7 and 8, we take account of the correlation between u and v due to unobserved heterogeneity—that is, correlation due to factors that are not included among the regressors, such as unobserved preferences that affect both HEI and information. If corr(u,v) $\neq$ 0, the OLS estimator of parameters in equation 7 suffers from simultaneous equations or endogeneity bias and is inconsistent. We tested whether NCK and DHA in models 2 and 3 are endogenous by applying a test of simultaneity. The test involves adding OLS residuals from equation 8 as an explanatory variable in equation 7 and applying a t-test to the estimated coefficient of the residual (Pindyck and Rubinfeld, 1991, pp. 303-304). The t-values for NCK and DHA residuals in such a test were -5.69 and

Table 7—Determinants of diet quality (absolute t-values in parentheses)—continued

Explanatory variable	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	LV (6)
Female head	-.004 (.01)	.165 (.25)	.070 (.11)	1.079 (1.24)	.466 (.59)	1.239 (1.33)
Male head	.787 (.64)	.615 (.51)	.971 (.80)	-.316 (.19)	1.963 (1.35)	.493 (.30)
BMI	-.128* (3.00)	-.139* (3.26)	-.130* (3.03)	-.200* (3.62)	-.139* (2.69)	-.199* (3.40)
Smoker	-4.592* (8.12)	-4.492* (8.01)	-4.447* (7.89)	-3.952* (5.45)	-3.667* (5.05)	-3.477* (4.41)
Quit smoking	-.151 (.23)	-.243 (.37)	-.185 (.28)	-.738 (.88)	-.366 (.48)	-.801 (.92)
Vegetarian	1.711 (1.23)	1.508 (1.06)	1.789 (1.28)	.407 (.21)	2.211 (1.32)	.841 (.47)
FSP	-.328 (.39)	-.335 (.40)	-.367 (.44)	-.380 (.34)	-.582 (.56)	-.524 (.47)
WIC	-.181 (.16)	-.022 (.02)	.159 (.14)	.840 (.56)	1.988 (1.23)	2.004 (1.16)
Midwest	-1.473** (2.05)	-1.762** (2.48)	-1.488** (2.08)	-3.332* (3.58)	-1.568*** (1.87)	-3.188* (3.14)
South	-2.094* (3.08)	-1.970** (2.94)	-2.036** (3.01)	-1.297 (1.50)	-1.725** (2.16)	-1.166 (1.28)
West	-1.150 (1.46)	-1.313*** (1.69)	-1.26 (1.60)	-2.016** (2.25)	-1.824** (2.00)	-2.476** (2.38)
Suburban	.172 (.30)	.179 (.32)	.230 (.40)	.212 (.30)	.542 (.81)	.425 (.57)
Nonmetro	.056 (.09)	.054 (.08)	.090 (.14)	.041 (.05)	.271 (.36)	.169 (.20)
1990	.183 (.39)	.009 (.02)	.090 (.20)	-.932 (1.53)	-.407 (.72)	-1.158*** (1.74)
Intercept	40.780* (9.30)	35.810* (8.22)	40.013* (9.19)	8.843 (1.07)	35.885* (6.86)	.000
R^2	.213	.229	.222	.114	.124	.229
Test of schooling restrictions	—	—		1.243 (.29)	.673 (.61)	9.098 (.059)

Note: *, **, and *** indicate coefficient estimates significant at 1-, 5-, and 10-percent levels, respectively, under two-sided t-test. All models also included six dummy variables indicating whether each of the 3-day intakes was less than usual or more than usual; usual is the omitted category; see table 4. The R^2 for 2SLS and LV models are the squared correlations between observed and predicted HEI values

-3.37, respectively. Therefore, the exogeneity of nutrition information is strongly rejected.

