Journal of Agricultural and Resource Economics, 20(1):165–177

Determinants of Nutrient Demand: A Nonparametric Analysis

Cyrus A. Ramezani

The influence of socioeconomic variables on nutrient intake is studied using nonparametric procedures that admit estimation of multivariate functions. The analysis indicates a nonlinear relation between intake, age, education, and income. Specifically, intake rises with income reaching an inflection point beyond which it is essentially flat. Socioeconomic variables influence intake primarily at lower-income levels. Nonparametric procedures prove useful in avoiding ad hoc specifications that would fail to uncover these findings.

Key words: additive models, AVAS, nonparametric methods, nutrient demand

Introduction

Understanding the link between the nutritional well-being of individuals and their socioeconomic status is important for influencing consumer behavior through public policy. Interest in understanding this relationship dates back to the 1940s and the development of the "minimum cost" diet plans (Senauer, Asp, and Kinsey). Adrain and Daniel conducted one of the first comprehensive studies of the relationship between nutrient intake and socioeconomic status. Since then, numerous researchers have reexamined this relationship using newer data and more innovative techniques. Senauer, Asp, and Kinsey and Morgan provide a complete survey and appraisal of this literature.[1] They indicate that while research in this area has progressed considerably, the influence of some key socioeconomic variables needs further investigation.

In particular, Morgan and Davis suggest a need for more interdisciplinary research to better understand the relationship between nutrient intake and income, household size, and education. The literature contains a diversity of findings with respect to these variables (Morgan). The variation in findings are reminiscent of similar analysis in the economic development literature, where researchers have obtained a wide range of estimates for income elasticity of nutrient intake, even among the very poor. This has led to an important debate regarding the effectiveness of increasing incomes in alleviating malnutrition in poor countries (Behrman, Deolalikar, and Wolfe). Recent studies by Strauss and Thomas and Subramanian and Deaton have shown that the low elasticity estimates are partly due to model misspecification. Measurement error may also cause the differences in these findings.

The author is an assistant professor in the Department of Consumer Science at the University of Wisconsin at Madison
Earlier versions of this paper were presented at the American Agricultural Economic Association's meetings and seminars at the University of California-Berkeley and Stanford University's Food Research Institute.
Comments from J. Chalfant, S. Lane, M. Carter, H. Jensen, K. Holden, B. Wolfe, the editors, and two anonymous referees improved the paper considerably. Any errors remaining are the author's responsibility.
[1] The reference section of Senauer, Asp, and Kinsey provides a complete overview. For developing countries, nutrient demand, particularly for calories, has been widely studied (Behrman and Doelalikar).

The variation in findings based on U.S. data reported in the literature may also be a result of model "misspecification." The aim of this article is to investigate this likelihood using recent data and new methods of functional-form estimation. Nonparametric procedures have been used to determine the interaction between calorie intake and income, but past studies have been primarily bivariate and, therefore, may be biased due to omitted variables. A "kinked" relationship between caloric intake and income was identified using locally weighted regressions (Ramezani; Ramezani and Roeder; Strauss and Thomas) and kernel estimators (Subramanian and Deaton). These studies, however, fail to control for the influence of other socioeconomic variables. The estimation procedures we use simultaneously control for all variables in the model. Hence, the intake-income relationship estimated is conditional on the appropriate functional specification for other explanatory variables. To assess the importance of specification, the selected independent and explanatory variables are defined as in previous studies. U.S. data provide a larger sample with more socioeconomic variables, more intake measures, and less severe "measurement error."

Perhaps the most useful aspect of nonparametric procedures is that they can suggest parametric specifications that best approximate the nonlinearities specific to any data. This is the spirit in which nonparametric procedures are used in this article. We rely on a relatively new nonparametric functional-form estimation technique—additivity and variance stabilization (AVAS)—to uncover the nonlinearities that are inherent to nutrient intake data. The exploratory analysis of the data using AVAS provides the basis for specifying a parametric model that is estimated using ordinary least squares (OLS). Policy discussions are based on the estimated parametric models.

Model and Data

Household decision theory suggests that demand for nutrient N is related to n demographic and economic variables, including composite variables representing their interaction, $S_1, \ldots, S_n$:

$$N = f(S_i, \ldots, S_n) + \varepsilon, \tag{1}$$

where ε is an error term added for estimation purposes, and the distributional assumptions about ε depend upon the estimation methods used (Behrman and Deolalikar). Before estimating (1), researchers must select the socioeconomic variables, nutrient intake measures, and the functional form. Policy interest has an important bearing on the choice of intake measure and socioeconomic variables (Morgan). The choice of functional form, however, has generally been arbitrary. The standard practice is to assume a linear form and add quadratic (income) term or to use the double-logarithmic specification. Here a nonparametric estimate of the multivariate function represented by (1) is obtained.

The individual portion of the Nationwide Food Consumption Survey (NFCS 1987–88) is used for this study (USDA). This survey of over 10,000 individuals' consumption is collected by the U. S. Department of Agriculture once every decade. The data consist of detailed information on three days of food and nutrient intake by individuals in all age groups and socioeconomic categories, and on all days of the week, during all seasons of the year (Hama and Riddick). The socioeconomic variables selected—age, gender, marital status, race, education, urbanization, regional location, household size, and income—facilitate comparisons with previous studies.

