Nonparametric Functional Form Estimation: Application to Nutrient Demand*

Cyrus A. Ramezani

Department of Consumer Science

University of Wisconsin-Madison

February 19, 1995

*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. Comments from Michael Carter, Helen Jensen, Jean-Paul Chavas, Karen Holden, Barbara Wolfe, two referees, and others improved the paper considerably. Any errors remaining are the author's responsibility.

Nonparametric Functional Form Estimation:
Application to Nutrient Demand

ABSTRACT

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. Regression analysis for income groups below and above the inflection point shows that 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.

Keywords: Nutrient Demand, Nonparametric Methods, Additive Models, AVAS

Nonparametric Functional Form Estimation:

Application to Nutrient Demand

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 & Kinsey 1991). Adrian & Daniel (1976) conducted one of the first comprehensive studies of this relationship. Since then, numerous researchers have revisited this question using newer data and more innovative techniques. Senauer et al. (1991) and Morgan (1986) provide a complete survey and appraisal of this literature.[1] They indicate that while research in this area has progressed considerably, the influence of a number of key socioeconomic variables needs further investigation.

In particular, Morgan (1986) and Davis (1982) 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 1986). 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 (Behramn, Deolalikar & Wolfe 1988). Recent studies by Strauss & Thomas (1990) and Subramanian & Deaton (1992) have shown that the low elasticity estimates are partly due to model misspecification. Measurement error may also cause the differences in these findings.

The hypothesis pursued in this study is that the variety of findings reported based on U.S. data is also a result of model 'misspecification'. The aim is to conduct a careful examination of this conjecture using recent data and new methods of functional form estimation. Previous application of nonparametric procedures in this context has focused on the interaction between calorie intake and income, and has been primarily bivariate in scope. A "kinked" relationship between calorie intake and income was identified using the Locally Weighted Regressions (Ramezani 1993, Strauss & Thomas 1990) and Kernel estimators (Subramanian & Deaton 1992). These studies, however, fail to control for the influence of other socioeconomic variables. The estimation procedures we utilize circumvents this problem: 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. Using U.S. data offers the advantage of larger samples with more socioeconomic variables, more intake measures, and less severe 'measurement error'.

Model and Data

Household decision theory suggests that nutrient demand N is related to a vector of demographic and economic variables, $S_1, ..., S_n$:

$$N = f(S_1, \ldots, S_n) + \epsilon \tag{1}$$

where ϵ is an error term added for estimation purposes and the distributional assumptions about ϵ depend upon the estimation methods utilized (Behrman & Deolalikar 1988). Before estimating equation (1), researchers must select the socioeconomic variables, nutrient intake measures, and a functional form for this relationship. Policy interest has an important bearing on the choice of intake measure and the socioeconomic variables (Morgan 1986). The choice of functional form, however, has been generally arbitrary. The standard practice is to assume a linear form and add quadratic (income) terms, or to use the double-logarithmic specification. This study uses the variables commonly utilized in the literature to provide nonparametric estimate of the *multivariate* function represented by equation (1).

The individual portion of the Nationwide Food Consumption Survey (NFCS 1987-88) is used for this study. This comprehensive 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 & Riddick 1988). 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 utilized here consists of 5870 individuals. The data excludes pregnant women, lactating mothers, and individuals less than 19 years of age. The nutritional needs of these individuals are markedly different than the rest of the sample (Murphy, et al. 1992). 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 were available in this sample. Capps, Jr & Schmitz (1991) 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 utilized: total calorie intake; calorie, protein, and calcium intake as percent of the recently revised individual specific RDA; and percent of calorie intake from fats. These measures correct for differences in dietary quality and requirements across individuals. Previous studies suggest that these measures are useful for assessing diet quality (Murphy, Rose, Hudes & Viteri 1992).

Table 1 provides definition and basic statistics on the dependent and the explanatory variables considered.[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 calorie 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 paper is that the influence of most socioeconomic variables will be dissimilar 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 under consideration. Consumer and production theory provide several examples. Nutrient demand functions – equation (1) – present a case for potential nonlinear interaction between intake, age, and income. The specification of a functional form is not a straightforward decision and, as is evident in applied literature, different functional forms are selected on the basis of their tractability rather than a priori knowledge of the true functional relation. The usual practice is to take the data as given and to impose a structure that is sufficiently general, as with flexible functional forms, or to search for the appropriate structure within a narrow class of specifications, as with the Box-Cox transformation. These methods impose structure upon the data. An alternative approach, advanced by nonparametric techniques, is to let the data determine the best specification.[4]

This section introduces a specific nonparametric curve estimation technique – the Additivity and Variance Stabilization (AVAS) method (Tibshirani 1988) – that can prove useful in the analysis of agricultural economic data. Given space limitations, the presentation will be brief, emphasizing the intuition and usefulness of AVAS rather than technical details, for which the interested readers should consult Tibshirani (1988). The proceeding discussion aims to elucidate the potential utility of these methods in uncovering nonlinearities in the data. As shown, AVAS is useful for specifying a parametric model that is simple to interpret.