2SLS Models

To account for the endogeneity of the information variables, we estimated equation 7 by two-stage least squares (2SLS). The results are reported as models 4 and 5 in table 7. The information variables are regressed on all the determinants of HEI and four indicators of the use of different sources of information: excess television watching, whether respondent received advice from physician or dietitian about dieting, and whether respondent compares (always/sometimes) nutrients in foods while shopping (Gould and Lin, 1994; Kenkel, 1991; Moorman and Matulich, 1993). Estimates from these first-stage regressions are reported in table 8, columns 1 and 2. In each case, the standard errors of the coefficient estimates were corrected for heteroskedasticity using White's procedure (Greene, 1995).

Estimated coefficients for NCK and DHA in table 7, models 4 and 5, are considerably higher than corresponding estimates in models 2 and 3.[4] Both the NCK

[4] We estimated a 2SLS model that included both NCK and DHA. While the coefficient for NCK came out positive and significant in this model, the estimate of the DHA coefficient was insignificant. This is due to the very high correlation between predicted values of NCK and DHA; the estimated correlation was 0.904. Therefore, we do not report the results of this model. We also examined possible nonlinearities in the information effects by estimating the HEI equations with the square of the predicted values of NCK and DHA added as regressors. The results showed that a small nonlinearity may exist at the lower ranges of NCK and DHA, but for the most part the relationships are linear. For example, the average derivative of HEI with respect to NCK is 4.143, which is almost identical to the linear estimate of 4.131 in model 4, table 7.

and DHA estimates are more than six times their OLS estimates. The elasticity of HEI with respect to nutrition information is close to one (0.995) for NCK and one-third (0.316) for DHA. The increase in size of the estimated information effects under 2SLS is an indication that much of the effect of the sociodemographic variables on HEI occurs through nutrition information. This is most clearly illustrated by the estimated effects of income and education in models 4 and 5. Most of these estimates are insignificant. At the same time, the first-stage estimates in table 8, models 1 and 2, show that income and education have highly significant positive effects on NCK and DHA.

In model 1, table 7, the education effects show a clear monotonic pattern, increasing with larger effects for higher education. The first-stage results in table 8 show that this monotonic pattern is generated by the steadily increasing informational effects of higher education. In the last row of table 7, we report F-statistics for excluding the education dummy variables from the HEI equation, with p-values given in parentheses. These restrictions cannot be rejected for both NCK and DHA.

These results suggest that the indirect effects of income and education occurring through information almost completely explain their net effects on HEI. Thus, the role of education and income in determining diet quality, at least as measured by HEI, appears to be wholly information-related. Our findings support the view proposed in the health economics literature that education influences both the choice of health inputs and health by increasing information about these inputs. Increased information leads to greater allocative efficiency, that is, the ability to select and use a better input mix (Grossman and Kaestner, 1995). Our findings also suggest that the favorable effects of schooling on health inputs (in our case, overall diet quality as measured by HEI) persists even at higher levels of schooling, thus supporting Grossman's finding about positive education effects on health.

The estimated effects of gender, race, and ethnicity in the 2SLS models provide additional evidence regarding the informational effects of sociodemographic variables. When NCK is the information variable, the coefficient for the female dummy variable becomes insignificant in the 2SLS model. This contrasts with the significant positive effects in the OLS models 1 through 3. In the 2SLS model, the entire effect of being female on HEI occurs through NCK, as evidenced by the significant positive effect of the female dummy variable on NCK in table 8, column 1. This result implies that, holding all sociodemographic and household characteristics constant, a male meal planner and a female meal planner, both possessing the same level of nutrient content knowledge, do not significantly differ in their HEI's. However, on average, female meal planners possess higher nutrient content knowledge than male meal planners, explaining part of the positive effect of NCK on HEI.

The informational effects of race and ethnicity are equally striking. The relatively small or insignificant effects of these variables in models 1 through 3 change to large positive effects in model 4. Thus, when the endogeneity of information is taken into account, all other things equal, the HEI for Black meal planners is about three points higher than that for White meal planners, meal planners of other races have HEI scores four points higher than White meal planners, and Hispanic meal planners have HEI scores that are eight points higher than non-Hispanic meal planners. However, when nutrition information is allowed to vary, the higher HEI's of these groups decline, disappear, or turn negative as in table 7, model 1. As table 8, column 1 estimates show, all other things equal, Black and other non-White meal planners have significantly lower nutrient content knowledge than White meal planners and the same holds for Hispanic meal planners compared with non-Hispanic meal planners. These results suggest that, while the tastes and preferences of non-White and Hispanic meal planners lead them to choose a more healthful diet, their relative lack of nutrition information reduces their ability to choose a better quality diet.