The final sample used here consists of 5,870 individuals. The data exclude pregnant women, lactating mothers, and individuals less than 19 years old. The nutritional needs of these individuals are markedly different from the rest of the sample (Murphy et al.). The sample also excludes the small portion of the survey—110 individuals—who received assistance through various government programs. It seems more appropriate to treat this group separately, particularly since the focus of this analysis is on the determinants of intake rather than assessment of specific government programs. A final caveat to note is that no direct measure of an individual's nutritional knowledge was available in this sample. Capps and Schmitz indicate that the level of nutritional knowledge may be an important determinant of intake. Educational attainment is used as an indirect measure of nutritional knowledge.

Five measures of intake are used: total calorie intake, calorie, protein, calcium intake as a percentage of the recently revised individual specific Recommended Dietary Allowance (RDA), and percentage of calorie intake from fats. These measures correct for differences in dietary quality and requirements across the sample. Previous studies suggest that these measures are useful for assessing diet quality (Murphy et al.).

Table 1 provides definitions, means, and standard deviations of the dependent and explanatory variables.[2] It is possible to form clear hypotheses regarding the influence of some but not all explanatory variables in the table. For example, we would expect that caloric intake would rise with age below a certain threshold, rise and then fall with income, rise and then perhaps fall with education, and rise with household size up to a limit. Without a behavioral model, it is difficult to form hypotheses with respect to other variables. The premise of this article is that the influence of most socioeconomic variables will vary for individuals in different income strata. Income is important because it determines access to quantity and quality of foods consumed.[3]

Nonparametric Functional-Form Estimation

There are various occasions when economic theory, the data, or both suggest a nonlinear relationship between the dependent and the explanatory variables. Consumer and production theory provide several examples. In nutrient demand functions the variables may interact nonlinearly. The specification of a functional form is not straightforward. The usual practice is to take the data as given and to impose a structure that is sufficiently general, as with flexible functional forms. An alternative approach is to use nonparametric procedures that permit the data to determine the best specification.[4]

This section introduces a specific nonparametric curve estimation technique—the additivity and variance stabilization (AVAS) method (Tibshirani). The presentation will be brief, emphasizing the intuition and usefulness of AVAS rather than technical details, for which the interested reader should consult Tibshirani. The proceeding discussion aims to explain the potential use of these methods in uncovering nonlinearities in the data. As shown, AVAS is useful for specifying a parametric model that is simple to interpret.

---

[2]The mean calorie intake as a percentage of RDA reported in the table appears low. This is partly due to a systematic underreporting of intake that often accompanies dietary recall surveys. In a recent study of U.S. population, Mertz et al. found mean underreporting as high as 18% of actual intake.

[3]The previous year's before-tax income is used so as to avoid the problem associated with the simultaneous determination of intake and current income (expenditures).

[4]Kernel estimators are developed in detail by Härdle and Silverman and are nicely surveyed by Altman.

Table 1. Description of Intake Measures and Socioeconomic Variables

Variable	Definition	Mean	SD
TOTNER	Total calorie intake	1,717.00	672.00
RDAENER	Calorie intake as percent of the individual's RDA	73.00	25.10
RDAPROT	Protein intake as percent of the individual's RDA	128.00	45.90
RDACALC	Calcium intake as percent of the individual's RDA	81.00	44.60
FATPERC	Percent of total calorie intake from fats	37.00	7.00
AGE	Individual's age (years)	45.60	17.50
SEX	Dichotomous: male = 1	0.45	0.50
CITY	Dichotomous: city dweller = 1	0.21	0.40
RURAL	Dichotomous: nonmetropolitan dweller = 1	0.29	0.45
SUBURB	Dichotomous: suburban dweller = 1	0.50	0.50
HHSZ	Household size (number of individuals)	2.90	1.40
INCOME	Household income before taxes ($)	29,275.00	23,832.00
PCINC	Income divided by HHSZ ($)	11,419.00	9,616.00
PCTPOV	Income as percent of federal poverty level	330.00	267.00
RACE	Dichotomous: white = 1	0.87	0.30
NEAST	Dichotomous: Northeast = 1	0.20	0.40
SOUTH	Dichotomous: South = 1	0.34	0.47
WEST	Dichotomous: West = 1	0.18	0.38
MWEST	Dichotomous: Midwest = 1	0.27	0.44
AVGEDUC	Average education of both household heads (years)	12.30	2.90
DUOHEAD	Dichotomous: double-head household = 1	0.75	0.43

AVAS belongs to a broad class of functional specifications called the generalized additive models (GAM) (Hastie and Tibshirani). The additive specification and AVAS are well suited to the analysis of economic data because they allow estimating multivariate functions, admit nonlinearity of both the dependent and the explanatory variables, and accommodate interaction effects. Under the GAM specification, an arbitrary function of the dependent variable is related to the sum of arbitrary functions of the independent variable(s). The approximation provided is superior to a linear specification. The costs of gaining these flexibilities are in terms of statistical inference (no parameter estimates and standard errors), added modeling and interpretive efforts, and computational costs.

