THE ESTIMATION OF FOOD STAMP SELF-SELECTION MODELS USING THE METHOD OF SIMULATION

Keane, Michael and Moffitt, Robert

1992

CONTENTS

TABLES

EXECUTIVE SUMMARY

An important issue associated with the Food Stamp Program (FSP) concerns the magnitude of its effects on the food expenditures, nutrition, and other outcomes of recipients. What must be considered in estimating the magnitudes of those effects is a well-known, but difficult, statistical problem arising from what is called "self-selection" into the program. The problem arises when the effects of the program are gauged by a comparison of the outcomes of recipients to those of eligible nonrecipients. Such comparisons will be in error if the values of the outcomes observed for nonrecipients are not the same as the outcomes that recipients would experience were they off the program. This discrepancy will occur if recipients are a "self-selected" group from the total population of eligibles. For example, if, as a group, recipients would have lower food expenditures if they were off the program than current nonrecipients are observed to have, the observed difference in food expenditures between recipients and nonrecipients would either be too small, if positive, or possibly negative, and the estimated effect of the FSP would be biased.

While statistical solutions to this problem have been developed to be able to obtain correct comparisons for households and individuals that participate in the FSP alone, only limited progress has been made in developing solutions for the more common case in which households and individuals are recipients of benefits from multiple programs. The problem in this case arises when attempting to gauge, for example, the effects on food expenditures of receiving both food stamps and Aid to Families with Dependent Children (AFDC) (or some other program benefit). Comparisons of the food expenditures of those receiving benefits from both programs to the food expenditures of either those receiving only food stamps, only AFDC, or no benefits at all may all be incorrect if those who participate in both the FSP and AFDC are a self-selected group whose food expenditures differ from those of the other recipient and nonrecipient groups independent of the programs per se (that is, if those who are in both programs have especially low food expenditures in the absence of program participation). For example, data may show that FSP recipients who are also AFDC recipients have lower food expenditures than FSP recipients who are not on AFDC, but this may be only because FSP recipients who are also on AFDC are worse off than FSP recipients not on AFDC and would have lower food expenditures than those non-AFDC-recipients even they were not on AFDC.

This report details a technique for solving this more general problem of self-selection into multiple programs. We apply recently developed methods for the estimation of "large" numbers of choice equations (e.g., more than two) to the problem of estimating the true effect of participation in the FSP and other programs on an outcome variable. The new technique is more computer-intensive than the prior techniques developed for the FSP-only case, but can still be handled by modern computers. We present an illustration for the case of three possible programs and report the computer times required for estimating the model with the Survey of Income and Program Participation (SIPP) data. We also include a diskette with the software capable of estimating models with up to four possible programs and technical documentation for its use.

I. INTRODUCTION

Much research sponsored by the U.S. Department of Agriculture's Food and Nutrition Service (FNS) has evaluated the effects of the Food Stamp Program (FSP) and other food and nutrition programs on outcomes of interest (for example, dietary intake or food expenditures). The problem of "self-selection" frequently arises in evaluations of assistance programs in general and in analyses of food and nutrition programs in particular. Self-selection occurs when participants in a program differ from eligible nonparticipants in ways that are (1) related to the outcome variable of interest but (2) are not measured in the data available to the analyst. The result of self-selection is that conventional estimates of program effects are biased.

An example of bias arising from the self-selection of eligibles into a program is the estimation of the effect of food stamps on food expenditures. If that effect is estimated by comparing the difference in food expenditures between eligible recipients and eligible nonrecipients, the danger of self-selection bias arises because recipients and nonrecipients might have different food expenditures even in the absence of the FSP. It may be the case that households that apply for benefits and become food stamp recipients have below-average food expenditures in the first place--indeed, they may have applied for food stamps because they were in need of food assistance (perhaps because they have high nonfood expenses). If so, then the observed difference in food expenditures between recipients and nonrecipients will either be too small, if positive, or it may even be negative, and the estimated effect of the program will be biased. In this example, recipients are a "self-selected" group with lower-than-average food expenditures in the absence of the FSP. The problem has arisen because (1) recipients and nonrecipients differ in a way that is related to food expenditures, an outcome of interest, and (2) that difference is not measurable, since we do not know what the food expenditures of each recipient household would be if it were not receiving food stamps.