The 2SLS estimates for gender, race, and ethnicity also highlight the sensitivity of results to the information measures used in the analysis. For example, the female dummy variable is insignificant in the 2SLS model with NCK, but it is positive and significant in the 2SLS model with DHA. For the Black binary variable, the result is the opposite, and the Hispanic effect in model 5 is half of that in model 4 (table 7). The reason for these differences is that these variables are capturing some of the variation that is specific to the information measure. For example, although the gender effect is positive on both NCK and DHA, women have relatively more nutrition information as measured by NCK than as measured by DHA. At the sample means of NCK and DHA, the female coefficient estimates in table 8, columns 1 and 2, suggest that women have 8.3 percent higher NCK scores and 7.8 percent higher DHA scores than men.[5]

[5]The female coefficient estimates for NCK and DHA in table 8 show that, all else equal, women answer 1.28 more NCK questions and 0.413 more DHA questions correctly than men. Therefore, at the sample means of NCK and DHA, 15.4 and 5.3 respectively (table 6), women have 8.3 percent higher NCK score and 7.8 percent higher DHA score than men.

Table 8—Coefficient estimates from first-stage reduced-form information equations (absolute t-values in parentheses)

Exogenous variable	2SLS NCK (1)	2SLS DHA (2)	LV INFO (3)
Log income	.562* (6.86)	.324* (4.53)	.562* (7.23)
High school	.642* (4.49)	.284** (2.36)	.596* (4.73)
Some college	1.088* (6.37)	.756* (5.25)	1.132* (7.28)
College	1.563* (6.33)	1.340* (7.12)	1.735* (7.97)
Postgraduate	1.928* (9.20)	1.710* (10.71)	2.171* (9.59)
Age	.000 (.04)	-.003 (.91)	-.001 (.29)
Female	1.280* (4.73)	.413** (2.30)	1.103* (5.34)
Black	-1.030* (5.74)	-.483* (3.40)	-.971* (6.25)
Other	-1.276** (2.21)	-1.045** (2.50)	-1.400* (3.42)
Hispanic	-1.574* (6.73)	-.588* (3.30)	-1.403* (7.00)
Part-time employed	.188 (1.22)	.451* (3.36)	.330** (2.10)
Not employed	-.004 (.03)	.246** (2.13)	.072 (.56)
BMI	.013 (1.44)	.003 (.46)	.009 (1.07)
Smoker	-.136 (1.08)	-.193*** (1.76)	-.205*** (1.74)
Quit smoking	.121 (.82)	.066 (.57)	.100 (.76)
Vegetarian	0.222 (.72)	-0.196 (.78)	0.083 (.31)
FSP	-.015 (.07)	.076 (.45)	.003 (.018)
WIC	-.252 (.88)	-.547** (2.19)	-.440*** (1.71)
TV ≥ 5 hours	-.235*** (1.68)	-.497* (4.27)	-.216** (2.29)
Diet advice	.401** (2.47)	.462* (3.34)	.481* (3.76)
Compare nutrients always	.566* (3.75)	.211 (1.58)	.665* (5.30)
Compare nutrients sometimes	.720* (6.42)	.079 (.84)	.537* (5.77)
Constant	7.706* (8.20)	1.539** (1.94)	.000
R^2	.260	.169	.425

Note: *, **, and *** indicate coefficient estimates significant at 1-, 5-, and 10-percent levels, respectively, under two-sided t-test. Each reduced-form information equation was regressed on all the independent variables in the model. For brevity, only selected first-stage estimates are reported here.