The AVAS specification assumes that an arbitrary function of the dependent variable, $\theta(N)$, is related to functions of the independent variables, $F_i(S_i)$:

$$\theta(N) = \sum_{i=1}^{n} F_i(S_i) + \varepsilon, \tag{2}$$

where the subscript i refers to the ith explanatory variable. It is assumed that $\theta(N)$ is monotone and strictly increasing, and ε is normally distributed with mean zero, which implies that $\theta(N)$ and $F_i(S_i)$s have a multivariate normal distribution, since by definition $\varepsilon = \theta(N) - \theta\Sigma_i\ F_i\,(S_i\,)$. Finally, ε is assumed independent of the S_is. Both the dependent and the explanatory variables may be categorical or continuous. An additional advantage, as shown by Tibshirani, is that AVAS is a generalization of the Box-Cox transformation. AVAS is, however, superior to Box-Cox in the sense that the transformations it provides are not limited to the logarithmic class of functions (Hastie and Tibshirani, p. 187). In this study, the representation in (2) is used as an approximation to the nutrient demand function in (1). The AVAS algorithm nonparametrically estimates functions $\theta(N)$ and $F_i(S_i)$s for given data. Fortran implementation of AVAS is available from Hastie and Tibshirani. The analysis presented below was conducted using "S-Plus" statistical package (Statistical Sciences, Inc.), which contains a "canned" implementation of AVAS.

Despite its additive structure, the formulation in (2) is quite general: each term can be a simple or complex function and perhaps a function of more than one explanatory variable. Thus $F_k(S_k)$, where $S_k = S_i\,S_j$ and $F_k(\cdot)$ is an unspecified function, may be a term in the model. At first glance, the underlying distributional assumptions may appear stringent. However, the fact that AVAS permits the independent and the explanatory variables to be jointly distributed may be desirable for economic data. As for normality, the reasonableness of this assumption can be tested using standard statistical tests.

The following example demonstrates the steps in interpreting the AVAS output. Suppose that the "true" relationship between nutrient intake, education, and income is

$$N = \gamma\, S_1^{\alpha_1} S_2^{\alpha_2}, \tag{3}$$

where N is total calorie intake, S_js are income and education (j=1,2), γ is the intake for a reference individual in the economy, and α_is are parameters. Reasonable values for these parameters could be $\gamma =$ 1,717 (the mean caloric intake in data described above), $\alpha_1 = 0.02$ for income, and $\alpha_2 = 0.01$ for education.

The relationship in (3) and the associated parameter values are, of course, unknown. The researcher obtains data on intake, income, and education and proceeds to estimate a linear relationship—clearly not a very good approximation if the true relationship is equation (3). The correct linear representation, following the addition of an error term , ε that accounts for potential measurement error in the dependent and the explanatory variables is

$$\log(N) = \log(\gamma) + \sum_{j=1}^{2} \alpha_j \log(S_j) + \varepsilon\ . \tag{4}$$

The AVAS algorithm applied to actual data for N, S_1, and S_2 will yield an estimate of the true underlying functions, $\theta(N) = \log(N)$ and $F_i(S_i) = \alpha_i \log(S_i)$, evaluated at each data point. The output is a vector of values for each function representing the "optimal" transformations obtained by the AVAS algorithm. By plotting each variable against the values generated by AVAS, for example N versus $\theta(N)$ and S_is versus $F_i(S_i)$s, one attempts to identify a reasonable transformation of the variables. Following the selection of a good transformation, we can proceed to obtain estimates of the parameters γ and α_is using OLS. We should note that when the "optimal" transformations are uncovered, the residuals from the AVAS should be orthogonal to the transformation suggested for the explanatory variables, $F_i(S_i)$s. A simple plot is used to verify this assumption. The interpretation of output of the AVAS algorithm is further clarified below.

AVAS and Regression Results

Scatter plots of each intake measure against income and other continuous variables did not reveal a discernible structure. The AVAS method helps elicit potential nonlinearity in the intake-socioeconomic relationship. The results—presented in graphical form below—will be used to specify a parametric model that is estimated through regression analysis. The presentation focuses on the relation between caloric intake, age, income, and education. The findings for other intake measures are similar.[5]

The reported plots show a particular functional relationship conditional on the "optimal" transformation for the dependent and all other explanatory variables. This is similar to multiple regression, where a specific parameter shows the marginal influence of that variable conditional on other explanatory variables. In the case of AVAS, functions replace coefficients. For example, the estimated relationship between intake and income is conditional on the optimal function relating intake to household size. Therefore, it is important to use explanatory variables in their "raw" form rather than arbitrarily transforming them, for instance to per capita values, prior to AVAS analysis.

The analysis was conducted for each intake measure and the set of explanatory variables defined above. Figure 1 presents a three-dimensional surface plot of the relationship between total calorie intake, age, and income while controlling for all other explanatory variables. At first intake smoothly rises with age and income, then it increases at a faster rate in the midrange of both variables, and although somewhat erratic, it becomes essentially flat at the top of the range. The key feature of the data is the peak in the midrange of income and age.