To control for and eliminate self-selection bias from estimates of program effects, most analysts use a variant of the adjustment technique developed by Heckman (1979) and discussed extensively in a textbook by Maddala (1983). This technique requires that an extra equation be estimated in addition to the main equation for the outcome of interest. The main equation relates program participation to food expenditures, nutrient availability, or some other outcome variable of interest; the second equation is designed to estimate the determinants of program participation itself--for example, by linking the likelihood of program participation to the potential benefit level, household income and size, and other variables. The procedure requires estimating the main equation and the second equation simultaneously (that is, jointly). The technique solves the selection-bias problem because by incorporating the determinants of participation into the estimation process, the second equation "adjusts" the estimate of program effects for nonprogram-related differences between program participants and nonparticipants.

This report addresses the phenomenon of multiple program participation. For example, to study the effect of the FSP on the food expenditures of households headed by a single woman, one must also control for the effects of Aid to Families with Dependent Children (AFDC) receipt, since so many female heads receive both food stamps and AFDC. Similarly, a study of the effects of the School Breakfast Program (SBP) should consider the effects of National School Lunch Program (NSLP), since many students qualify for and receive benefits from both. In fact, multiple program participation may encompass three or more programs, as is the case for families who receive benefits from the SBP, NSLP, and the FSP, or families who receive AFDC, Special Supplemental Food Program for Women, Infants and Children (WIC), and FSP benefits. When an analysis involves two or more programs, severe technical difficulties arise in applying the conventional selection adjustment procedure. An extra equation must be added for each new program, each specifying the determinants of participation in that program; thus, two or three equations must be considered along with the main equation. All these equations must be estimated simultaneously because it is necessary to estimate

the determinants of participation in each combination of programs. This is a formidable problem that has thus far limited the estimation of multiple-program selection models.

In past work with one or two programs only, the problem of self-selection bias has been shown to be important. In studies of the effect of the FSP on the work effort of recipients, Fraker and Moffitt (1988, 1989) found evidence that the work levels of FSP recipients were lower than those of nonrecipients for reasons related to sample selection, not to the FSP itself. In a study of the effects of the NSLP and SBP on food expenditures, Long (1988) found that households with recipient children were self-selected into the programs. Fraker et al. (1989) found self-selection into the WIC program in a study of the effect of WIC and FSP on dietary adequacy. Furthermore, in a study of the effect of the FSP on nutrient availability, Devaney and Moffitt (1990) studied two different types of selection bias. The first type was the standard type, which tends to make observed, measured effects of the FSP too small--recipients tend to have lower levels of outcomes (including nutrient availability) than nonrecipients because recipients are worse off overall. The second, new type of bias arises if those households who participate in the FSP who are those who "get the most out of it" by increasing their food expenditures after enrolling in the program more than other households would. This type of bias would tend to make the observed effect of the FSP too large because those on the program are again a "self-selected" group with higher-than-average food expenditures.

There have been no studies to date involving three or more programs because it has not been possible with existing software and techniques. Yet many FSP households participate in both AFDC and WIC, and others participate in both SBP and NSLP. Our example in the next section is to a case in which many FSP households participate in both AFDC and public or subsidized rental housing. FSP households who participate in three programs other than the FSP is rarer but still occasionally occurs, and can do so for any three of these programs (AFDC, public or subsidized rental housing, WIC, the SBP, and the NSLP).

Fortunately, a promising new econometric methodology has recently been developed to resolve the technical problem of controlling for self-selection into as many as three or more programs. In two papers widely discussed in the academic community, Daniel McFadden of MIT (McFadden, 1989) and Ariel Pakes and David Pollard of Yale University (Pakes and Pollard, 1989) have developed a new technique for estimating large numbers of simultaneous equations of the type generated by the self-selection problem in program evaluations. The "method of simulated moments" (MSM) technique, as it is termed, is designed for a broader set of problems than the self-selection problem, but it is applicable to it as a special case. The MSM technique has attracted attention because it appears to be relatively easy to implement; it involves a simple "simulation" of the simultaneous-equations model and the application of a "method-of-moments" estimation method. The technique is sufficiently new that very few researchers have yet applied it, one exception being a study by Keane (1990).

In this report we discuss the adaptation of this technique to the problem of evaluating self-selection bias in the FSP when multiple program participation is present. In Section II, we discuss the prototype model that we have developed for the application and the issues that arose in applying it. In Section III, we report the results of an illustrative estimation of the model with the new MSM technique, using Survey of Income and Program Participation (SIPP) data on female heads of families who are faced with the choice of three possible programs (FSP and two others). We discuss the computational burden of the technique as well. In the final section, we summarize the results of the estimation. Included as an attachment to the report is a copy of software that can apply the technique to problems with up to four possible programs, and documentation for its use.