LV Model

A conceptually attractive and statistically meaningful way to reconcile these differences is to use a latent variable (LV) model where NCK and DHA serve as imperfect measures of an unobserved information variable. This unobserved information variable captures the common variation in NCK and DHA that best predicts HEI. Variation specific to each measure is relegated to error terms. Empirically, this view is supported by the relatively high correlation between NCK and DHA (correlation=0.462) and the much higher correlation between their predicted values from equation 8 (correlation=0.904). Under the LV model, the equations for the unobserved information variable (INFO) can be written as:

$$\text{(9)} \qquad NCK = \lambda_1 INFO + e_1$$
$$DHA = \lambda_2 INFO + e_2$$

where λ's are unknown coefficients to be estimated and e's are uncorrelated measurement errors.

Equations 7 through 9 constitute a latent variable model where INFO cannot be observed directly but is measured indirectly by NCK and DHA. If one assumes the error terms of the three equations to be jointly normal, the unknown parameters of the equations can be estimated by maximum likelihood (Bollen, 1989). The estimation procedure is described in the appendix. During the estimation, all observed variables are expressed as deviations from their means, so the intercept coefficients in equations 7 through 9 are zero.

The scale of the latent variable INFO in equation 9 is undetermined and a restriction is required to identify it (Bollen, p. 239). We identify the scale of INFO by imposing the restriction $\lambda_1 = 1$ so that NCK and INFO have the same scale, that is, a unit change in INFO causes a unit change in NCK. The coefficient for DHA in equation 9, λ_2, is free. The λ_2 estimate is 0.61 with a t-value of 20.98. The R^2 for NCK and DHA are 0.6 and 0.4, respectively. Thus, both NCK and DHA contribute significantly to the measurement of INFO, although, as previously found, NCK is a better measure of nutrition information than DHA. The estimates of the HEI equation under the LV model are reported in table 7, column 6 and the estimates of the INFO equation under the LV model appear in table 8, column 3.

The R^2 for the HEI equation is 0.23, nearly double that of the 2SLS models. The R^2 for the INFO equation is also nearly double that of the first-stage estimates of NCK and DHA under 2SLS. Thus, the LV model appears to have a relatively good fit. The estimated coefficient for the unobserved information variable (INFO) in the HEI equation is 5.05. This estimate is larger than both the NCK and DHA estimates in the 2SLS models.

The estimated effects of income and education in the LV model are negative and significant. This finding implies that, holding all else constant including information levels, meal planners with higher incomes or higher education tend to have lower HEI's. This is in contrast to the 2SLS results, which suggested that higher income or education, all other things equal, has no significant effect on HEI. The LV results, therefore, suggest an even larger informational role for income and education than the 2SLS results. The LV results are not surprising given that similar effects for income and education, conditional on identical information levels, have been found for individual HEI components such as fat, saturated fat, and cholesterol by Carlson and Gould (1994), Gould and Lin (1994) and Variyam, Blaylock, and Smallwood (1997).

Under the LV model, the gender effect is purely informational. Conditional on a constant information level, the HEI's of women do not significantly differ from those of men. On the other hand, if information is held constant, Blacks and respondents of other races as well as Hispanics have significantly higher HEI's than Whites and non-Hispanics. Holding information constant at the sample mean of the HEI's, Blacks have 5 percent higher HEI's than Whites, respondents of other races have 10 percent higher HEI's than Whites, and Hispanics have 14 percent higher HEI's than non-Hispanics.

From the LV model estimates, a year of age adds about one-fifth of a point to the HEI, an additional unit of BMI reduces HEI by a similar amount, and smokers' HEI's are about 3.5 points lower than nonsmokers'. In contrast to income, schooling, or gender effects, however, the effects of age, body mass, and smoking are almost entirely due to the different tastes and preferences associated with these characteristics and not due to any informational differences. The first-stage coefficient estimates for age and BMI are insignificant, while smoking has a small but significant negative effect on information at the 10-percent level (table 8). The absence of strong informational effects explains the relative stability of the coefficient estimates for these variables under OLS, 2SLS, and LV in table 7. The results for age agree with similar effects found for fat, saturated fat, and cholesterol by Carlson and Gould (1994) and by Variyam, Blaylock, and Smallwood (1997). For BMI, a negative diet quality effect has also been obtained by Variyam, Blaylock, and Smallwood (1997) with respect to cholesterol; by Yen, Jensen, and Wang (1996) for eggs; and by Guthrie and Fulton (1995) for the consumption of food groups conforming to the *Food Guide Pyramid*. The results for smoking confirm previous findings by McPhillips, Eaton, and Gans (1994) that smokers eat a less healthful diet than nonsmokers. Given the weak informational effect related to smoking, such dietary behavior on the part of smokers may be related to their tendency to underestimate health risks more so than nonsmokers do (Viscusi, 1990). This view is supported by the lack of significance of the quit-smoking variable, since those who have quit smoking are likely to have done so due to a higher perceived health risk from smoking compared with current smokers.

