

The Calorie-Income Link in the United States*

CYRUS A. RAMEZANI

Department of Consumer Science

University of Wisconsin, Madison

First draft: August 1991

Last revised: June 18, 1993

*Earlier versions of this paper were presented at the American Agricultural Economic Association's meetings and seminars at the University of California-Berkeley and the Stanford University. Comments provided by J. Chalfant, F. Viteri, S. Lane, S. Murphy, J. Csete and others improved the paper considerably. I alone are responsible for any errors. Financial support from the U.S. Department of Agriculture (CA #58-3198-9-064) is gratefully acknowledged.

The Calorie-Income Link in the United States

Abstract

This paper is concerned with the relationship between nutrient intake and socioeconomic characteristics of individuals in the United States. Two measures of intake are utilized; caloric intake as percent of the Recommended Dietary Allowance (measures underconsumptions) and percent of caloric intake from fats (accounts for differences in quality of diets). Regression analysis and two nonparametric methods are used to study these relationships. The nonparametric techniques indicate a nonlinear shape – resembling an inverted ‘U’ – where intake rises and then falls with income. Regression analysis based on total sample would lead to erroneous estimates of income elasticity of caloric intake. Accounting for nonlinearity leads to larger elasticity estimates at lower incomes.

Keywords: Nutrient Demand, Caloric Intake, Nonparametric Methods.

The Calorie-Income Link in the United States

Increasing affluence, rising costs of health care, and the overall aging of the population have raised health consciousness and concern for better nutrition in the United States. This prevalent concern has focused attention on the qualitative characteristics of foods and has resulted in an extensive public policy debate on how to influence consumer behavior and improve diets (National Research Council 1989*a*, Institute of Medicine 1991).

In light of these developments, researchers have sought to establish a link between the nutritional well-being of individuals and their socioeconomic status.¹ Indeed, information about the interaction between income, education, living arrangements, the degree of urbanization, and the dietary intake of individuals is considered fundamental to the design and implementation of effective public health policies (Institute of Medicine 1991).

For developing countries, nutrient demand, particularly for calories, has been widely studied (Behrman & Deolalikar (1988), Alderman (1989), and Behrman, Deolalikar & Wolfe (1988)). Surprisingly, however, less is known about the nature of such relationships in the developed countries. The purpose of this paper is to partially fill this gap. Specifically, I investigate the interaction between caloric intake and income, race, age, education, household size, marital status, and degree of urbanization of a sample of the U.S. adult population. A comparison of these findings with those from the developing countries may provide insights into the impact of economic development on human nutrition.

Because caloric intake is highly correlated with most dietary components of foods, it is believed to be a good measure of overall nutritional quality. Evidence suggests that diets which achieve the recommended energy intake are likely to meet, or even

exceed, the recommended guidelines for other nutrients (Pao & Mickle 1981). For the sample population studied here, Murphy, Rose, Hudes & Viteri (1992) conclude that energy intake is a better predictor of the nutritional adequacy of the U.S. adult diet than a variety of socioeconomic variables examined. Their analysis shows that energy intake alone can predict approximately 50% of the variation in the number of nutrients in individual diets falling below 67% of Recommended Dietary Allowance (RDA). This is the rationale for our focus on caloric intake as the key nutrient. ²

The calorie-income Engel curve I investigate has been widely studied since the early 1970s (Behrman & Deolalikar 1988, Alderman 1989, Deaton & Muellbauer 1980). The interaction between income and nutrient intake is of particular importance for alleviating malnutrition (Timmer, Falcon & Pearson 1983, Timmer & Alderman 1979).

³ The conventional wisdom on the calorie-income link suggests that at lower income levels, energy intake rises with income. In the middle range of income levels, the response is likely to be flat, and then it may be reversed at higher income levels.⁴ One explanation of this relationship may be that with increases in income households switch to higher-value food items, which may not necessarily be of better nutritional quality (Behrman & Deolalikar 1987, Behrman & Deolalikar 1988).

Despite some empirical evidence in support of this hypothesis, questions regarding its plausibility have been raised. Some have suggested that even among the very poor, income increases may influence the demand for taste only, rather than calories and nutritional quality (Behrman & Deolalikar 1987). This suggests that the caloric-income relationship may be generally flat and at best mildly increasing. The validity of this conjecture has important policy implications, particularly since considerable resources (domestically and internationally) are devoted to alleviating undernutrition problems through income transfers.

Other economic variables, e.g., education, urbanization, etc., have provided similar points of contention, though to a lesser extent (Behramn & Wolfe 1984, Behrman & Deolalikar 1988). The majority of these have often been studied individually and this may influence the results of the analysis. Our aim is to investigate the *combined* effects of these variables.

