CORNELL AGRICULTURAL ECONOMICS STAFF PAPER

Sample Partitioning Bias in Estimating the Effects of the Food Stamp Program

by

Christine K. Ranney

March 1985

No. 85-2

Department of Agricultural Economics
Cornell University Agricultural Experiment Station
New York State College of Agriculture and Life Sciences
A Statutory College of the State University
Cornell University, Ithaca, New York, 14853

It is the policy of Cornell University actively to support equality of educational and employment opportunity. No person shall be denied admission to any educational program or activity or be denied employment on the basis of any legally prohibited discrimination involving, but not limited to, such factors as race, color, creed, religion, national or ethnic origin, sex, age or handicap. The University is committed to the maintenance of affirmative action programs which will assure the continuation of such equality of opportunity.

FOREWORD

Christine K. Ranney is an Assistant Professor in the Department of Agricultural Economics, Cornell University.

This research was supported by Regional Research Project NC152, Research Agreement No. 58-9AHZ-1-14 with the Western Human Nutritional Research Center, and Regional Research Project S165.

The author is grateful to John Kushman and Sylvia Lane for their help in the development of this study and to Bill Greene for providing the Probit Computer program. The author would also like to thank Linda Johnson for her patience in typing the manuscript.

Copies of this paper may be obtained from:

Office of Publications
Department of Agricultural Economics
Warren Hall - Cornell University
Ithaca, NY 14853-7801

Abstract

This paper addresses the possibility of biased estimates of the determinants of food expenditures of households eligible for the Food Stamp Program. The source of such bias, sample partitioning, is described and a correction procedure is utilized. For eligible nonparticipants' food expenditure estimates, bias was found but not for participants.

TABLE OF CONTENTS

LIST OF TABLES

The Food Stamp Program (FSP) is one of the largest domestic redistributive transfer programs. Approximately 22 million persons participated in the program in 1983 at a cost of approximately $12 billion. Given the magnitude of the program, accurate methods are needed for evaluating the impact of the existing program structure or proposed structural changes on the achievement of program objectives and for forecasting program costs. Of particular concern is the ability to accurately predict how many eligible households will participate in the Food Stamp Program and the impact of the program on food expenditures of eligible households. The purpose of this paper is to address the possibility of biased food expenditure parameter estimates due to sample partitioning and to correct for such bias, if present.

The paper is organized as follows: the data utilized for this study are described first, followed by a discussion of the sample partitioning problem. The next section presents the empirical model and results and the final section presents conclusions.

The Data

This study is based on household survey data from one metropolitan and one nonmetropolitan county in each of four states, California, Indiana, Ohio, and Virginia. Eight hundred ninety-six households eligible for the Food Stamp Program were interviewed between July 1979 and May 1980. Five hundred thirteen of the interviewed households indicated they participated in the program during the month of the interview. At the time of the survey participants were no longer paying cash for stamps. Their entire allotment was so-called "bonus stamps."

Sample Partitioning--The Econometric Problem and Remedy

The nature of econometric problems is often dictated by the characteristics of the economic data. The sample partitioning problem, a variant of the sample selection problem, can occur when estimation is based upon a sample that is nonrandom due to the manner in which the sample is selected by the analysts, or due to the fact that the units of observation, through their own behavior, can self-select or partition themselves into a particular sample. The sampling procedure for this study selected households from the population of households eligible to participate in the Food Stamp Program. Eligible households were, however, not randomly assigned to program participant or nonparticipant groups. The sampled households self-selected or partitioned themselves into potentially nonrandom subsamples of participants and nonparticipants.

Suppose we wish to predict what average food expenditures would be under various program structures given our household survey data. For example, how would expenditure behavior change if bonus stamp allotments were increased? Using ordinary least squares estimation on the subsample of participating households would tell us how already participating households may respond to such a program change. However, all eligible nonparticipants are potential participants. Increasing bonus stamp allotments would be likely to encourage some nonparticipants to become participants. Thus, the true effect on average expenditures of this program change would depend upon the expenditure behavior of both continuing participants and new participants. In other words, the observed "average" relationship between the allotment and food expenditures will depend on who is included in the group being averaged. If there is bias in the self-selection process, the observed relationship between allotment and expenditures of participating households will not be representative of the relationship if all eligible households or a random selection of households were included in a Food Stamp Program experiment as participants.

We cannot directly observe the expenditure behavior of households newly participating due to a hypothetical program change. If the participants and nonparticipants in our sample behave, on average, in systematically different ways, estimates of the effects of changes in the FSP on food expenditures based upon the nonrandom subsample would be biased and inconsistent. The problem of possible sample selection bias with respect to estimation of the expenditure relations must be resolved.