The estimated household food program participation and time and resource allocation effects are either insignificant or inconclusive. The presence of children, household size, and sex of the head of the household have no significant effects on HEI or information (from results not reported in table 8). Similarly, both the FSP and WIC coefficients are insignificant in the HEI equation. Interestingly, the WIC coefficient is relatively large (2.004 in table 7) in the LV model. Its lack of significance may be related to the small proportion of WIC households, about 4 percent of the sample. WIC participation also has a small but negative effect on nutrition information. These WIC effects clearly need further investigation because of the implied benefits of targeted nutrition education.

The evidence on employment status is mixed, with a largely insignificant direct effect on HEI, but some positive informational effects indicated by the significant coefficient for part-time employment on INFO under the LV model. The insignificant effects for those not employed outside the home is surprising given that the expected effect is through greater time available for shopping and food preparation (Horton and Campbell, 1991). Also surprising are the insignificant effects of being a vegetarian, since strong beneficial diet quality effects have been found previously by Variyam, Blaylock, and Smallwood (1996) for dietary fiber; and by Variyam, Blaylock, and Smallwood (1997) for fat, saturated fat, and cholesterol.

The first-stage effects of nutrition information sources are as expected. Excess television watching (TV5) has a significant negative effect on nutrition information, while the effects of receiving diet advice from a physician or a dietitian and the use of nutrition labels are all positive. These results are similar to the findings of Carlson and Gould (1994) and Gould and Lin (1994). There is virtually no effect for the degree of urbanization on HEI. Region variables and the year variable show some effect, indicating some regional and temporal variations in prices, tastes, or preferences.

Conclusions

Our results suggest that nutrition information affects overall diet quality, even after controlling for individual differences in a host of personal and household characteristics including income, education, age, gender, race, ethnicity, smoking behavior, and body mass. This evidence adds to previous findings for individual nutrients such as fat, saturated fat, and cholesterol (Carlson and Gould, 1994; Gould and Lin, 1994; Variyam, Blaylock and Smallwood, 1997); dietary fiber (Variyam, Blaylock, and Smallwood, 1996); and for food groups (Guthrie and Fulton, 1995; Yen, Jensen, and Wang, 1996). This micro-level evidence complements evidence of informational effects at the aggregate level found by Brown and Schrader, 1990; Capps and Schmitz, 1991; and Chern, Loehman, and Yen, 1995. Together, these findings show how consumers have absorbed and used the information linking diet and health in their gradual shift toward more healthful diets.

These findings also suggest a continued role for nutrition education efforts to close the persistent gap between actual and healthful diets. In this regard, our study has placed a special focus on isolating the role of nutrition information on overall diet quality from the role of various consumer characteristics that affect both information and diet quality. Because sociodemographic variables and nutrition information affect diet quality separately, previous studies linking diet quality to sociodemographic characteristics without taking nutrition information into account may be questionable. For example, without controlling for nutrition information, females tend to have higher HEI scores (table 5; table 7, column 1). They also tend to possess more nutrition information than men, other things equal (table 8). When the effects of nutrition information are properly accounted for, men and women of similar characteristics have similar HEI scores (table 7, column 6). Thus, increasing male meal planners' stock of nutrition information should result in increased overall diet quality. Similarly, the results for race and ethnicity show that main meal planners who are Black, of other race, or Hispanic will benefit greatly from additional nutrition information. Targeting these main meal planner groups for nutrition education and promotion efforts should result in a significant improvement of their overall diet quality. Given the limited resources available for nutrition education and promotion, our results suggest guidelines for the efficient allocation of these resources.

