7. CRITERIA FOR DISTINGUISHING GROUPS FOR PARAMETER ESTIMATION

This report has been prompted by the need to identify the demand parameters of subsets of a population for use in food policy analysis. It is worthwhile, then, to consider how the criteria for subdividing the population are determined. Such criteria are needed regardless of the functional form; they are needed whether or not a systems approach is used, and whether the subdivisions will be applied with covariance analysis or with separate estimations of separate samples. Even when the parameters vary continuously rather than discretely, any table or scenario used to present the findings to policymakers will report the parameters by subgroups. The division is, of course, influenced by the availability and expected application of the data, but frequently the choice is less clear-cut than a division by a discrete characteristic such as region, employment category, or sex of the household head.

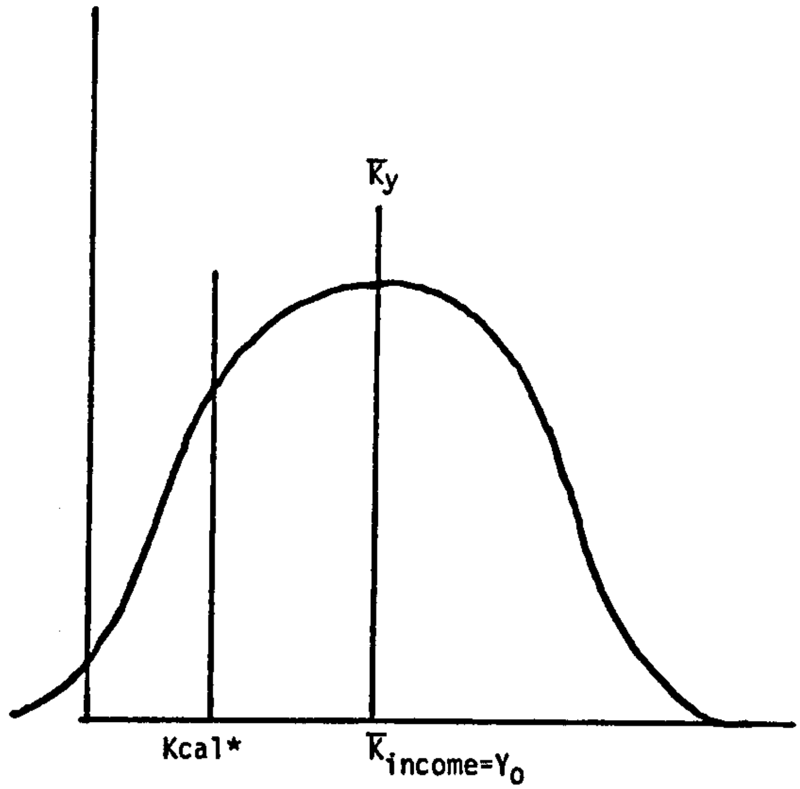
As the goal here is to formulate nutrition policy, one seemingly logical choice is to subdivide the sample by calorie consumption and estimate separate parameters for severe or moderately calorie-deficient households. There are, however, statistical and practical limitations to this.

From a statistical standpoint, subdividing the sample using the dependent variable for truncation introduces the same nature of bias that stimulated the development of the Tobit model. Suppose that Observed Kcal = True Kcal ± error. Presume this error is a recording error and equal for all levels of Kcal, distributed normally with the mean at zero. Suppose also that the analyst picks a cutoff at Kcal*, as in Figure 4. Given the normal distribution, if Kcal* is less than $\bar{K}y$, more cases in which Observed Kcal < Kcal* < True Kcal will be included in the subsample than cases in which Observed Kcal > Kcal* > True Kcal are excluded. Under such conditions, the error in the truncated sample is not distributed normally and has a mean value less than zero. Furthermore, the distribution of the error will be heteroskedastic as it will vary by income. This is so because $\bar{K}y$ increases with income while Kcal* is a constant. The bias in the error due to truncation will vary by income groups even though, under the original assumptions, the error was unaffected by the level of Kcal.

The discussion has been in terms of measurement error but, given the known importance of individual variation of calorie requirements, it can be restated.[103] A sample divided by calorie consumption, in

[103]The importance of individual variation is discussed by a number of authors. See George H. Beaton, "Energy in Human Nutrition: Perspective and Problems," Nutrition Reviews 41 (1983): 325-340. Beaton also makes the point that requirement is a normative concept.