Figure 2 presents plots of the pair-wise relationship between intake and each continuous variable. Panels A and B are graphs of the optimal transformations of age, $F_1(AGE)$, and income, $F_2(INCOME)$, for caloric intake. The transformations suggested by AVAS (the y axis) are "scale-indeterminate": they are unique representations of the "true" relationship up to a scale factor. The AVAS method begins with an arbitrary but reasonable choice of scale by "normalizing" the estimated transformations—mean zero and unit variance. Normalization insures that the outcome of the analysis is not sensitive to changes in the units of measurement. The y axis therefore represents the response of the dependent variable—in units of standard deviations—to changes in the explanatory variables.

Focusing on panel A, the figure shows that total caloric intake initially rises and subsequently falls with income. A similar pattern is found for other intake measures. Panel C shows that for calcium, however, beyond its initial rise, intake is invariant to changes in income. Overall, the AVAS analysis suggests that the intake-income relationship resembles an inverted V. Similar but more extreme nonlinearities have been reported by Strauss and Thomas based on Brazilian data and bivariate analysis of intake and income.

The striking feature common to all intake measures is that the inflection point occurs at the same income level—approximately \$30,000 household income (\$10,000 per capita).[6]

[5]The plots for other intake measures and the programs that generated these output are available from the author upon request.

[6]Using per capita income instead of income and household size indicates a kink in the data at the same absolute income level. However, the intake and per capita income relationship is extremely nonlinear and difficult to interpret. As argued above, since the AVAS method aims to uncover optimal transformations of the data, it is more appropriate to use variables in their level form. Additionally, use of income and household size instead of per capita income is standard to this literature as researchers are interested in the effect of each variable separately.