References

Basiotis, P. P., S. Johnson, K. Morgan, and J. Chen. "Food Stamps, Food Costs, Nutrient Availability, and Nutrient Intake," *Journal of Policy Modeling*, 9 (1987), pp. 383-392.

Basiotis, P. P., J. D. Hirschman, and E. T. Kennedy. "Economic and Sociodemographic Determinants of 'Healthy Eating' As Measured by USDA's Healthy Eating Index," *Consumer Interests Annual*, 42 (1996), pp. 81-88.

Becker, G. S. "A Theory of the Allocation of Time," *Economic Journal*, 75 (1965), pp. 493-517.

Behrman, J. R., and A. B. Deolalikar. "Health and Nutrition," in *Handbook of Development Economics*, Vol. I, eds. H. Chenery and T. N. Srinivasan. Amsterdam: North-Holland, 1988, pp. 631-771.

Bollen, K. A. *Structural Equations with Latent Variables*. New York: Wiley, 1989.

Brown, D. J., and L. F. Schrader. "Cholesterol Information and Shell Egg Consumption," *American Journal of Agricultural Economics*, 72 (1990), pp. 548-55.

Butler, J. S., and J. E. Raymond. "The Effect of the Food Stamp Program on Nutrient Intake," *Economic Inquiry*, 34 (1996), pp. 781-98.

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

Carlson, K. A., and B. W. Gould. "The Role of Health Knowledge in Determining Dietary Fat Intake," *Review of Agricultural Economics*, 16 (1994), pp. 373-86.

Chern, W. S., E. T. Loehman, and S. T. Yen. "Information, Health Risk Beliefs, and the Demand for Fats and Oils," *Review of Economics and Statistics*, 77 (1995), pp. 555-64.

Connor, J. M. "North America as a Precursor of Changes in Western European Food-Purchasing Patterns," *European Review of Agricultural Economics*, (1994), pp. 155-73.

Dattilo, A. M. "Dietary Fat and its Relationship to Body Weight," *Nutrition Today*, 27 (1992), pp. 13-20.

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

Gawn, G., R. Innes, G. Rausser, and D. Zilberman. "Nutrient Demand and the Allocation of Time: Evidence from Guam," *Applied Economics*, 25 (1993), pp. 811-30.

Gould, B. W., and H. C. Lin. "Nutrition Information and Household Dietary Fat Intake," *Journal of Agricultural and Resource Economics*, 19 (1994), pp. 349-65.

Greene, W. H. *LIMDEP Version 7.0 User's Manual*. New York: Econometric Software, 1995.

Grossman, M. "The Correlation between Health and Schooling," in *Household Production and Consumption*, N. E. Terleckyj, ed., New York: Columbia University Press, 1975, pp. 147-211.

Grossman, M., and R. Kaestner. "Effects of Education on Health." Presented at the conference, "Social Benefits of Education: Can They be Measured?" Meridian International Center, Washington, DC, 1995.

Guthrie, J. F., and L. H. Fulton. "Relationship of Knowledge of Food Group Servings Recommendations to Food Group Consumption," *Family Economics and Nutrition Review*, 8 (1995) pp. 2-17.

Horton, S., and C. Campbell. "Wife's Employment, Food Expenditures, and Apparent Nutrient Intake: Evidence from Canada," *American Journal of Agricultural Economics*, 73 (August 1991), pp. 784-94.

Ippolito, P. M., and A. D. Mathios. "Information, Advertising, and Health Choices: A Study of the Cereal Market," *Rand Journal of Economics*, 21 (1990), pp. 459-80.

Jensen, H. H., T. Kesavan, and S. R. Johnson. "Measuring the Impact of Health Awareness on Food Demand," *Review of Agricultural Economics*, 14 (1992), pp. 299-312.