The recent literature on the presence of and correction for such bias was developed by James Heckman (1979) in the context of estimating labor supply equations. Heckman has made two major contributions with regard to eliminating this bias. First, he has shown that sample selection or partitioning bias can be viewed as arising from an ordinary omitted variable problem. The omitted variable is a function of the probability that each observation self-selects into the subsample. Estimates based upon the subsample will be biased as long as the omitted variable are not included in the equation. Second, Heckman developed a method that provides consistent estimates of the omitted variable from probit analysis of the sample selection process, i.e., from an estimate of the probability of self-selecting into the subsample. Estimates of the omitted variable can then be included as a regressor when estimating a relationship using the self-selected subsample observations.

More formally, consider a three equation model and a random sample of $i = 1, 2, \ldots, N$ observations.

$$Y_{1i} = X_{1i}B_1 + u_{1i} \tag{1a}$$

$$Y_{2i} = X_{2i}B_2 + u_{2i}, \text{ and} \tag{1b}$$

$$Y_{3i} = X_{3i}B_3 + u_{3i}. \tag{1c}$$

where the X_{ji} are vectors of exogenous regressors and the B_j are vectors of coefficients. The u_{ji} are the disturbances which are assumed to be bivariate normal for $j = 1, 2, 3$ and have zero means with no correlation across observations, but there is correlation for a given observation between u_{1i} and u_{3i} and between u_{2i} and u_{3i}. In the context of the Food Stamp Program equation (1c) represents the participation equation while (1a) and (1b) are food expenditure equations given participation and nonparticipation, respectively.

Now, suppose we would like to estimate food expenditures given participation (1a). Following Heckman, the regression function for the selected sample, given the selection rule $Y_{3i} > 0$ is:

$$E(Y_{1i} | X_{1i}, Y_{3i} > 0) \tag{2a}$$

$$= X_{1i}B_1 + E(U_{1i} | Y_{3i} > 0) \tag{2b}$$

$$= X_{1i}B_1 + E(U_{1i} | U_{3i} > -X_{3i}B_3) \tag{2c}$$

$$= X_{1i}B_1 + (\sigma_{13}/ \sigma_{33}^{1/2}) \lambda_i. \tag{2d}$$

For (2d) the assumption that $h(u_{1i}, u_{3i})$ is a bivariate normal density, is required and

$$i = \frac{\phi(Z_i)}{1-\Phi(Z_i)} = \frac{\phi(Z_i)}{\Phi(-Z_i)} \tag{3}$$

with ϕ and Φ the density and distribution functions, respectively, for the standard normal variable, and $Z_i = -X_{3i}B_3/(\sigma_{33})^{1/2}$. The λ_i is the inverse of the Mills Ratio, which is often referred to as the hazard rate in reliability studies.

The selected sample regression function, therefore, depends upon both x_{1i} and x_{3i}, and sample selection bias can be viewed as having arisen from an omitted term, $(\sigma_{13}/\sigma_{33}^{1/2})\lambda_i$. Inserting a consistent estimate of λ_i, say $\tilde{\lambda}_i$, into the food expenditure equation will yield consistent estimates of both $\sigma_{13}/\sigma_{33}^{1/2}$ and B_1. Heckman further shows that the consistent estimates of λ_i can be obtained from probit analysis of the participation equation.

Application of the Heckman model usually occurs when observations are missing on Y_{1i} unless $Y_{3i} > 0$. In the case of the Food Stamp Program, expenditures on food (Y_{1i}) are observed whether or not the household participates. The impact of program benefits and costs on food expenditures, however, are not observed unless the household participates in the program ($Y_{3i} > 0$). In such a case, where there are no actual missing observations, the Heckman approach has been referred to as correction for "the sample partitioning problem" rather than the "sample selection problem." The extension of the Heckman model to the sample partitioning case is quite simple and has been outlined concisely by Greene (1979). Greene shows that omitted term relevant for estimating food expenditures of nonparticipants is $(\sigma_{23}/\sigma_{33}^{1/2})\lambda_i$, where $= \phi(Z_i)/\Phi(Z_i)$ with Z_i as defined for (3).

Note that the estimates of B_1 and $\sigma_{13}/\sigma_{33}^{1/2}$ and of B_2 and $\sigma_{23}/\sigma_{33}^{1/2}$ will be consistent but inefficient. The inefficiency arises due to the heteroskedasticity due to the correlation of u_{1i} and u_{3i} and u_{2i} and u_{3i}, respectively. A standard generalized least squares procedure can be used to obtain appropriate standard errors. Further, if u_1 and u_3 (u_2 and u_3) are independent, then $\sigma_{13} = 0$ ($\sigma_{23} = 0$) and the "omitted variable" drops out of the regression function for the selected sample. Similarly, if X_{3i} and X_{1i} (X_{3i} and X_{2i}) are uncorrelated, no bias will be present. A test of the hypothesis that $\sigma_{13}/\sigma_{33}^{1/2} = 0$ ($\sigma_{23}/\sigma_{33}^{1/2}$) is a test for the absence of selection bias.