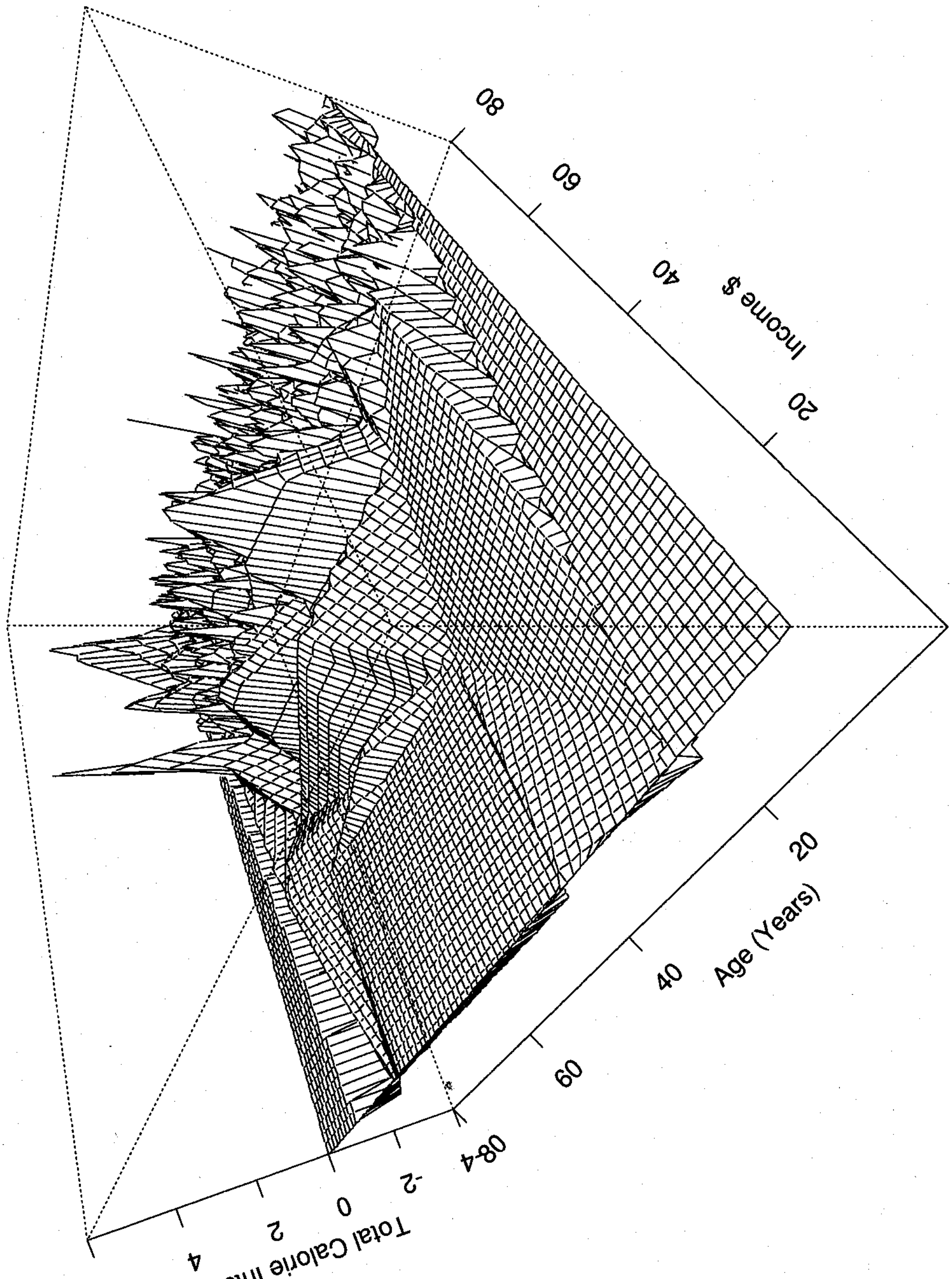

Figure 1. The AVAS surface plot for total calorie intake, age, and income

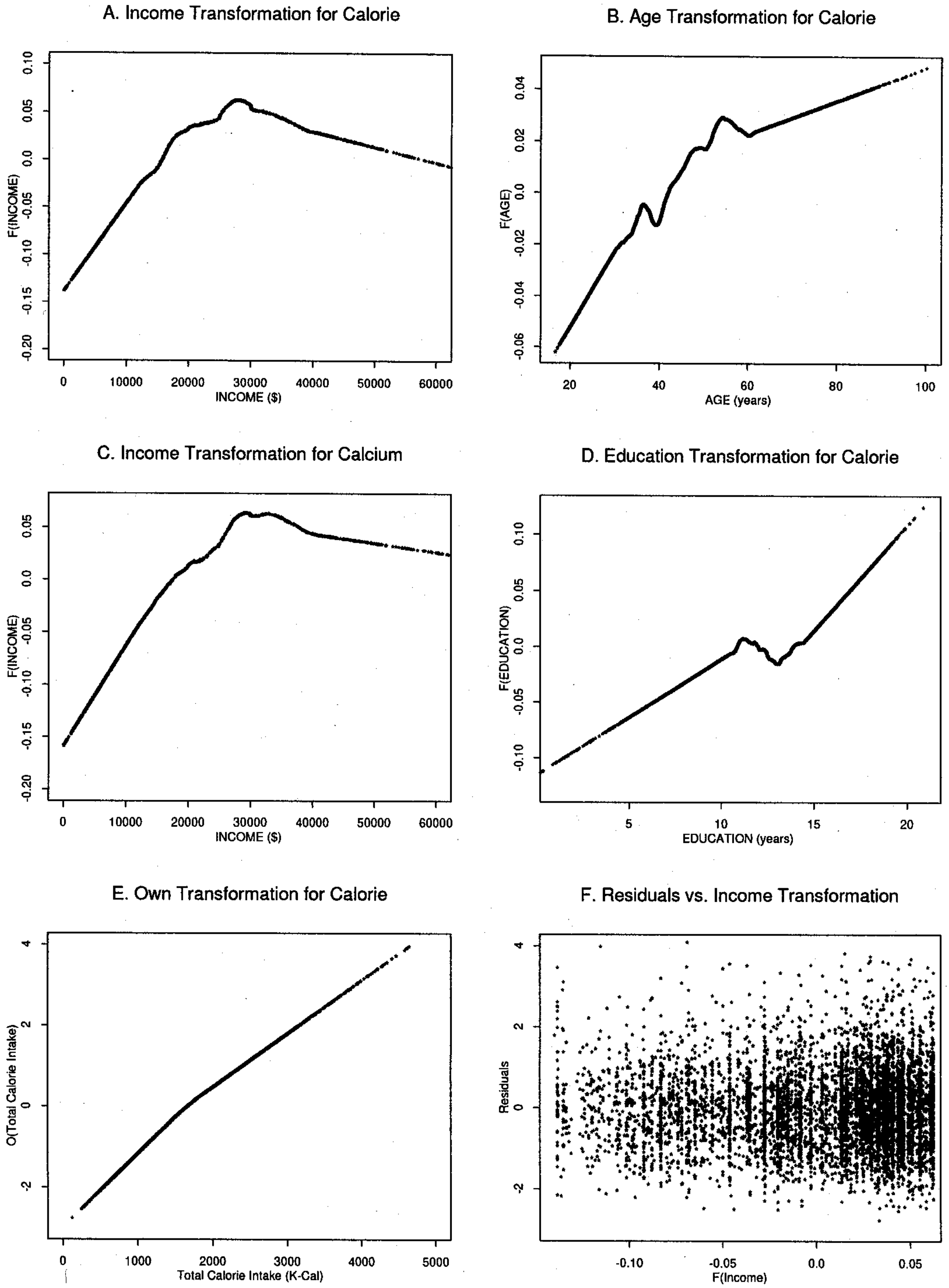

Figure 2. The AVAS transformations for selected nutrient intake measures

Given other explanatory variables and their relationship to intake, it is clear that at the lower income levels nutrient intake is more responsive to changes in income.

Panels B and D depict the optimal transformations for age and education. For age, a logarithmic transformation appears appropriate. This is reasonable since caloric intake at both extremes of age is markedly different from intake of individuals at the mean age. The logarithmic transformation would reduce the extreme values, leading to a more linear relationship between intake and age. Panel B also suggests that in the regression of intake on age, the slope coefficient will be larger for individuals below the mean age (45) than those above.