Kenkel, D. S. “Health Behavior, Health Knowledge, and Schooling,” *Journal of Political Economy*, 99 (1991), pp. 287-305.

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), pp. 1103-08.

Kushi, L. H., A. R. Folsom, D. R. Jacobs, R. V. Luepker, P. J. Elmer, and H. Blackburn. “Educational Attainment and Nutrient Consumption Patterns: The Minnesota Heart Survey,” *Journal of the American Dietetic Association*, 88 (1988), pp. 1230-36.

Lancaster, K. *Consumer Demand*: *A New Approach*. New York: Columbia University Press, 1971.

McGinnis, J. M., and W. H. Foege. “Actual Causes of Death in the United States,” *Journal of the American Medical Association*, 18 (1993), pp. 2207-12 .

McPhillips, J. B., C. B. Eaton, and K. M. Gans. “Dietary Differences in Smokers and Nonsmokers from Two Southeastern New England Communities,” *Journal of the American Dietetic Association*, 94 (1994), pp. 287-92.

Moorman, C., and E. Matulich. “A Model of Consumers’ Preventive Health Behaviors: The Role of Health Motivation and Health Ability,” *Journal of Consumer Research*, 20 (1993), pp. 208-28.

Morgan, K. J. “Socioeconomic Factors Affecting Dietary Status,” *American Journal of Agricultural Economics*, 68 (1986), pp. 1240-46.

Pindyck, R. S., and D. L. Rubinfeld. *Econometric Models & Economic Forecasts*. New York: McGraw Hill, 1991.

Pitt, M. M., and M. R. Rosenzweig. “Health and Nutrient Consumption Across and within Farm Households,” *Review of Economics and Statistics*, 67 (1985), pp. 212-23.

Putler, D. S., and E. Frazão. “Consumer Awareness of Diet-Disease Relationships and Dietary Behavior: The Case of Dietary Fat,” *Journal of Agricultural Economics Research*, 45 (1994), pp. 3-17.

Senauer, B., and M. Garcia. “Determinants of Nutrition and Health Status of Preschool Children: An Analysis with Longitudinal Data,” *Economic Development and Cultural Change*, (1991), pp. 371-89.

Singh, G. K., K. D. Kochanek, and M. F. MacDorman. “Advance Report of Final Mortality Statistics, 1994.” *Monthly Vital Statistics Report*, Supplement, Vol. 45, U.S. Dept. Health and Human Services, National Center for Health Statistics, September 1996.

Thomas, P. R. *Improving America’s Diet and Health*. Washington, DC: National Academy Press, 1991.

U.S. Department of Agriculture, Center for Nutrition Policy and Promotion. *The Healthy Eating Index*. Washington, DC, October 1995.

U.S. Department of Agriculture. *The Food Guide Pyramid*. Home and Garden Bulletin No. 252, 1992.

U.S. Department of Agriculture and U.S. Department of Health and Human Services. *Nutrition and Your Health: Dietary Guidelines for Americans*. Fourth edition, Home and Garden Bulletin No. 232, 1995.

U.S. Department of Health and Human Services, Public Health Service. *Healthy People 2000*. DHHS (PHS) Publication No. 91-50212, 1991.

Variyam, J. N., J. Blaylock, and D. Smallwood. “A Probit Latent Variable Model of Nutrition Information and Dietary Fiber Intake,” *American Journal of Agricultural Economics*, 78 (1996), pp. 628-39.

Variyam, J. N., J. Blaylock, and D. Smallwood. *Diet-Health Information and Nutrition: The Intake of Dietary Fats and Cholesterol*. U.S. Dept. of Agr., Econ. Res. Serv., TB-1855, 1997.

Viscusi, K. W. “Do Smokers Underestimate Risks?” *Journal of Political Economy*, 98 (1990): pp. 1253-1269.

Yen, S. T., H. H. Jensen, and Q. Wang. “Cholesterol Information and Egg Consumption in the US: A Nonnormal and Heteroskedastic Double-Hurdle Model,” *European Review of Agricultural Economics*, 23 (1996), pp. 343-56.