II. APPLYING THE MSM TO A MODEL OF MULTIPLE PROGRAM SELF-SELECTION

We have applied the new MSM method to a prototype model drawn from past work on self-selection into the FSP and other programs. Our example has three possible programs, although the software we are providing permits up to four. The mathematical representation of the model is as follows:

$$(1) \quad Y_i = X_i\beta + \alpha_1 B_{1i} + \alpha_2 B_{2i} + \alpha_3 B_{3i} + \epsilon_i$$

$$(2) \quad P^*_{1i} = Z_{1i}\gamma_1 + \nu_{1i}$$

$$(3) \quad P^*_{2i} = Z_{2i}\gamma_2 + \nu_{2i}$$

$$(4) \quad P^*_{3i} = Z_{3i}\gamma_3 + \nu_{3i}$$

The variables in these equations have the following meanings:

Y_i = outcome variable of interest (food expenditures, dietary intake, etc.) for individual i

X_i = variables determining Y, excluding program benefits themselves

B_{1i} = benefit received from program 1 (=0 for nonrecipients)

B_{2i} = benefit received from program 2 (=0 for nonrecipients)

B_{3i} = benefit received from program 3 (=0 for nonrecipients)

P^*_{1i} = variable representing the "propensity" to be a recipient of program 1

P^*_{2i} = variable representing the "propensity" to be a recipient of program 2

P^*_{3i} = variable representing the "propensity" to be a recipient of program 3

Z_{1i} = variables affecting the propensity to be a recipient of program 1 (including the program benefit)

Z_{2i} = variables affecting the propensity to be a recipient of program 2 (including the program benefit)

Z_{3i} = variables affecting the propensity to be a recipient of program 3 (including the program benefit)

The variables v_{1i}, v_{2i}, and v_{3i} represent the effects of unobserved determinants of participation in three programs, while ϵ_i represents the effects of unobservables on the outcome of interest. The coefficients in the model that we wish to estimate are β, α_1, α_2, α_3, γ_1, γ_2, and γ_3.

Equation (1) is the main equation for the outcome variable of interest. Past studies have usually included in this equation program benefits received as well as other variables such as age, household size, and so on (which we represent as "X"). The variables in X may include other program-related variables as well as non-program-related variables--we focus on the benefit variables B because they are the easiest to illustrate. Because we are considering three programs, variables for three program benefits appear in the equation.

Equations (2), (3), and (4) are the equations that determine participation in each of the three programs. The variables that affect participation in each, which we represent as "Z", usually include the potential program benefit as well as other variables (age, household size, and so on) that are thought to affect families' likelihood of receiving benefits.

In most models at least one variable must be in each of the participation equations, equations (2)-(4), that is not in the main equation, equation (1). That is, there must be at least one factor that affects participation in a program that does not directly affect the outcome variable of interest. Access to the program--distance from the nearest office, for example--is an example of such a variable. The presence of such variables permits the effects of participation on the outcome variable to be disentangled from the "self-selection" into the program. For example, an examination of the food expenditures of families who live different distances from the nearest program office allows us to determine the effect of the self-selection because such families will have different participation rates but not different values of Y, such as food expenditures (we are operating under the assumption that distance does not enter the main equation; that is, distance does not affect food expenditures per se).[1]

[1] We might note that this point that our new estimation method does not eliminate this necessity (continued...)

The problem of self-selection bias arises when the determinants of participation, as shown in equations (2)-(4), are related to the unobserved and unmeasured determinants of Y_i, which are denoted in equation (1) by ϵ_i. If, for example, program participants also have below-average values of ϵ_i, then this implies that participants would have lower food expenditures than nonparticipants even if they did not participate.

If the variables represented by Z are correlated with ϵ, this would cause no problem since those variables are, by definition, observed in the data and could just be added to equation (1). But if the unmeasured and unobserved determinants of participation, which are denoted by the terms ν_{1i}, ν_{2i}, and ν_{3i} in equations (2)-(4), are correlated with ϵ_i, then effects of self-selection cannot be controlled for directly.

