Model Specification, Estimation, and Forecasting

The ultimate objective of the study was to project commodity consumption, both at home and away from home, for every 5 years, ending with the year 2020. Because we utilized data from the 1994-96 and 1998 CSFII, our consumption projections began with the year 2000.

The analysis involved several tasks. The first task was to model eating-out behavior and household heads' diet-health knowledge. Estimated eating-out and knowledge variables were then combined with economic, social, and demographic data to estimate the consumption of 25 food groups. Using the projected values for the economic, social, and demographic characteristics, future food consumption was predicted for the years 2000, 2005, 2010, 2015, and 2020.

The EPA's FCID and ARS's PSD were used, together with the ERS conversion-factor database, to create a food-commodity translation database so that projected food consumption could be converted to commodity consumption.

Indirect and Direct Approaches

There are two approaches to estimating commodity consumption using food consumption data—indirect and direct. The indirect approach begins with the estimation of food consumption, then derives commodity consumption using conversion factors. The direct approach first converts foods to commodities and then estimates commodity consumption. Under the argument that consumers make food choices rather than commodity choices, we employed the indirect approach. The econometric model for ith food consumed by an individual j can be expressed as:

(1) $$F_{ij} = f(P_1, P_2, \ldots, P_n, P_{gs}, Y_j, X_j) \quad = 1, 2, \ldots n$$

where F_{ij} is the amount of ith food consumed by jth individual; P_i is the price of ith food; P_{gs} is a price index for nonfood goods and services, Y_j is income, and X_j is a vector of social and demographic characteristics of the individual j.

Then the projected values of the exogenous variables (P, X, and Y) are plugged into the estimated food consumption equations to forecast food consumption. Using the food-commodity translation database (equation 2), the consumption of n foods can be converted to the consumption of m commodities (equation 3).

(2) $$Q_k = \Sigma_i T_{ki} F_i$$

where T_{ki} is a translation coefficient indicating the units of the commodity Q_k contained in each unit of the food F_i.

(3) $$Q_{kj} = g(P_1, P_2, \ldots, P_n, Y_j, X_j) \quad k = 1, 2, \ldots m.$$

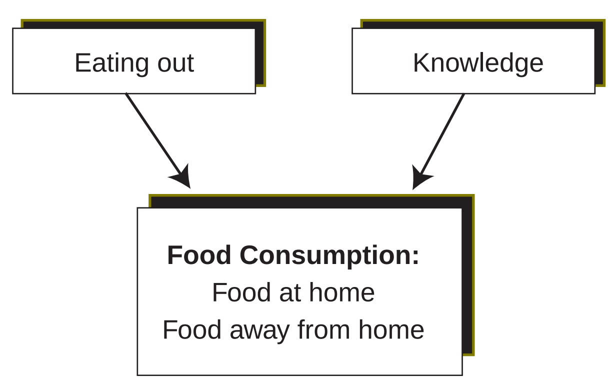
The direct approach is to apply the food-commodity translation database to derive the consumption of commodity k by individual j, Q_{kj}, from the consumption of food i consumed by individual j, F_{ij}. Then equation 3 is estimated.

Model Structure and Specification

We specified a two-step econometric system to model consumers' food consumption (fig. 1). In the first step, we estimated consumers' eating-out habits and their diet-health knowledge. In the second step, food consumption, both at home and away from home, were related separately to consumers' eating-out habits, knowledge, income, and social and demographic characteristics.

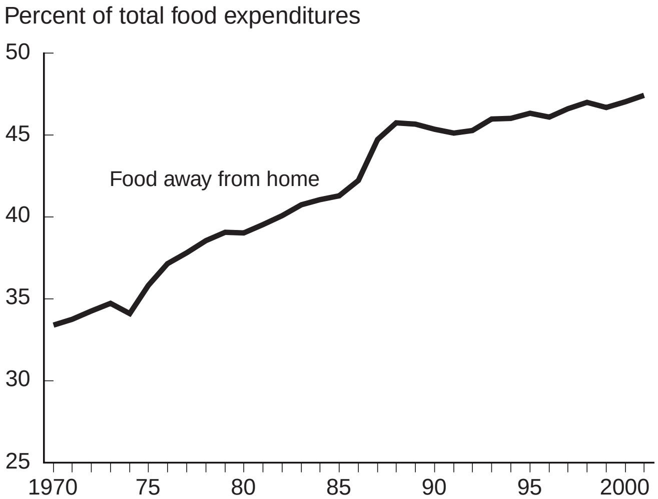
Over the past three decades, eating out has become increasingly popular for Americans. In 2000, Americans spent 47 percent of their food expenditures away from home, up from 33 percent in 1970 (fig. 2). Americans

Figure 1
Two-stage econometric modeling



Source: Economic Research Service, USDA.