Panel D shows a similar but reversed effect for education. The response of intake to increased education is small for those below a high school education and higher for those above. This suggests a reinforcing effect of education. Of course, age, education, and income are highly related and interact in a complex manner. The multivariate analysis shows that over the range of the data, interaction among these variables affect intake in opposite directions. This is because with AVAS, unlike linear regression, the influence of each variable on intake could vary over its own range, and the range of other variables.

Panel E provides the means to address the question of whether the dependent variables should be transformed. It plots caloric intake against its optimal transformation $\theta(N)$: the shape is a straight line, indicating that the dependent variable does not require a transformation and that the correct model should be in terms of the intake level.

The final step in the analysis is to plot the model's residuals ($\varepsilon = \theta(N) - \Sigma_i F_i(S_i)$) against the transformations of the explanatory variables, $F_i(S_i)$. Recall that, by assumption, the residuals and each transformation of the explanatory variables should be orthogonal. This means the plot of residuals against the transformations of the explanatory variables should appear as noise. Panel F suggests that overall this assumption is satisfied, though the mean of the residuals seems to fall with income.

To summarize, the AVAS results indicate that for all dietary measures, nonlinearity is an important and persistent feature of the data and that linear regressions would represent a misspecification of this relationship. The most useful aspect of the AVAS analysis is that it can suggest parametric models that best approximate these nonlinearities. Considering the preceding findings, we ask what is a reasonable parametric specification for this data. There are at least three possibilities for modeling the intake-income relationship. First, a quadratic specification in income could generate the relationship depicted in panel A. Such an exercise using the present data leads to statistically insignificant income coefficients, as is the case in other studies. The AVAS plots suggest that a quadratic specification may be inappropriate because, unlike a quadratic curve, the relationship in panel A is essentially linear on both sides of the inflection point.

A second method of accounting for the "kinked" relationship is to use the switching regression techniques that allow for consistent estimation of the coefficients over different ranges of the explanatory variables (Judge et al.). The AVAS is particularly useful for identifying the inflection point specific to any data. A third approach, which is equivalent to the switching regression technique but is easier to interpret, is to partition the sample at the inflection point of income and estimate a linear model for each income group. Again, the partition point is not chosen arbitrarily but is suggested by the data at approximately $30,000.

For simplicity, the third approach is adopted in this study. To correct for other nonlinearities, logarithm of age is used rather than its level and a dichotomous variable is used for education ($EDUC = 1$, if $AVGEDU \leq 12$, and 0 otherwise). The latter specification is inspired by the plot in panel D and is consistent with the usual practice of using educational categories (primary, high school, college, etc.). The analysis leads to a linear model containing income, the logarithm of age, and categorical variables for degree of urbanization, regionality, race, education, marital status, and household size. This model is then estimated using the total data and sample partitions of those below and above $30,000 income.

Table 2 presents the regression results for the total sample and each income partition. The results for each subsample are strikingly different from one another and from the sample as a whole. In all regressions income and other socioeconomic variables become significant when the sample is partitioned. These results are invariant to changing the size of the partition by 5% on both sides of the inflection point.[7]

The consequence of modeling the data in this manner is generally surprising. For instance, the estimated coefficients show that age, which in previous studies had a significant positive effect on intake, is generally only significant at lower incomes. This is also true of variables that measure the influence of urbanization and geographical regions. The subsamples show that spatial effects are uniform across income strata for percentage of calories from fat but quite varied for other nutrients.

Among the explanatory variables considered, income, household size, race, education, and the presence of both household heads are the most significant predictors of all intake measures, and their influence for the lower income group is quite distinct. Income has a consistently significant and positive effect on intake at the lower income levels and is insignificant otherwise. This confirms the functional form suggested by the AVAS analysis.

As income goes up nutrient intake increases, but the increase in caloric intake results from increased consumption of fats. One explanation of this phenomenon is that as income increases, taste, convenience, and other qualitative characteristics of food become important factors in consumer choices resulting in higher consumption of high-fat items, such as meats and processed foods. This suggests that income and food assistance programs should be supplemented with nutrition education.

Race is also generally significantly influential on nutrient intake at lower income levels, while at higher levels it is not. The coefficients suggest that intake is independent of race at high income levels, but at lower incomes, whites have higher intake of all nutrients relative to nonwhites. Even at high income levels, differences between whites and nonwhites become apparent when qualitative measures such as percentage of calories from fats are considered. Note that the regression based on the full sample masks these differences.

The dichotomous education variable (*EDUC*) takes on the value of one when the average education of the household heads (or the actual value in the case of single-head households) is less than 12 years. The estimated coefficient for each income group is consistent with panel D: all else equal, those with less than a high school education have lower caloric intake.

The regression coefficients indicate that household size and the presence of two household heads are important for lower-income groups but insignificant for the higher. Table 2 indicates a significant and substantial effect of household size on intake, providing strong evidence of economies of scale at lower income levels. These estimates, however, may be upwardly biased as fertility and nutrient-intake choice are simultaneously determined (see Behrman and Deolaikar, p. 678). The existence of double-headed households is by far most

[7] A set of diagnostic tests for the influence of multicollinearity were also conducted for each regression. To assess the extent of collinearity among the explanatory variables the procedures suggested in Belsley, Kuh, and Welsch were followed. Overall, the pair-wise correlation among the explanatory variables did not exceed 0.30. The condition number (the ratio of the largest to the smallest eigenvalue) obtained was 20, which falls short of the value suggested by Belsley, Kuh, and Welsch as an indicator of strong collinearity (30).