Economists have long recognized that the detected relationships among economic variables may be an artifact of the specified functional form or the statistical methods utilized. In the present context, the choice of functional specification may be critical in linking caloric intake and economic status (Strauss & Thomas 1990). Nonparametric methods provide a useful tool to model complex data in a flexible manner and therefore reduce the uncertainty associated with model specification. I rely on two new techniques which require minimal distributional assumptions and permit exploratory analysis of this type of data (Efron & Tibshirani 1991). As in other applications, our study confirms the usefulness of these techniques for uncovering the relationship among economic variables.

Nonparametric Analyses of Calorie Intake

In this and other similar data, as much as 80% of the variation in caloric intake can be explained by intake of other nutrients.⁵ The remainder of the variation in intake has been linked to socioeconomic characteristics using linear specifications. This parametric approach to estimating the relation between caloric intake and socioeconomic variables has the drawback that a functional form, usually the double-log, must be specified in advance. 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. Since this choice is arbitrary, it has a high probability of being incorrect. Nonparametric methods offer a way to avoid this problem. This is particularly true when no stringent distributional assumptions are invoked.

I utilize the Alternating Conditional Expectations (ACE) method of Breiman & Friedman (1985) and the Locally Weighted Regression (LWR) technique of Cleveland (1979), and Cleveland & Devlin (1988). Both methods aim to ascertain the best functional relation between the explanatory variables and the response variable. Both are suited for exploratory analysis since they provide graphical aid in determining the appropriate parametric specification and subsequent estimation. Applications in other economic studies indicate that the plots generated by these methods are highly suggestive of the functional forms present in the data (Berck & Chalfant (1990), Meese & Rose (1991)).

I hypothesize that the caloric intake n is related to a vector of demographic and economic variables, e :

$$n = f (e) + \epsilon \tag{1}$$

where ϵ is the error term and the distributional assumptions about ϵ depend upon the nonparametric methods utilized. The only specification involved is the choice of

the explanatory and the explained variables.

The ACE algorithm finds the optimal transformations (i.e. functions) $\Phi(n)$ and $\phi_i(e_i)$ that minimize the unexplained variation in $\Phi(n)$ and produces the maximal correlation for an additive function of the form:

$$\Phi(n) = \sum_i^n \phi_i(e_i) + \epsilon \quad (2)$$

where the subscript i refers to the i th explanatory variable and ϵ is an error term with zero expectation. The statistical procedures utilized to arrive at estimates of $\Phi(n)$ and $\phi_i(e_i)$ are very general and can help determine any functional relationship.⁶ The cost of gaining this generality is that the methodology yields only a graphical representation rather than parameter estimates.

A feature of ACE, which is particularly useful in the context of economic data, is that, unlike standard regression models, ACE does not treat stochastic predictors as if they were fixed and independent. Indeed the algorithm is based on the assumption that the response and the explanatory variables have a multivariate distribution. Therefore ACE formally permits for simultaneity and multicollinearity in the data. This is a desirable property in relating caloric intake to correlated socioeconomic variables.

The Locally Weighted Regression (LWR) is also a very general technique for estimating regression surfaces. Considering equation (1), suppose elements of e are weakly exogenous variables, $f(\cdot)$ is smooth and ϵ is distributed *i.i.d* normal with mean zero and finite variance. To obtain a best estimate of the pairwise functional relation of n and any e_i , the LWR algorithm operates in the same manner as a moving average, where a pre-specified portion of the data nearest (in a Euclidian sense) to each observation is used to estimate a locally weighted regression line, where closer observations receive higher weights. Plots of the smoothed regression surface may

then be used as an aid in determining the appropriate form. In general the larger the fraction of data used to fit the regression surface the smoother the resulting plots.

In our analysis, regression and standard diagnostic tests are conducted first. The results from the application of ACE and LWR are evaluated relative to those from multiple regression. The LWR analysis is used to investigate the pairwise relationship between the measures of caloric intake and the income indicators alone. The ACE methodology provides a multivariate test of the functional relationship. Hence, ACE enables us to assess the relation between the intake and income while controlling for other socioeconomic variables.

Data and Results

The energy-socioeconomic relationship was explored using the individual portion of the Nationwide Food Consumption Survey (NFCS 1987-88). 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 and seasons of the year (Peterkin, Rizek & Tippett 1988, Hama & Riddick 1988).