While the Heckman technique as extended by Greene does correct for possible sample partitioning bias, there are some drawbacks to using the technique. Ordinary least squares estimation of participants' or nonparticipants' food expenditures requires estimation of the probability of participation, using probit analysis to calculate the sample selection or partitioning bias correction factor (λ_i). Computational practicality requires specifications that are linear in the parameters though not necessarily in the variables. The form of the actual food expenditure equations to be estimated should be determined by the form of the indirect utility function from which it is derived which may not, in fact, be linear in parameters.

Empirical Model and Results

As discussed in the previous section, the Food Stamp Program participation and food expenditure decisions are not statistically independent. On theoretical grounds as well, the decisions should not be considered independently. For FSP eligible households, utility maximization involves making the participation decision and then determining their level of food demand/expenditures given their participation decision.

Before they can use food stamps, eligible households must undertake certain activities. They must apply, be certified as eligible, and procure the stamps. These activities are not cost free and households may have different preferences regarding them. For instance, households may feel stigmatized by the certification process and by using stamps and, thereby, experience a loss of prestige when participating. Additionally, since food stamps are not cash but are restricted purchases, households may feel constrained by this restriction. Costs and preferences could exist such that eligible households would rationally choose nonparticipation and forego an increase in resources in the form of food stamps. Formally, FSP eligible households can be viewed as comparing the value of their indirect utility function, the maximum utility possible, given participation to that given nonparticipation. Their food demand/expenditure relations are derived, in turn, from the relevant indirect utility relationship.

We are primarily interested in testing for the presence of sample partitioning bias in estimating the food expenditures of FSP participants and nonparticipants and in correcting for such bias if necessary. In Ranney (1983) the same data used for this study was used to analyze the participation decision. The participation equation was specified as a function of the difference between the value of the logarithmically transformed indirect utility function when the household participated and the value of the corresponding function when the household did not participate in the Food Stamp Program. Probit analysis of the probability that a household participated in the program was utilized to estimate a participation equation. Estimated values of the extent to which the stamp allotments are cash equivalent (c) were calculated for each household with estimates of the relevant coefficients from the probit analysis of the participation equation. The calculated C's will be used to define two explanatory variables, Ca and (1 - C)a included in the specification of participant food expenditure equations. The probit participation results also generated sample partitioning bias correction factors, λ_P and λ_{NP}, that will be included as regressors when estimating the food expenditure equation of participants and of nonparticipants, respectively.

A combination of theoretical and pragmatic factors contributed to the specifications of the food expenditure equations. From general theoretical considerations, we know that the demand for food by nonparticipants is a function of relative prices, real income (total resources) and household characteristics. In addition to those variables, participants' food demand is a function of the real food stamp allotment. Because the sample was gathered over a relatively short time span, approximately ten months, and the regional price indices for those months exhibit very little variation, the food expenditure equations will not include prices as exploratory variables. Based on the exploratory nature of this study and the requirement of the sample partitioning bias correction procedure that the functions be linear in parameters, a simple linear specification of the food expenditure equations has been chosen.

For nonparticipants, (NP) the food expenditure specification is

$$EXP^{NP} = b_o + b_1y + b_2HSEX + b_3ED1 + b_4ED2 + b_5ETH1 + b_4ETH2 + b_5HAGE + b_6\lambda + u, \qquad (4)$$

where EXP_{NP} is monthly food expenditures, including the value of home-produced food; y is monthly income per index person; HSEX is equal to one if the sex of the household head is male (=0, otherwise); ED1 is the number of years of education of the household head if the number of years is less than or equal to twelve (=0, otherwise); ED2 is equal to 1 if the number of years of education of the household head is greater than twelve (=0, otherwise); ETH1 is equal to one of the ethnicity is Black (=0, otherwise); ETH2 is equal to one if ethnicity is "other" (=0, otherwise); HAGE is the age of the household head; is the sample partitioning correction factor; and u is the error term.

For participants (P), the food expenditure specification is

$$EXP^P = b_o + b_1y + b_2Ca + b_3(1-C)a + b_4HSEX + b_5ED1 + b_6ED2 + b_7ETH1 + b_8ETH2 + b_9HAGE + b_{10}\lambda + u, \qquad (5)$$

where all but the third and fourth terms are as defined for nonparticipants. The second, third, and fourth terms together represent total household resources per index person. They have been specified separately to allow for differences in their effects on food expenditures.