Table 2. Regression Coefficients for Nutrient Intake by Income Level

Variable Name		Calorie Intake	Calorie % RDA	Fat % RDA	Protein % RDA	Calcium % RDA
INTERCEP	Total Sample	1,400.80**	65.70**	35.00**	114.90**	68.00**
	Low Income	1,230.70**	60.40**	33.70**	108.70**	63.60**
	High Income	1,640.90**	71.00**	38.80**	121.20**	71.10**
AGE (logarithm)	Total Sample	39.30*	1.60**	0.40*	2.60*	0.80
	Low Income	48.60*	2.00*	0.32	2.00	0.81
	High Income	44.90	1.45	0.50	4.20*	1.50
SEX	Total Sample	– 2.60	– 0.50	– 0.3*	–1.00	–1.90*
	Low Income	11.90	0.50	– 0.10	–1.10	– 0.40
	High Income	–21.20	–1.80*	– 0.80**	– 0.60	– 3.80**
CITY	Total Sample	–39.20*	–1.20	–1.00**	– 0.20	1.20
	Low Income	–51.30*	–1.80	– 0.90**	1.00	1.40
	High Income	–13.10	– 0.30	–1.10**	–2.00	1.46
RURAL	Total Sample	8.20	0.80	0.50**	1.80	2.40*
	Low Income	–19.80	– 0.05	0.65**	– 0.30	1.70
	High Income	66.00*	2.30*	0.40	6.70**	5.60**
INCOME	Total Sample	– 0.3E–03	– 0.3E–04**	0.2E–05	– 0.8E–03	0.25E–04
	Low Income	0.5E–02**	0.1E–03**	0.4E–04**	0.3E–03**	0.3E–03**
	High Income	– 0.5E–03	– 0.1E–04	– 0.9E–05	– 0.4E–04	0.1E–05
HHSZ	Total Sample	18.30**	–1.20**	0.16**	0.90**	– 0.80**
	Low Income	25.80**	–1.10**	0.20**	1.60**	–1.60**
	High Income	5.20	–1.20**	0.03	– 0.20	0.10
RACE	Total Sample	86.30**	3.20**	0.40	– 0.60	16.30**
	Low Income	78.40**	3.20**	0.70**	– 0.30	13.00**
	High Income	63.50	2.40	–2.00**	– 0.50	22.50**
NEAST	Total Sample	–17.60	0.30	–1.00**	–1.00	–3.20*
	Low Income	–16.00	1.60	–1.00**	0.50	0.40
	High Income	–41.00	–1.00	–1.00**	–2.60	–6.50**
SOUTH	Total Sample	32.00	1.70*	– 0.90**	0.75	–7.20**
	Low Income	50.60*	1.80	– 0.80**	2.40	–7.00**
	High Income	14.00	1.20	–1.00**	–1.50	–6.80**
WEST	Total Sample	–58.70**	–2.00**	–1.00**	– 4.50**	0.50
	Low Income	–54.00	–1.95	– 0.50	– 4.70**	1.00
	High Income	–64.20*	–2.10	–1.70**	– 4.10	0.30
EDUC	Total Sample	–36.70**	0.70	0.40**	–1.50**	–7.00**
	Low Income	–17.90*	1.50*	0.55**	0.50	–6.00**
	High Income	–58.30**	– 0.90	0.50*	–3.90**	–8.20**
DUOHEAD	Total Sample	102.00**	2.90**	0.60**	4.00**	5.00**
	Low Income	134.90**	4.00**	0.50*	4.40**	8.30**
	High Income	–101.70**	–2.20	– 0.02	–2.70	–8.40**

Note: For a definition of variables see table 1. High (low) income partition of the sample excludes individuals with household before tax annual income below (above) $30,000. The *F*-value for all regressions are significant at 0.01 and higher.

*Indicates significance between 0.05 and 0.10 level.

**Indicates significance at less than 0.05 level.

significant for the lower-income strata, possibly indicating the important influence of time availability, a constraint particularly binding for single-headed households that make up a large portion of the low-income population.

Conclusions

Previous analyses of nutrient intake assign a bewildering and often contradictory impact to various socioeconomic variables. For instance, the influence of income on various measures of intake is found to be positive by some (Adrian and Daniel; Horton and Campbell; Basiotis et al.), negative by others (Murphy et al.), and statistically insignificant in other studies (Davis). This is a likely consequence of model misspecification, as the most prevalent approach is to add squared income terms (Adrian and Daniel; Horton and Campbell; Basiotis et al.) or to take logarithms of continuous variables (Scearce and Jensen).

Other important socioeconomic variables also appear contradictory. Education seems to influence intake negatively (Adrian and Daniel), very little (Horton and Campbell), and positively (Murphy et al.). The effect of household size is the most indeterminate among the studies in this area, including those cited above. Effects of race, age, and regional variables also vary. The AVAS analysis shows that allowing nonlinearity is an important step in removing specification error. The results obtained show that controlling for specification not only improves the explanatory power of the parametric models but provides evidence that the intake-socioeconomic link is stronger than anticipated.