The sample utilized here consists of 5868 non-pregnant, non-lactating adults aged 19 years and older. The basic socioeconomic variables considered are age, gender, marital status, race, education, urbanization, regional location, household size, and income. In addition to the absolute income level, per capita income (income deflated by household size) and income relative to the federal poverty levels based on household size (U.S. Bureau of the Census 1989) were also considered.

Two measures of energy intake are utilized: caloric intake as percent of the recently revised Recommended Dietary Allowance for energy (National Research Council 1989*b*), and percent of caloric intake from fat. The first measure corrects for differences in requirements across individuals; low caloric intake is an indicator of underconsumption. The second measure is an indicator of diet quality; a high percent of calories from fat is undesirable. Table 1 provides definitions and descriptive statistics for all variables utilized.

The mean energy 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. Additionally,

since the RDA of energy is a relatively large multiple of the absolute minimum of individual needs, the reported mean does not imply that individuals are at risk.

I begin the discussion of our findings with the results of regression analysis. Table 2 contains the parameter estimates for measures of caloric intake, for the total sample and by sex. Using the R-squared as a measure of association, the linear relationship between caloric intake and economic status does not appear to be significant. However, the use of R-squared is particularly misleading in this case because a large portion of variation in caloric intake is due to intake of other nutrients and environmental factors, which are absent from this regression. This implies that socioeconomic variables may at best explain a small fraction of variation in caloric intake, a point which has important bearing on the interpretation of the results presented below.

Most socioeconomic variables are statistically significant and have plausible signs; the percent of RDA energy rises with age and education and falls with income. The proportion of energy derived from fats declines with education but rises with income. Household size adversely affects both measures. Region and degree of urbanization are significant predictors of intake. All else the same, the rural population appears to meet a greater percent of the RDA energy, though a greater proportion derives from fat. Non-whites and single-headed households are at lower levels for both intake measures. Finally, the determinants of caloric intake vary by gender. Overall, fewer variables are significant for males, indicating that the data generating process may be distinct for each sex.

The relationship of both intake measures to income seems precarious and counter to prior beliefs. Caloric intake as percent of RDA falls with income while percent of calories from fat rises with income. Moreover, income, while statistically significant for the first measure, is insignificant for the second.

In light of this finding, I explore whether (a) these unexpected results may be due to multicollinearity, and (b) the likelihood that the relationship between caloric intake and income may be nonlinear, as evidence from the LDCs suggests (Behrman & Deolalikar 1988, Strauss & Thomas 1990).⁷

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 explanatory variables did not exceed 0.3. 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). In the present context, although multicollinearity is present, we believe its effects are minimal. As noted earlier, the ACE technique offers the advantage that the explanatory variables are not assumed to be linearly independent.

Scatter plots of income against the intake measures reveal no discernible patterns. In fact, the data are uniformly distributed over the range of values of both variables.⁸ The nonparametric methods employed will help assess whether the energy-income relation is nonlinear and, if so, whether the widely used elasticity form (the double-log) is an adequate representation of its shape. This latter specification has been widely used to obtain estimates of income elasticity of nutrient intake (Behrman & Deolalikar 1988). This arbitrary specification may be one reason for the varying, and often low, elasticity estimates reported in the literature (see Strauss & Thomas (1990) for a comparison of reported estimates).

Figure 1 contains plots of the relationship between energy intake (as percent of RDA) and three measures of household purchasing power: income, per capita income (PCINC), and income as percent of federal poverty levels (PCTPOV). The first column of graphs are the plots generated by the LWR technique while the second

column are those from ACE. The LWR depicts the pairwise relationship while the ACE accounts for the influence of other explanatory variables in addition to income. Both techniques utilize the total sample.⁹

Focusing first on the LWR generated plots, it appears that energy intake initially rises and subsequently falls with all income measures. The plots provide evidence of a nonlinear relationship, similar but more pronounced than those reported by Strauss & Thomas (1990).¹⁰ This same shape is present when the dietary variable considered is the percent of calories from fat.¹¹ Moreover, the inflection point (the ‘kink’) occurs at the same income levels as the RDAENER. Overall, for both dietary measures, nonlinearity appears to be an important feature which is not captured by regression analysis. To my knowledge, these results represent the first finding of this kind for the United States.¹²

The LWR plots suggest that although caloric intake may rise with income, the magnitude of this increase may be small. One possible explanation for this phenomenon holds that with increases in income, tastes rather than high caloric value play a greater role in individual’s consumption decisions (Behrman & Deolalikar 1987, Behrman & Deolalikar 1988). This explanation seems plausible for the present data as few individuals in the sample are nutritionally at risk.