Heuristically, the extent to which food stamps increase expenditures on food depends on (1) the effects of stamps through an expansion of general purchasing power and (2) the extent to which stamps constrain part of total household resources to be spent on food. These different effects are reflected in the third and fourth terms of the food expenditure equation Ca and (1 - C)a, respectively. An increase in stamp allotments is hypothesized to increase food expenditures, but less so the more nearly cash equivalent or constraining stamps are. This taken together with the hypothesis that stamps will have a greater effect than income (y) on increasing food expenditures require $b_3 > b_2 > b_1 > 0$. These signs and differences are adopted as hypothesis in estimating the food expenditure functions of participants.

Each equation was estimated with and without the sample partitioning correction factor to facilitate comparisons. Included in the respective subsamples were 310 participating households and 346 nonparticipating eligible households that completed all relevant questions. The estimation results are presented in Table 1.

Recall that the test for the presence of sample partitioning bias is a t-test of whether the coefficient for the bias correction factor (λ) is significantly different from zero. Based on these results, bias is probably present in the case of nonparticipants but not for participants. In keeping with the insignificance of λ in the participants' equation, the OLS and S-P coefficients are nearly the same. Note, however, the differences between coefficients from S-P and OLS for nonparticipants. All except that for income differ and some differ substantially. Because the inclusion of the sample partitioning bias correction factor (λ) in the specification not only allows one to test for such bias, but also removes the bias from the other coefficients, the results from the S-P procedure should be used for predicting the behavior of nonparticipants. Similarly, in the absence of evidence of bias, the OLS or S-P estimates would be appropriate for the purposes of prediction.

Based on the OLS estimates for participants and the S-P estimates for nonparticipants the economic variables are clearly important in predicting their food expenditures. For participants, the results show that stamps are more efficient or powerful in increasing expenditures on food by a factor of six to one. Further, b_1, and b_2 and b_3 are significantly in the hypothesized relationship the results also have the intuitively appealing feature that the coefficient for the most

Table 1
Food Expenditure Estimation Results for
Participants and for Nonparticipants:
OLS and S-P Coefficients and Standard Errors

Variable	Coefficient (Standard Error)			
	Participants		Nonparticipants	
	OLS	S-P	OLS	S-P
Constant	36.09 (30.65)	27.91 (30.13)	86.59 (23.42)	111.12 (20.56)
y	0.11 (0.16)	0.10 (0.02)	0.10 (0.01)	0.09 (0.01)
Ca	0.52 (0.13)	0.58 (0.12)	-- --	-- --
(1 - C)a	1.12 (0.03)	1.17 (0.26)	-- --	-- --
HSEX	4.69 (9.45)	1.67 (9.21)	5.84 (6.85)	1.37 (6.62)
ED1	23.15 (12.41)	22.15 (12.60)	-7.44 (9.46)	-9.83 (9.52)
ED2	2.93 (1.91)	2.72 (1.93)	0.79 (1.40)	0.59 (1.42)
ETH1	1.23 (8.88)	1.72 (9.01)	-5.81 (7.43)	-2.60 (7.34)
ETH2	-16.30 (17.99)	-13.88 (17.15)	64.47 (19.05)	72.56 (21.22)
HAGE	-0.33 (0.31)	-0.36 (0.31)	-0.88 (0.19)	-1.00 (0.18)
λ	-15.09 (13.89)	-- --	-21.00 (10.15)	-- --

constrained portion of stamps, (1 - C)a, the coefficient is nearly one. That is, an additional dollar added to the most constrained portion of stamps results in one dollar of additional food expenditures. For nonparticipants and participants a one dollar increase in income increases food expenditures by ten cents. With the exception of having less than 13 years of education (ED1) for participants and the proportion of household with "other" ethnicity (ETH2) the household characteristics have little impact on expenditures.

Conclusions

Estimation of the impact of Food Stamp Program on eligible households' food expenditure decisions should account for the possibility of sample partitioning bias. While the bias problem arose due to the way the household survey data used here was collected, other widely-used survey data, such as the Nationwide Food Consumption Survey (NFCS) and the Consumer Expenditure Survey (CES), suffer from the same problem. In terms of estimation, the primary implication for other FSP researchers is that food expenditures should not be estimated in isolation. The participation decision must be estimated as well to capture correlation between the food expenditure and program participation decisions and to correct for possible sample partitioning bias. Indeed, any researchers utilizing samples where the units of observation can self-select into the sample should recognize the potential for bias and estimate the selection rule or participation decision as a necessary first step.

References

Greene, William. Manual for LIMDEP Computer Package. Ithaca, New York: Department of Economics, Cornell University. (mimeographed) 1979.

Heckman, James. "Sample Partitioning Bias as a Specification Error." Econometrica. 74 (January 1979): 153-161.

Ranney, Christine K. "Food Stamp Program Participation and Cash Equivalent Benefits." Cornell Agricultural Economics Staff Paper No. 83-24. Ithaca, New York: Department of Agricultural Economics, Cornell University, 1983.