Figure 2
U.S. expenditures for food away from home as a share of total food expenditures



Source: Economic Research Service, USDA.

tend to eat different kinds of foods and different amounts when eating out than eating at home (Lin, Guthrie, and Frazao). If the upward trend in eating out continues, we expect food and commodity consumption will change accordingly. Therefore, we paid special attention to modeling the decision to eat out and its effect on the type and amount of food and commodities consumed at home and away from home. We also examined the growth of the at-home and away-from-home segments of the food and commodity markets.

In addition to eating out more frequently, U.S. consumers have also changed their food consumption patterns in reaction to the flood of diet and health information coming out of the Nation's laboratories and research institutions (Variyam and Golan). We therefore tried to model how diet-health knowledge and attitudes are reshaping food consumption.

To our knowledge, no research has been conducted to model the effects of both eating out and diet-health knowledge on food consumption. In this study, eating out and consumers' diet-health knowledge were modeled as two separate equations in the first step of a two-step recursive econometric system. Eating-out behavior was hypothesized to be affected by household characteristics, race and education of the household head, and individual characteristics, as shown below.

(4) FAFH = f(Y, Hsize, Htype, Tenancy, Region, Metro, Edu, Race, Sex, Age, School, Employ, Weekend)

where FAFH measures eating out by the individual (subscript is dropped for simplicity). Y is per capita household income. Hsize is the household size. Htype is the type of household structure (single or dual heads, with or without children). Tenancy is the household housing arrangement (own, rent, etc.). Region is the Census region. Metro indicates if the respondent resides in a city, suburb, or rural area. Edu is the education level of the respondent. Race is the race and ethnicity of the household head. Sex is the gender of the respondent. Age is a vector of age classes for the respondent. School indicates if the respondent was in school. Employ indicates the employment status of the respondent. Weekend indicates if the day of food intake was a Saturday or Sunday.

In the 1994-96 CSFII, an adult age 20 or above who provided the first-day intake data was randomly chosen from each household to provide information related to his/her diet-health knowledge and attitude. The data were used to construct a knowledge variable (KNOW) for household heads. This knowledge variable was hypothesized to be affected by the characteristics of the household and the household head (Variyam et al.), as shown below:

(5) KNOW = f(Y, Htype, Region, Metro, Edu, Race, Sex, Hhage, Hemploy)

where Hhage is the age of the household head and Hemploy is the percent of household heads employed; other variables were already discussed in equation 4.

The fitted values of the two dependent variables (FAFH and KNOW) from the first step were treated as exogenous variables in the second step, in which a system of food consumption was estimated. We separated food at home and food away from home and estimated two sets of food consumption equations: one set for 25 food groups at home and the other for 25 food groups away from home. By doing so, we were able to estimate the differing effects of exogenous variables (including eating out, knowledge, income, social, and demographic variables) on food consumption at home and away from home. For example, as consumers eat out more, they are likely to eat more french fries away from home but fewer french fries at home. Another advantage of separating food consumed at home and food consumed away from home is to enable us to forecast the growth in food and commodity consumption in these two market segments.

Because CSFII respondents only reported the amount of foods they ate but not the price they paid, price variables in equation 1 were dropped from equation 6.

(6) F = f(FAFH, KNOW, Y, Hsize, Htype, Season, Region, Metro, Race, Sex, Age)

where Season is the season during which the individual's food intake occurred; other variables were already discussed.

For detailed definitions of all variables as well as their descriptive statistics (weighted means and standard deviations) see table 1.

Estimation Procedure

Food intake data in CSFII were collected for sample persons of all ages, while DHKS data were collected for those age 20 and above who completed the first-day intake recall. Therefore, the DHKS sample is a subset of the CSFII sample. The knowledge variable is included in the food consumption equation 6, which was estimated using the CSFII sample. So, the knowledge variable had to be generated for some CSFII respondents. This was accomplished by fitting the knowledge equation 5 for DHKS respondents who were also household heads. Because personal data for household heads were collected for each CSFII respondent, the data and the fitted knowledge equation were used to generate household heads' knowledge for the whole CSFII sample. The knowledge equation was estimated with the weighted ordinary-least-squares estimator, using the DHKS sample weight.

On any given day, not everyone consumed all the 25 food groups, either at home or away from home. Therefore, a cluster of zero consumption values for eating out and for various food groups was observed in the data—making it necessary to estimate a censored regression model. Any statistical procedure that does not account for zero observations produces inconsistent parameter estimates.