The second column of Figure 1 provides the graphical results for multivariate ACE analysis. Each plot is based on the same set of explanatory variables as in Table 2, but the income measure utilized varies. These plots of the income measures against their optimally transformed values are suggestive of the shape of the relevant functions, $\phi_i(\cdot)$. Since the ACE algorithm normalizes the coefficients of the transformed regressions to unity (hence the scale on y-axis), the slopes of these plots reveal the response of intake measures (i.e., the sign of the coefficient) over the range of income.

Note that a graph with a slope of one indicates that the variable in question requires no transformation. This was true of both intake measures used.

The ACE plots also indicate a nonlinear relationship between income and caloric intake, which is consistent with the LWR results. The multivariate ACE analysis, however, indicates a greater degree of nonlinearity relative to the LWR bivariate analysis. Both techniques suggest that misspecification may result in biased estimate of the income elasticity of caloric intake, which is an important tool for food policy analysis.

Are there transformations of the response and the explanatory variables that would lead to an improved linear relationship? The ACE plots suggest that the income measures should be transformed.¹³ A transformation that has been popularized by its ease of interpretation is the double-log form. I applied this transformation to the continuous variables and used the LWR and the ACE to assess whether such a transformation removes the nonlinearities in the data. The inflection point is again a persistent feature of the data, suggesting that elasticity estimates obtained from logarithmic transformation are likely to be biased.

As a final step, I partitioned the sample into households with incomes below and above \$25,000, which is the point of inflection indicated by the nonparametric techniques. The regression results for these subsamples are collected in Table 3. The sign of income coefficient is different for each group and is consistent with the nonparametric analysis. At the lower income levels (N=2917), a larger number of variables are statistically significant. For this group, with an average per capita income of \$6228, regional effects, degree of urbanization, marital status, and education are the most (statistically) significant factors influencing intake.

For individuals with household incomes in excess of \$25,000 (\$16,550 per capita),

degree of urbanization and regional effects dominate other variables in their significance. Interestingly, however, marital status and education are insignificant for this group.

The estimates of the income elasticity of caloric intake – calculated at the means of income and intake – are -0.02 (high income), -0.014 (total sample), and 0.01 (low income). Although these estimates are not statistically significant, they indicate that at lower income levels the elasticity may be positive.

Conclusions

This analysis of the NFCS data indicates that individuals' caloric intake in the United States is highly influenced by a variety of socioeconomic factors. The nonparametric analysis suggests that intake first rises and subsequently falls with income. Although regression analysis appears to confirm this hypothesis, the coefficients and the estimated income elasticity of caloric intake are both small in magnitude and statistically insignificant. However, this analysis shows that this elasticity may be positive at low income levels. This question could be resolved by conducting similar analyses for a sample of low-income individuals.

Other socioeconomic variables are found to be significant predictors of intake. Among these, regional effects and degree of urbanization are significant at all income levels. This analysis indicates that at low income levels, education and marital status are among the most significant factors affecting intake and offer the most important tools for public policy.

Notes

¹ Household production and nutrient demand models are surveyed in Behrman & Deolalikar (1988). The ability to obtain nutrients from foods, i.e., nutrient production, should be distinguished from the desire or the demand for nutrients. The latter is the focus of this and many other studies. Fewer applied studies have focused on nutrient production functions.

² The same justifications are provided in studies of the developing countries (Behrman & Deolalikar 1988, Alderman 1989, Behrman, Deolalikar & Wolfe 1988).

³ A distinction has been made between income and food expenditures as measures of purchasing power (Strauss & Thomas 1990). Although these measures are highly correlated, they are influenced by different factors and their manipulation requires different policy tools. The consensus appears to favor income because this choice reduces the simultaneity problems associated with intake and food expenditures. In this study income is also used.

⁴ The shape may resemble an inverted 'U' and the inflection point has been referred to as a 'kink', though the word does not imply a discontinuity (Strauss & Thomas 1990).

⁵ Regression of caloric intake on the intake of other nutrients and, say, age could yield an R-Squared in excess of 0.8. This is generally true of this type of data.

⁶ In fact all functions that can be written in additive form may be approximated. These functions need not be monotonic, so that $\phi(\cdot)$ could be quadratic, cubic, or have any other shape.

⁷ In separate regressions I utilized per capita income and income as percent of

poverty guideline and obtained essentially the same results. The square and logarithm of income were also included in the regression as a first test for nonlinearity. Although both were statistically significant, income still had the wrong sign.

⁸To save space we excluded these plots from the manuscript. They can be obtained by writing the authors.