The AVAS belongs to a broad class of functional specifications called the Generalized Additive Models (GAM) (Hastie & Tibshirani 1990). The additive specification and AVAS are well suited to the analysis of economic data because they allow estimation of *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, added modeling and interpretive efforts, and computational costs of implementing these techniques.

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

$$\Theta(N) = \sum_{i=1}^{n} F_i(S_i) + \epsilon \tag{2}$$

where the subscript i refers to the i-th explanatory variable, and it is assumed that $\Theta(N)$ is monotone and strictly increasing, $F_i(S_i)$'s have a multivariate normal distribution, ϵ is normally distributed with mean zero, and ϵ is independent of S_i's. Both the dependent and the explanatory variables may be categorical or continuous. An additional advantage, as Tibshirani (1988) has shown, 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 & Tibshirani 1990, page 187). In this study, the representation in (2) is used as an approximation to nutrient demand function in equation (1). The AVAS algorithm provides nonparametric estimates of the *functions* $\Theta(N)$ and $F_i(S_i)$'s for a given data.

Despite its additive structure, the formulation in equation (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 $F_k(\ . \ )$ is an unspecified function, may be a term in the model. At first glance, the underlying distributional assumptions may appear stringent. However, when viewed in the context of economic data, these assumptions may indeed be appropriate. For example, AVAS permits the explanatory variables to be jointly distributed, which is desirable in the case of most economic data. As for normality, the reasonableness of this assumption may only be assessed given the size and the context of the specific data analyzed.

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

(3) $$N_j = \gamma S_{1j}^{\alpha_1} S_{2j}^{\alpha_2}$$

where N_j is total calorie intake, S_{ij}'s are income and education for individual j, γ is the intake for a reference individual in the economy and α_i's are parameters. Reasonable values for these parameters could be $\gamma = 1770$ (the mean calorie intake in data described below), $\alpha_1 = .02$ for income, and $\alpha_2 = .01$ 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:

(4) $$Log(N) = Log(\gamma) + \sum_{i=1}^{2} \alpha_i * log(S_i) + \epsilon$$

The AVAS algorithm applied to actual data for N, S_1, and S_2 will yield an estimate (denoted by $\hat{}$) of the true underlying functions, $\hat{\Theta}(N) = log(N)$ and $\hat{F}_i(S_i) = \alpha_i * log(S_i)$'s, 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, i.e., N versus $\hat{\Theta}(N)$ and S_i's versus $\hat{F}_i(S_i)$'s, one attemps 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_i's using Ordinary Least Squares (OLS). These same steps are taken in the proceeding analysis. We should note that when the 'optimal' transformation are in fact uncovered, the residuals from the AVAS should be orthogonal to the transformation suggested for the explanatory variables, $\hat{F}_i(S_i)$'s. A simple plot is used to verify this assumption. The interpretation of output of the AVAS algorithm are further clarified below.

AVAS and Regression Results

Scatter plots of each intake measure against income and other continuous variables do 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 via regression analysis. To save space the presentation focuses on the relation between calorie intake, age, income, and education. The findings for other intake measures are similar.[5]

It is important to note that 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 pre-transform them – e.g., per capita values.

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. It is apparent that at first intake smoothly rises with age and income, then it increases at a faster rate in the mid-range 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 mid-range of income and age.

Figure 2 presents plots of the pairwise relationship between intake and each continuous variable. Panels A and B are graphs of the optimal transformations of age, $\hat{F}_A(age)$, and income, $\hat{F}_I(income)$, for calorie intake. The transformations suggested by AVAS (the y-axis) are "scale-indeterminate": they are unique representation 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 calorie intake initially rises and subsequently falls with income. 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 & Thomas (1990) 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] 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. This finding is robust in the sense that the relationship which emerges is less likely to be the result of misspecification because AVAS permits for large classes of functional forms.

Panels B and D depict the optimal transformations for age and education. For age, it appears that a logarithmic transformation is most appropriate. This is reasonable since calorie intake at both extremes of age is markedly different than 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 of small magnitude 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 could effects intake in an offsetting manner. This is because under AVAS specification, unlike linear regression, the influence of each variable on intake could vary over its own range.

Panel E provides the means to address the question of whether the dependent variables should be transformed. It plots calorie intake against its transformation $\hat{\Theta}(N)$. The shape is a straight line, indicating that the dependent variable does not require a transformation and that the correct model should describe the level of intake.