At low-income levels, nutrient intake is highly influenced by a variety of socioeconomic factors, the most significant being income, household size, race, and the presence of two household heads. The implication is that public policies aiming to influence consumption behavior—government transfer programs, nutrition education programs, and other efforts—should be designed to account for differences in behavior by income, household size and composition, race, and education of the poor. At higher-income levels, few variables influence intake. Specification of a model that is appropriate to these data was an important step to uncovering these findings.

[*Received July 1994; final version April 1995*]

References

Adrian, J., and R. Daniel. "Impact of Socioeconomic Factors on Consumption of Selected Food Nutrients in the United States." *Amer. J. Agr. Econ.* 58(1976):31–38.

Altman, N. "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression." *The American Statistician* 46(1992):31–38.

Basiotis, P., S. Johnson, K. Morgan, and J. Chen. "Food Stamps, Food Costs, Nutrient Availability, and Nutrient Intake." *Policy Modeling* 9,3(1987):383–92.

Behrman, J., and A. Doelalikar. "Health and Nutrition." In *Handbook of Development Economics*, eds. H. Chenery and T. Srinivasan. New York: North Holland, 1988.

Behrman, J., A. Doelalikar, and B. Wolfe. "Nutrients: Impacts and Determinants." *World Bank Review* 2(1988):299–20.

Belsley, D., E. Kuh, and R. Welsch. *Regression Diagnostics*. New York: John Wiley and Sons, 1980.

Capps, O., Jr., and J. D. Schmitz. "A Recognition of Health and Nutrition Factors in Food Demand Analysis." *West. J. Agr. Econ.* 16(1991):21–35.

Davis, C. G. "Linkages Between Socioeconomic Characteristics, Food Expenditure Patterns, and Nutritional Status of Low Income Households: A Critical Review." *Amer. J. Agr. Econ.* 68(1982):1017–25.

Hama, M., and H. Riddick. "Nationwide Food Consumption Survey, 1987." *Family Econ. Rev.* 2(1988):24–27.

Härdle, W. *Applied Nonparametric Regression*, New York: Cambridge University Press, 1989.

Hastie, T., and R. Tibshirani. *Generalized Additive Models*. New York: Chapman and Hall, 1990.

Horton, S., and C. Campbell. "Wife's Employment, Food Expenditures, and Apparent Nutrient Intake: Evidence from Canada." *Amer. J. Agr. Econ.* 73(1991):784–94.

Judge, G. G., W. E. Griffiths, R. C. Hill, and T. C. Lee. *The Theory and Practice of Econometrics*. New York: John Wiley and Sons, Inc., 1980.

Mertz, W., J. Tsui, J. Judd, S. Reiser, J. Hallfrisch, E. Morris, P. Steele, and E. Lashley. "What are People Really Eating? The Relation between Energy Intake Derived from Estimated Diet Records and Intake Determined to Maintain Body Weight." *Amer. J. Clinical Nutrition* 54(1991):291–95.

Morgan, K. J. "Socioeconomic Factors Affecting Dietary Status: An Appraisal." *Amer. J. Agr. Econ.* 68,5(1986):1240–246.

Murphy, S. P., D. Rose, M. Hudes, and F. E. Viteri. "Demographic and Economic Factors Associated with Dietary Quality for Adults in the 1987–88 Nationwide Food Consumption Survey." *J. Amer. Dietetic Association* 92,11(1992):1352–357.

Ramezani, C. A. "The Calorie-Income Link in the United States: A Nonparametric Analysis." Work. Pap. Dept. of Consumer Science, University of Wisconsin-Madison, 1993.

Ramezani, C. A., and C. Roeder. "Health Knowledge and Nutritional Adequacy of Female Heads of Households in the United States," *J. of Consumer Affairs* 29,2(1995): 60–73.

Scearce, W., and R. Jensen. " Food Stamp Program Effects on Availability of Food Nutrients for Low Income Families in the Southern Region of the United States." *S. J. Agr. Econ.* 11,2(1979):113–20.

Senauer, B., K. Asp, and J. Kinsey. *Food Trends and the Changing Consumer.* St. Paul MN: Eagan Press, 1991.

Silverman, B., *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall, 1986.

Statistical Sciences, Inc., "S-PLUS," Version 3.2. Computer Software Pkg., Statistical Sciences, Inc., Seattle WA, 1994.

Strauss, J., and D. Thomas. "The Shape of Calorie-Expenditure Curve." Discus. Pap. No. 595, Yale Economic Growth Center, 1990.

Subramanian, S., and A. Deaton. "The Demand for Food and Calories: Further Evidence from India." Work. Pap., Dept. of Agr. and Resour. Econ., University of California, Berkeley, 1992.

Tibshirani, R. "Estimating Optimal Transformations for Regression via Additivity and Variance Stabilization." *J. Amer. Statis. Assoc.* 83(1988):394–405.

U.S. Department of Agriculture. "Nationwide Food Consumption Survey, 1987–88 Individual Intake." Access No. PB90-504044. Springfield VA: National Technical Information Service, 1991.