Tobin was the first to propose a censored normal regression model (Tobit model) to deal with censored data in regression. Since the development of Tobit, many censored regression estimators have been proposed, including the maximum-likelihood procedure of Amemiya, Lee, Wales and Wood, and the two-step procedures of Heien and Wessells, Shonkwiler and Yen, and Perali and Chavas.

In this study, we chose the Tobit procedure for its ease of estimation. The Tobit model can be expressed as:

(7)
$$q_i = x'\beta_i + \varepsilon_i \quad \text{if } x'\beta_i + \varepsilon_i > 0$$
$$= 0 \quad \text{if } x'\beta_i + \varepsilon_i \le 0, \; i = 1, 2, \ldots, n$$

where q_i denotes the endogenous variable, x is the vector of exogenous variables, and ε_i is the error term.

The Tobit procedure was used to estimate 50 censored consumption equations (25 food groups, at home and away from home) as well as a separate censored eating-out equation, using the CSFII sample.

Table 1—Variables used in the two-step recursive system and their descriptive statistics

Variable	Definition	Mean	Standard deviation
Eating out	Percent of food consumption, in quantity terms, that was prepared away from home (%)	22.6	22.8
Knowledge	Scores of diet-health knowledge of the household head	17.2	4.1
Income	Household income, per capita (in $1,000)	14.7	11.6
Employment	The respondent's employment status (0, 1)	0.47	0.50
Hemploy	Percent of household heads employed (0, 0.5, 1)	0.57	0.44
Student	The respondent was attending school (0, 1)	0.18	0.39
Weekend	Number of intake days that fell on weekend (0, 1, 2)	0.57	0.57
High school	The respondent completed high school education but did not go to college (0, 1)	0.27	0.44
	Base = respondents who did not finish high school		
Some college	The respondent went to college but did not graduate with a degree (0, 1)	0.31	0.46
College	The respondent completed a college degree (0, 1)	0.22	0.41
Male	The respondent is male (0, 1)	0.49	0.50
Age 0-4	The respondent is age 0-4 (0, 1)	0.08	0.27
Age 5-9	The respondent is age 5-9 (0, 1)	0.07	0.26
Age 10-14	The respondent is age 10-14 (0, 1)	0.08	0.26
Age 15-19	The respondent is age 15-19 (0, 1)	0.07	0.25
Age 20-29	The respondent is age 20-29 (0, 1)	0.14	0.35
Age 30-44	The respondent is age 30-44 (0, 1)	0.24	0.43
Age 45-54	The respondent is age 45-54 (0, 1)	0.12	0.33
Age 55-64	The respondent is age 55-64 (0, 1)	0.08	0.27
Age 65-74	The respondent is age 65-74 (0, 1). Base = respondents age 75 and over	0.07	0.25
HH age 20-34	The household head is age 20-34 (0, 1)	0.26	0.44
HH age 35-54	The household head is age 35-54 (0, 1)	0.48	0.50
HH age 55-69	The household head is age 55-69 (0, 1). Base = respondents age 70 and over	0.15	0.35
Black	The respondent is non-Hispanic Black (0, 1). Base = non-Hispanic White	0.13	0.33
Hispanic	The respondent is Hispanic (0, 1)	0.11	0.31
Asian	The respondent is Asian/Pacific Islander (0, 1)	0.03	0.17
Other	The respondent's race/ethnicity is none of the above nor White (0, 1)	0.01	0.12
HH type1	The household is dual-headed, with children (0, 1)	0.48	0.50
HH type2	The household is dual-headed, without children (0, 1)	0.27	0.44
HH type3	The household is single-headed (either male or female), with children (0, 1)	0.10	0.31
	Base = single-headed households without children		
Midwest	The respondent resides in the Midwestern States (0, 1)	0.23	0.42
South	The respondent resides in the Southern States (0, 1)	0.35	0.48
West	The respondent resides in the Western States (0, 1). Base = Northeast	0.22	0.41
Nonmetro	The respondent resides in a rural area (0, 1)	0.21	0.41
Suburb	The respondent resides in a suburb (0, 1). Base = central city	0.47	0.50
HH size	Number of household members	3.42	1.67
Tenancy	The respondents own their homes (0, 1)	0.67	0.47
Quarter 1	The first day of intake falls in January - March (0, 1)	0.25	0.43
Quarter 2	The first day of intake falls in April - June (0, 1)	0.25	0.43
Quarter 3	The first day of intake falls in July - September (0, 1). Base = October - December	0.25	0.50