For a single program, the methodology developed by Heckman (1979) [see Maddala (1983) for a textbook exposition] requires that the equation for participation in the program be estimated jointly with the equation for the outcome of interest. In our case, equation (1) could be estimated jointly with equation (2) if there were just one program. The unobservables ϵ_i and ν_{1i} would be assumed to be correlated. Estimating the model with maximum likelihood would yield unbiased estimates of the coefficient on the program benefit amount (for example, α_1), which are free of self-selection bias. The presence of a variable in the participation equation that is not in the outcome equation is the key to being able to eliminate selection bias.[2]

Unfortunately, the estimation of the model becomes more difficult when multiple programs are present for reasons that are purely computational. The estimation of a single participation equation like equation (2) requires the computation of probabilities--on the computer--that follow the normal

---

[1](...continued)
of having variables in the participation equations that are not in the main equation. It is just as necessary as in models estimated by other means.

[2]There is also a two-step version of the technique in which the participation equation is first estimated alone, and the results are used to create a "selection bias correction" variable which is then entered into the outcome equation (1). Either technique can be used; they are equally acceptable for present purposes.

distribution (the probabilities of program participation are assumed to be normally distributed). However, to jointly estimate three participation equations (representing participation in three programs) requires the computation of a three-way, or trivariate, normal probability. Performing this computation is important because the unobservables for the three programs--ν_{1i}, ν_{2i}, and ν_{3i}--are expected to be correlated because the unmeasured influences of participation in one program are no doubt related to those that influence other programs. That is, even for families with the same income, potential benefit, household size, and other variables (that is, in Z), families that receive AFDC benefits are also very likely to receive FSP benefits, which would lead to a positive correlation between the propensity to participate in one program and the propensity to participate in another (or others).

When the three participation equations are estimated jointly with the outcome equation, four-way normal probabilities must be computed. Conventional computer techniques, which use types of "approximation" techniques for this evaluation, are not feasible for this large a compuation. McFadden (1989) and Pakes and Pollard (1989) have proposed an alternative method based on "simulation" techniques. The basic idea behind their method is as follows. In their proposed simulation method, the probabilities in any large set of equations such as ours are not mathematically approximated but are instead directly "simulated" by randomly generating values of the unobserved error terms on the computer. In our case, there are four such error terms (ϵ_i, ν_{1i}, ν_{2i}, and ν_{3i}). If these four error terms are normally distributed, then random, simulated "draws" must be taken from a four-way normal distribution. There are many "random number" generation methods available on all computers, and the creation of a large number of random "draws" from a four-way normal distribution, though not difficult, is moderately computer-intensive depending the number of random draws taken. Following this, a beginning, "trial" set of values is chosen for each of the coefficients in equations (1)-(4)--namely, β, α_1, α_2, α_3, γ_1, γ_2, and γ_3. For each set of draws of the error terms (i.e., for each set of four, one for each of the error terms), the values of the dependent variables--Y_i,

P_{1i}, P_{2i}, and P_{3i}--are determined for each family i by plugging into equations (1)-(4) the values of the independent variables for that individual, that family's draw of the four error terms, and the trial values of the coefficients. In our case, it is determined whether each family would or would not participate in each of the three programs as well as the value of Y_i. Once this determination has been made for each of a number of random draws (for example, 10, 20, or 100 sets of the four error terms), the fraction of the draws that result in each family being a "participant" is computed, and this value is used as the estimate of the probability that that family would participate. Thus, the probability that each family would participate is "simulated" by counting the number of times it would participate if its unobserved determinants (i.e., the four error terms) took on a randomly-drawn set of different values, values which we cannot observe but can simulate.

Once these probabilities are determined for a single trial value of the coefficients, the estimation of those coefficients proceeds by iteration as it does for maximum-likelihood estimation in the single-program case. A systematic search is taken over all possible values of all of the coefficients, and the set that generates predicted probabilities that are the closest to the probabilities observed in the data (i.e., which best "fit" the data) are chosen as the estimated coefficients. In the simulation method, this implies that the predicted probabilities for all families in the data set must be simulated for different possible values of all the coefficients.

Since the new method is designed to directly address a computational problem with existing methods, its success or failure must depend on whether it is computationally feasible and not burdensome. A liability of the new method is its computationally intensive requirement that repeated draws from a normal distribution must be generated to simulate the probabilities, a process that must be performed for each family and for a wide set of coefficient values. To determine the feasibility of the method, we have implemented it on the SIPP database and we have estimated the simple model described in the next section. As we shall discuss, we find the technique to be very feasible and not particularly burdensome for the four-equation case shown at the beginning of this section.