⁹ Both nonparametric methods discussed above are known to overfit and indicate erroneous functional relationship in the presence of outliers. To avoid this pitfall and also to strengthen the analysis, we first deleted the observations outside three standard deviations from their mean (381 observations). Analyses were also run with the full sample; the conclusions drawn here were not altered in any way.

¹⁰ I used the subjective methods suggested in Cleveland & Devlin (1988) to choose the fraction of the data used for fitting the curves. For the reported plots 80% of the data were utilized to fit the curves. Our experiments with other fractions revealed the same, though more jagged, overall pattern.

¹¹Figure is excluded to save space but can be obtained from the author.

¹² Using the British family expenditure survey data, Blundel & Ray (1984) rejected the linearity of *commodity* Engel curves. They indicated that the extent of nonlinearities was equally influenced by the choice of technique and the goods considered. Also, using nonparametric methods, similar results for commodity Engel curves have been reported for the Netherlands (Bierrens & Pott-Butter 1990).

¹³ The ACE algorithm applies different ‘optimal’ transformations at different ranges of data. Although it is not possible to recover the exact transformations applied by ACE, the plots are suggestive of possible functional specifications.

References

- Alderman, H. (1989), 'New research on poverty and malnutrition: what are the implications for research and policy?', *Mimeo, IFPRI*.
- Behrman, J. & Wolfe, B. (1984), 'More evidence on nutrition demand: Income seems overrated and women's schooling underemphasized', *Journal of Development Economics* **14**, 105–128.
- Behrman, J., Deolalikar, A. & Wolfe, B. (1988), 'Nutrients: Impacts and determinants', *World Bank Review* **2**, 299–320.
- Behrman, J. & Deolalikar, A. (1987), 'Is variety the spice of life? implications for nutrient response to income', *University of Pennsylvania*.
- 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*.
- Berck, P. & Chalfant, J. (1990), 'Forecasts from a nonparametric approach: ACE', *American Journal of Agricultural Economics* **72** (3), 799–803.
- Bierrens, H. & Pott-Butter, H. (1990), 'Specification of household Engel curves by nonparametric regression', *Econometric Review* **9**, 123–184.
- Blundel, R. & Ray, R. (1984), 'Testing for linear Engel curves', *Economic Journal* pp. 800–811.
- Breiman, L. & Friedman, J. (1985), 'Estimating optimal transformations for multiple regression and correlation', *Journal of American Statistical Association* **80**, 580–620.
- Cleveland, W. (1979), 'Regression by local fitting', *Journal of American Statistical Association* **74**, 829–836.
- Cleveland, W. & Devlin, S. (1988), 'Locally weighted regressions: an approach to regression analysis by local fitting', *Journal of American Statistical Association* **83**, 596–610.
- Deaton, A. & Muellbauer, J. (1980), 'An Almost Ideal Demand System', *American Economic Review* **70**(3), 312–326.
- Efron, B. & Tibshirani, R. (1991), 'Statistical data analysis in the computer age', *Science*.

- Hama, M. & Riddick, H. (1988), 'Nationwide food consumption survey, 1987', *Family Economic Review* **2**, 24–27.
- Institute of Medicine (1991), *Improving America's Diet and Health: From recommendation to action*, National Academy Press, Washington, D.C.
- Meese, R. & Rose, A. (1991), 'An empirical assessment of non-linearities in models of exchange rate determination', *Review of Economic Studies* **58**, 603–619.
- 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.
- 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.
- National Research Council (1989a), *Diet and Health: Implications for Reducing Chronic Disease Risk*, National Academy Press, Washington, D.C.
- National Research Council (1989b), *Recommended Dietary Allowances*, 10th Edition, National Academy Press, Washington, D.C.
- Pao, E. & Mickle, S. (1981), 'Problem nutrients in the United States', *Food Technology*.
- Peterkin, B., Rizek, R. & Tippett, K. (1988), 'Nationwide food consumption survey, 1987', *Nutrition Today* **23 (1)**, 18–24.
- Strauss, J. & Thomas, D. (1990), 'The shape of calorie-expenditure curve', Discussion Paper No. 595, *Yale Economic Growth Center*.
- Timmer, C. P. & Alderman, H. (1979), 'Estimating consumption parameters for food policy analysis', *American Journal of Agricultural Economics* pp. 982–987.
- Timmer, C. P., Falcon, W. P. & Pearson, S. R. (1983), *Food Policy Analysis*, World Bank, Washington D.C.
- U.S. Bureau of the Census (1989), *Poverty in the United States: 1987*, Current Population Report, Series p-60, No. 160, Government Printing Office.