The final step in the analysis is to plot the model's residuals ($\hat{\epsilon} = \hat{\Theta}(N) - \sum_i \hat{F}_i(S_i)$) against the transformations of the explanatory variables, $\hat{F}_i(S_i)$. Recall that, by assumption, the residuals and each transformation of the explanatory variables should be orthogonal. This would mean the plot of residuals against the transformations of the explanatory variables would appear as noise. Panel F suggests that overall this assumption is satisfied, though the mean of 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 provides a means for specifying 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 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 quadratic specification may be inappropriate since the relationship is essentially linear on both sides of the inflection point.

A second method for 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, Griffiths, Hill & Lee 1980). The AVAS is particularly useful in this context since it provides a means for *identifying* the inflection point. 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 AVAS.

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 variables is used for education (EDUC=1 if AVGEDU $\leq$ 12 and zero otherwise). The latter specification is inspired by plot in panel D and is consistent with the usual practice of using educational categories (primary, high school, college, etc.).

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 robust 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 income level. Partitioning the sample has similar consequence for the influence of the degree of urbanization and geographical regions, which capture the influence of the local cultures, variation in food characteristics, and other food supply conditions on intake. The subsamples show that spatial effects are uniform across income strata for percent 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.

While increases in income positively influence intake of the nutrients considered, we find the surprising result that increases in calorie intake result from increased consumption of fats, i.e., a decline in the quality of diet in terms of increased share of fats in the diet is seen initially as income rises. One explanation of this phenomenon is that as income increases, taste and other qualitative characteristics of food become important factors in consumers' choice resulting in higher consumption of high fat luxury 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 significant at lower income levels while at higher levels it does not influence intake. 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 non-whites. Even at high income levels, differences among whites and non-whites become apparent when qualitative measures such as percent of calories from fats are considered. Note that the regression based on the full sample masks these differences.

The dichotomous education variable takes on the value of one when the average education of the household heads (or the actual value in the case of single 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 calorie intake.

The regression coefficients indicate that household size and the presence of both household heads are important for the poor but insignificant for the higher income groups. These variables, particularly household size, are reflective of 'returns to scale', household socialization, and household time resources that could be devoted to food purchase and preparation activities. Table 2 indicates a significant and substantial effect of household size on intake, providing strong evidence of economies of scale at lower income level. These estimates, however, may be upwardly biased as fertility and nutrient intake choice are simultaneously determined (see Behrman & Deolalikar (1988), p. 678). The presence of both household heads is by far the most significant for the poor, 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. This variable also captures other differences attributable to households where both heads are present.

To summarize, correcting for specification considerably improves the analysis of determinants of nutrient demand. The analysis, however, can be further improved in a number of directions, most notably by including omitted variables. For example, economic theory suggests that nutrient demand functions include prices as their argument. The individual portion of the NFCS data, however, does not include such information. Other variables that are necessarily omitted from this and similar analyses include individual-specific factors such as metabolism and activity level as well as factors that determine the distribution of food among household members.

Despite these limitations, the ‘positive’ aspect of this analysis is to provide insights on factors that influence nutrient intake, a question that can be primarily addressed by empirical investigation. In this respect, the analysis has improved upon previous studies by controlling for the most probable difficulty – the specification choice. Care in specification has sharpened the ‘normative’ consequences of the analysis. In particular, this analysis provides strong evidence that even in relatively wealthy countries such as the United States, socioeconomic factors are important determinants of nutritional adequacy, particularly for the poor. For this group, income transfers could lead to improved diets. Such improvements could be further enhanced if welfare programs strengthen their educational component.

Conclusions

Previous analyses of determinant 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 & Daniel 1976, Horton & Campbell 1991, Basiotis, Johnson, Morgan & Chen 1987), negative by others (Murphy et al. 1992), and statistically insignificant in many studies. This is a likely consequence of model misspecification, as the most prevalent approach is to add squared income terms (Adrian & Daniel 1976, Horton & Campbell 1991, Basiotis et al. 1987) or to take logarithms of various continuous variables (Scearce & Jensen 1979).

Other important socioeconomic variables also appear contradictory. Education seems to influence intake negatively (Adrian & Daniel 1976), very little (Horton & Campbell 1991), and positively (Murphy et al. 1992). The effect of household size is the most indeterminate among the studies in this area, including those cited above. Race, age, and regional variables fluctuate considerably as well. The AVAS analysis shows that correcting for nonlinearity is an important step in removing specification error. The results obtained are important because they 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.