III. AN ILLUSTRATION WITH SIPP

We have implemented the MSM technique using the fourth wave of the first panel of the SIPP, which was administered in the fall of 1984. This data set was used by Fraker and Moffitt (1988) to study the effect of two programs, AFDC and FSP, on the labor supply of female heads of families. We use the same sample of female heads, but we analyze the effect of three programs--FSP, AFDC, and public housing--on rental expenditures instead of labor supply.[1] We use rental expenditures for three reasons: (1) the SIPP data do not include information on food expenditures (an outcome of greater interest than rental expenditures to FNS), (2) rental expenditures is more purely an "expenditure" variable than is labor supply, and (3) the distribution of rental expenditures is continuous, rather than having a concentration at zero, as is the case with labor supply.[2]

We use all female heads ranging in age from 18 to 64 years who have children younger than 18 present in the family. We exclude families with assets in excess of $4,500 because they are far above the program asset limits, and their behavior is likely to be very different from families with lower assets levels. There are 968 female heads in the sample. The reference month for the measurement of participation in the three programs--AFDC, FSP, and public housing (the last includes Section 8 housing)--is the month prior to the interview.

In the sample, 53 percent of the female heads do not participate in any of the three programs. About 30 percent participate in AFDC, and 40 percent participate in FSP. These participation rates are somewhat lower than participation rates calculated in other studies because we do not exclude all ineligibles--only those with high assets, as mentioned above. Twenty-six percent of the female heads participate in both AFDC and the FSP, which implies that virtually all women who receive

[1] Rental expenditures are imputed for those who are homeowners.

[2] All of the female heads have either a reported or an imputed rental expenditure, but not all of them work. Those who are not employed have zero hours of market labor.

AFDC also receive FSP benefits, and that over half of those who receive FSP benefits also receive AFDC. Thus, as is well known, participation in the two programs is strongly correlated.

About 17 percent of the sample participates in public or subsidized housing. About half of these cases also receive both AFDC and FSP benefits. Less than one-fifth of the cases that participate in public or subsidized housing receive only one of the two other kinds of benefits.

Table III.1 shows the results of an estimation of a model with the three participation equations only--no equation for rent is included. We show this model because in potential future applications it is likely to be of interest to estimate only those equations, and because we wish to examine the computational burden of such estimation by itself.[3]

The results in Table III.1 were obtained using 20 "draws," or simulations, of the three errors terms v_{1i}, v_{2i}, and v_{3i} [see equations (2)-(4)].[4] The run times for this model are given below. As the table shows, the estimates indicate that the potential AFDC benefit has a positive effect on AFDC participation, and the potential FSP benefit has a positive effect on FSP participation. However, the potential benefit in public or subsidized housing has no effect on participation in such housing. We interpret this as evidence that public or subsidized housing is rationed and not an entitlement program. The hourly wage rate has a negative effect on participation in all three programs, although the effect is again insignificant for housing.[5] Nonlabor income has a significantly negative effect on participation probabilities in all three equations. The other coefficients show that education, age, living in the South, and being white generally have negative, although not always significant, effects on participation. The number of children younger than 18 has a positive effect

[3]Such a model would be of interest, for example, in an analysis of participation in multiple assistance programs.

[4]That is, 20 sets of the three error terms were drawn for each of the 968 female heads in the sample.

[5]The wage rate for nonworkers was obtained from predictions from the wage equation reported in Keane and Moffitt (1991).

TABLE III.1

RESULTS OF THE ESTIMATION: THREE PARTICIPATION EQUATIONS ONLY

	AFDC Participation Equation	FSP Participation Equation	Housing Participation Equation
Program Benefit[a]	.065 *	.032 *	-.014
	(.011)	(.019)	(.016)
Hourly Wage Rate	-.151 *	-.108 *	-.082
	(.058)	(.058)	(.067)
Nonlabor Income[b]	-.058 *	-.068 *	-.057 *
	(.011)	(.009)	(.011)
Education	-.045	-.067 *	-.008
	(.029)	(.029)	(.034)
Age	-.026 *	-.023 *	-.019 *
	(.006)	(.006)	(.007)
South Dummy	.004	-.220 *	-.015
	(.086)	(.069)	(.086)
No. Children Younger Than 18	.188 *	.201 *	-.203 *
	(.045)	(.069)	(.061)
White Dummy	-.448 *	-.474 *	-.719 *
	(.067)	(.066)	(.081)
Constant	1.250 *	1.939 *	.868 *
	(.333)	(.312)	(.434)
Correlation Coefficients:			
Between AFDC and FSP			.946 *
			(.012)
Between AFDC and housing			.429 *
			(.037)
Between FSP and housing			.407 *
			(.038)

NOTE: Standard errors in parentheses.