Figure 4 — Truncation of normally distributed calorie variable



effect, aggregates two types of consumers. Suppose the population has four subgroups, low-income consumers with and without high biological needs for food and high-income consumers with and without high biological needs for food. (Family aggregation complicates matters but is tangential here. As changes in activity change requirements, the term "requirements" is used here in the sense of what would be consumed if activity were not curtailed because too few calories were available.) The low Kcal group will include individuals with high requirements for food but limited resources and individuals with low demand but adequate resources. The income elasticities generated reflect what happens to consumption in the group when income increases. They will be averages of the low response of the low-requirement/high-income group and the average response of the bulk of the low-income group. This is, again, a missing variable bias, for a variable denoting requirements is missing. One can expect ∂Kcal/ ∂requirements > 0 and, since the truncated sample includes mainly high-income/low-requirement individuals and low-income individuals with varied requirements, requirements in the truncated sample (though not in the overall sample) will be negatively correlated with income. This will bias income elasticities downward.[104]

[104]This point is made more formally in A. S. Goldberger, "Linear Regression after Selection," Journal of Econometrics 15 (April 1981): 356-366.

An analogous point can be made with price responses in which the truncated sample includes both high- and low-requirement groups at high prices and only low-requirement groups with low prices. In such a case, requirements and price in the truncated sample will be positively correlated. While missing variable bias is a general problem with estimation, the essential point of this discussion is that truncation by the value of the dependent variable can exacerbate the problem by introducing correlation of the missing variable and the independent variable where none had existed before. The same kind of problem exists if, instead of separate regressions made on a subsample, different parameters are estimated by covariance techniques.

The analysis is more commonly done on individual commodities than on calories. The problem with using a Kcal cutoff is, then, somewhat mitigated, but as Kcal is actually a linear transformation of the sum of the commodities, the error in a few commodities may dominate the error in Kcal and the truncation can still introduce estimation errors.

REAL INCOME AND HOUSEHOLD DEMOGRAPHICS

The most common alternative to subdivision of the sample by caloric intake is subdivision by income.[105] Alternatively, one can subdivide by employment category, region, or sex of household head, depending, in part, on the expected use of the data. For example, if the data will be used for a program to lower the consumer price by improving or subsidizing transportation of food to a region in which post-harvest prices are remarkably high, then a regional breakdown may be sufficient, although the remaining cross-sectional price variability will be limited. More commonly, however, the targeting or planning refers to the "poor" defined either by consumption of certain goods, especially food and shelter, or by the income that makes it possible to purchase a certain amount of goods. These two concepts are different.[106] The former assumes that different goods have social weights that reflect a concept of one-dimensional equity--that is, the view that society considers equity on the consumption of a certain bundle of goods more important than equity overall. The latter is founded more on the concept of consumer sovereignty, assuming basically that, since a household maximizes its utility subject to price and budget constraints, it is sufficient to look at the size of income at which nutritional goals or other consumption goals can be met by the household. As mentioned, using the former criterion in estimation is likely to create econometric biases. Even when it is administratively more convenient to identify low-consumption groups than low-

---

[105]Frequently the sample frame of a household survey is based on a stratified design. It may be possible, then, to separate households from different strata if that information is available.

[106]For a discussion of poverty indicators, see Michael Lipton, Poverty, Undernutrition and Hunger. Also see A. K. Sen, Levels of Poverty: Policy and Change, World Bank Staff Working Paper 401 (Washington, D.C.: World Bank, 1980).

income groups, it may be advisable to estimate income and price parameters by income group and apply them to the low-consumption groups.

If income is used to distinguish subdivisions of the population, then the question remains whether household income, income per capita, or some other standardized income term is more relevant. The potential for misclassification is readily apparent if household income is used; a large poor household may rank above a single individual with a moderate income. The alternative of per capita income, however, has problems with economies of scale and the percentage of children in a household. If economies of scale exist for some commodities then a household with N individuals and M units of earning will be less well off than a household with 2N members and 2M units of earnings. Note that this is true even if there are no economies of scale for the consumption of foods, which are the central focus of the study. If some households have economies of scale in housing or transportation, then they have more disposable income for food items at the same per capita income. The issue here is the possibility of welfare comparisons based on some normalized income and not the importance of demographic variables in predicting consumption of a given good. The latter use of demographics can be achieved ad hoc, without theoretical underpinnings, by including such variables in any demand equation.[107] If, however, one wishes to go from a commodity-specific analysis to a method of comparing household income, some assumptions about consumer behavior are necessary.