Moreover, the analysis of the NFCS data indicates that 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 both spouses. The implication is that public policies aiming to influence consumption behavior – government transfer programs, nutrition education programs, and other efforts – should be designed in such way to account for difference in behavior by income, household size, race, education, and the marital status of the poor. At higher income levels, very few variables influence intake. Neglecting to correct for specification error would have precluded many of these findings.

Notes

[1] Space limitation does not permit even a partial listing of the extensive research in this area. The reference section of Senauer et al. (1991) provides a complete overview. For developing countries, nutrient demand, particularly for calories, has been widely studied (Behrman & Deolalikar 1988).

[2] The mean calorie intake as percent of the RDA reported in the table appears to be 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. (1991) 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] Recently popularized Kernel estimators are developed in detail by Härdle (1989) and Silverman (1986) and are nicely surveyed by Altman (1992).

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

[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 & Welsch (1980) were followed. Overall, the pairwise 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 & Welsch as an indicator of strong collinearity (30). It is therefore unlikely that collinearity will influence the validity of the reported estimates.

Table 1:

Table 2:

References

Adrian, J. & Daniel, R. (1976), 'Impact of socioeconomic factors on consumption of selected food nutrients in the United States', *American Journal of Agricultural Economics* **58(1)**, 31–38.

Altman, N. (1992), 'An introduction to Kernel and nearest-neighbor nonparametric regression', *The American Statistician* **46**, 175–186.

Basiotis, P., Johnson, S., Morgan, K. & Chen, J. (1987), 'Food stamps, food costs, nutrient availability, and nutrient intake', *Journal Policy Modeling* **9(3)**, 383–392.

Behramn, J., Deolalikar, A. & Wolfe, B. (1988), 'Nutrients: Impacts and determinants', *World Bank Review* **2**, 299–320.

Behrman, J. & Deolalikar, A. (1988), Health and nutrition, *in* Chenery & Srinivasan, eds, 'Handbook of Development Economics', North Holland.

Belsley, D., Kuh, E. & Welsch, R. (1980), 'Regression diagnostics', *Wiley, New york*.

Capps, Jr, O. & Schmitz, J. D. (1991), 'A recognition of health and nutrition factors in food demand analysis', *Western Journal of Agricultural Economics* **16**, 21–35.

Davis, C. G. (1982), 'Linkages between socioeconomic characteristics, food expenditure patterns, and nutritional status of low income households: A critical review', *American Journal of Agricultural Economics* **68**, 1017–1025.

Hama, M. & Riddick, H. (1988), 'Nationwide food consumption survey, 1987', *Family Economic Review* **2**, 24–27.

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

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

Horton, S. & Campbell, C. (1991), 'Wife's employment, food expenditures, and apparent nutrient intake: Evidence from canada', *American Journal of Agricultural Economics* **73**, 784–794.

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

Mertz, W., Tsui, J., Judd, J., Reiser, S., Hallfrisch, J., Morris, E., Steele, P. & Lashley, E. (1991), 'What are people really eating? The relation between energy intake derived from estimated diet records and intake determined to maintain body weight', *American Journal of Clinical Nutrition* **54**, 291–295.

Morgan, K. J. (1986), 'Socioeconomic factors affecting dietary status: An appraisal', *American Journal of Agricultural Economics* **68(5)**, 1240–1246.

Murphy, S. P., Rose, D., Hudes, M. & Viteri, F. E. (1992), 'Demographic and economic factors associated with dietary quality for adults in the 1987-88 Nationwide Food Consumption Survey', *Journal of the American Dietetic Association* **92(11)**, 1352–1357.

Ramezani, C. A. (1993), 'The calorie-income link in the united states: A nonparametric analysis', Working Paper*, Dept. of Consumer Science, Univ. of Wisconsin-Madison.*

Scearce, W. & Jensen, R. (1979), 'Food stamp program effects on availability of food nutrients for low income families in the Southern region of the United States', *Southern Journal of Agricultural Economics* **11(2)**, 113–120.

Senauer, B., Asp, K. & Kinsey, J. (1991), *Food Trends and the Changing Consumer*, Eagan Press, St. Paul, Minnesota.

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

Strauss, J. & Thomas, D. (1990), 'The shape of calorie-expenditure curve', Discussion Paper No. 595*, Yale Economic Growth Center.*

Subramanian, S. & Deaton, A. (1992), 'The demand for food and calories: Further evidence from india', Working Paper*, Department of Agricultural and Resource Economics, University of California, Berkeley.*

Tibshirani, R. (1988), 'Estimating optimal treansformations for regression via additivity and variance stabilization', *Journal of American Statistical Association* **83**, 394–405.

Figure 1: Surface Generated by AVAS: Calorie Intake, Age, and Income

Figure 2: The AVAS Transformations for Selected Nutrients