Appendix

Let $\mathbf{y}_{(3\times1)}$ = (HEI, NCK, DHA)´, $\boldsymbol{\eta}_{(2\times1)}$ = (HEI, INFO)´ and let $\mathbf{x}_{(R\times1)}$ be a vector of all the independent variables X_p and X_q. Then, assuming that all variables are mean-centered, equations 7 through 9 can be written in the general latent variable model form as

(10) $$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}$$

(11) $$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}$$

where

$$\boldsymbol{\Lambda}_{(3x2)} = \begin{bmatrix} 1 & 0 \\ 0 & \lambda_1 \\ 0 & \lambda_2 \end{bmatrix}, \quad \mathbf{B}_{(2x2)} = \begin{bmatrix} 0 & \beta \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{\Gamma}_{(2xR)} = \begin{bmatrix} \boldsymbol{\alpha}_{(1xR)} \\ \boldsymbol{\gamma}_{(1xR)} \end{bmatrix},$$

$$\boldsymbol{\epsilon}_{(3x1)} = (0, e_1, e_2)', \quad \boldsymbol{\zeta}_{(2x1)} = (u, v)',$$

such that $\boldsymbol{\alpha}_{(1\times R)}$ contains the $\boldsymbol{\alpha}_p$ parameters of equation 7 and $\boldsymbol{\gamma}_{(1\times R)}$ contains the $\boldsymbol{\gamma}_q$ parameters of equation 8. Note $\boldsymbol{\alpha}$ that $\boldsymbol{\gamma}$ and will contain zero restrictions so as to exclude those variables in **x** that do not enter the right-hand side of equations 7 and 8 respectively.

The maximum likelihood (ML) estimation of equations 10-11 is based on the idea that for multivariate normally distributed **y**, the variance-covariance matrix of the observables is a sufficient statistic. The variance-covariance matrix of the observables (**y x**) can be written as a function of the unknown parameters of equations 10-11. This can be seen by substituting equation 11 into equation 10:

(12) $$\mathbf{y} = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\zeta} + \boldsymbol{\epsilon},$$

so that,

(13) $$E(\mathbf{y}\mathbf{y}') = \boldsymbol{\Lambda}(I - \mathbf{B})^{-1}\boldsymbol{\Gamma}\mathbf{S}_{xx}\boldsymbol{\Gamma}'(I - \mathbf{B})^{-1}\boldsymbol{\Lambda}'$$
$$+\boldsymbol{\Lambda}(I - \mathbf{B})^{-1}\boldsymbol{\Psi}(I - \mathbf{B})^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$$

where E(**xx'**) = $\mathbf{S}_{xx}$, E($\boldsymbol{\zeta\zeta}'$) = $\boldsymbol{\Psi}$, and E($\boldsymbol{\epsilon\epsilon}'$) = $\boldsymbol{\Theta}$. Note that E($\mathbf{x}\boldsymbol{\zeta}'$) = **0**, E($\mathbf{x}\boldsymbol{\epsilon}'$) = **0**, and E($\boldsymbol{\zeta\epsilon}'$) = **0**. Similarly,

(14) $$E(\mathbf{y}\mathbf{x}') = \boldsymbol{\Lambda}(I - \mathbf{B})^{-1}\boldsymbol{\Gamma}\mathbf{S}_{xx}.$$

Let E[(**y x**) (**y x'**)] = $\boldsymbol{\Sigma}(\boldsymbol{\delta})$ where $\boldsymbol{\delta}$ is a vector of the unknown parameters. Therefore, $\boldsymbol{\delta}$ can be estimated by fitting the variance-covariance matrix implied by the model, $\boldsymbol{\Sigma}$, to the empirical variance-covariance matrix **S**. This is achieved by minimizing the ML fitting function

(15) $$F_{ML} = 1n\left|\boldsymbol{\Sigma}(\boldsymbol{\delta})\right| + tr\left[\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\delta})\right] - 1n\left|\mathbf{S}\right| - (m + R)$$

where m is the number of variables in **y** (Bollen, 1989, pp. 319-338).