For example, adult equivalency scales based on caloric requirements are commonly used to standardize households. Although this distinguishes the requirements for children and has some justification in aggregate food demand equations, it does not allow for economies of scale (after all, two adult males are exactly twice one), and it has no foundation in utility theory. Utility is not likely to be a function of calorie requirements.

Another means of scaling household income can be based on Engel's law, which holds that the budget share of food declines monotonically with increases in real income. It can be assumed, then, that two households that allocate identical budget shares to food have identical real incomes. This information can be used to compare the real incomes of families with different sizes and composition.[108]

Another method is based on a comparison of purchases of two bundles of goods--one that children share and another, containing such goods as tobacco and alcoholic beverages, consumed primarily by adults. It is assumed that the latter purchases indicate welfare,

---

[107]For reviews, see Deaton and Muellbauer, <u>Economics and Consumer Behavior</u>, Chapter 8, and Pollak and Wales, "Demographic Variables <u>In Demand Analysis</u>."

[108]Angus Deaton, "Inequality and Needs: Some Experimental Results for Sri Lanka," <u>Population and Development Review</u>, Supplement to Volume 8, (1983): 35-49.

independent of the number of children, and that differences in expenditures on these goods, controlled for nominal income, will reflect the costs of children. Note that this method does not handle the economies of scale of adult members and, as it reflects a partitioning of commodities, indirectly includes an assumption of additivity.

These methods, as well as others reported in the literature, do not necessarily scale household income equivalents identically.[109] It is beyond the purpose of this report to delve deeply into the implications of various methods of comparing household income when family sizes change. However, an analyst who intends to report parameters by various groups should have some understanding of the implications of the classifications. Another justification for the discussion of ranking of incomes is that some transfer programs base benefits partially on family size, thereby making implicit assumptions on the scale economies of children.

Having ranked households by some income measure--whether per capita or per household or some other scaled measure--a cutoff point for the target group is still needed. Occasionally, an official poverty line is given or defined in terms of either calories consumed or the ability to consume a certain amount. Alternatively, an arbitrarily determined percentile can be the focus--the poorest quartile or decile, for instance. There is one technique that allows the data to determine the group classifications internally. Rather than focusing on, say, the ith percentile of scaled income, the parameters can be allowed to switch at an internally determined point, using a maximum likelihood algorithm to search for the best switching point.[110] This will give disjointed rather than continuous changes of parameters. Unfortunately, it cannot rigorously test whether any two coefficients are statistically different, although Akin et al. used such a method to determine a stratification and then made separate regressions on both strata. They then used a Chow test to determine statistical differences.

On the other hand, testing for statistical differences using covariance techniques is fairly straightforward. One can create dummy variables for any group and use a multiplication interaction in the regression for dummy x price. The test of whether the derivative with respect to price for that group differs from the general population will be the t-statistic for the parameter of the interaction variable. The total response of the group will be the sum of the general population response and the coefficient of the interaction term, while the variance will be the sum of the variance of the two

---

[109] See, for example, Angus Deaton, Three Essays on a Sri Lanka Household Survey, Living Standards Measurement Study, Working Paper 11 (Washington, D.C.: World Bank, 1981).

[110] See John S. Akin, David K. Guilkey, and Barry M. Popkin, "The School Lunch Program and Nutrient Intake: A Switching Regression Analysis," American Journal of Agricultural Economics 65 (August 1983): 477-485. Note, however, that switching regressions is expensive and not easily done.

parameters and twice their covariance. Alternatively, one can include interaction terms for all groups and then exclude an independent price term. The problem of collinearity when dummy variables for all subgroups are included will be avoided, as the intercept will not be a linear combination of the interactions. This gives the price parameter for each group; the difference between the price parameters can be tested, as above, with the standard error derived from the covariance of the parameters.