[a]Weekly. Measured at zero hours of work. Coefficient is multiplied by 10.

[b]Weekly. Coefficient is multiplied by 10.

*Statistically significant at the 90 percent level.

on participation in AFDC and the FSP, but a negative effect on housing participation; the reasons for this are unclear.

The correlations between the error terms in the participation equations are shown at the bottom of the table. Strong positive correlations are observed, especially between the error terms in the AFDC and the FSP participation equations.

Table III.2 shows estimates of the full model, including the rent equation (ignoring, for the moment, the last column). The coefficients on the variables in the participation equations are generally of the same sign and significance as reported in Table III.1, which should be the case since there is no "feedback" from the rent equation to the participation equations in this simple model. In the rent equation, rental expenditures are seen to be positively affected by the wage rate and nonlabor income. Moreover, those expenditures are positively affected by the amount of program benefits received from each of the three programs. The error terms in the participation equations are positively correlated with each other, but are negatively correlated with the error term in the rent equation. All of the correlation coefficients are statistically significant. Thus, female heads with higher rental expenditures are less likely to participate in these programs.[6]

These last correlations are important because they are an indication of self-selection bias. The fact that they are significant implies that self-selection bias is present. In addition, their negative values indicate the direction of such bias. Specifically, they indicate that families with low rental expenditures are more likely to participate in AFDC, FSP, and housing programs *independent* of the direct effects of benefits in those programs. Thus, the types of recipients in these programs are "self-selected" by their rent levels. This suggests, in turn, that a simple comparison of rent levels of

[6] As the table indicates, only one variable (number of children younger than 18) appears in the participation equation and not in the rent expenditure equation. Preferably, there should be three such variables, one for each equation. In addition, it is certainly possible that this particular variable has direct effects on rental expenditure, in which case a different type of variable should be used. One category of variables that might be appropriate is that which consisting of variables that affect the "costs" of participation, such as the "access" variable we mentioned previously in the report. Unfortunately, our data set contains no direct measures of access or other cost.

TABLE III.2

RESULTS OF THE ESTIMATION: THREE PARTICIPATION EQUATIONS AND RENT EQUATION

	AFDC Part. Eqn.	FSP Part. Eqn.	Housing Part. Eqn.	Rent Eqn.	OLS Rent Eqn.
Program Benefit[a]	.081 *	.054 *	-.038 *	--	--
	(.011)	(.020)	(.017)		
Hourly Wage Rate	-.308 *	-.270 *	-.208 *	15.274 *	5.899 *
	(.067)	(.065)	(.074)	(2.323)	(.1950)
Nonlabor income[b]	-.042 *	-.060 *	-.056 *	1.321 *	.261
	(.013)	(.010)	(.011)	(.231)	(.195)
Education	.046	.027	-.035	-4.310 *	-.690
	(.033)	(.032)	(.037)	(1.162)	(.976)
Age	-.012 *	-.009	-.001	-.777 *	-.401 *
	(.007)	(.007)	(.008)	(.248)	(.209)
South Dummy	-.108	-.364 *	-.023	-5.802 *	-5.837 *
	(.098)	(.082)	(.096)	(3.092)	(2.678)
No. Children Younger Than 18	.207 *	.195 *	-.142 *	--	--
	(.044)	(.053)	(.052)		
White Dummy	-.312 *	-.353 *	-.537 *	10.572 *	7.461 *
	(.081)	(.079)	(.092)	(3.016)	(2.572)
SMSA Dummy	--	--	--	6.301 *	10.015 *
				(2.007)	(2.491)
Fair Market Rent in Area[c]	--	--	--	.249	2.656 *
				(.488)	(.531)
AFDC Benefit	--	--	--	2.195 *	-.043
				(.293)	(.377)
FSP Benefit	--	--	--	1.774 *	-2.417 *
				(.442)	(.564)
Housing Benefit	--	--	--	2.725 *	-1.172
				(.215)	(.271)
Constant	.272	.953 *	.458	45.510 *	16.346
	(.375)	(.351)	(.476)	(14.086)	(12.599)
Correlation Coefficients:					
Between AFDC and FSP				.962 *	
				(.010)	
Between AFDC and housing				.450 *	
				(.045)	
Between FSP and housing				.500 *	
				(.044)	

TABLE III.2 (continued)

	AFDC Part. Eqn.	FSP Part. Eqn.	Housing Part. Eqn.	Rent Eqn.	OLS Rent Eqn.
Between AFDC and rent				-.706 * (.026)	
Between FSP and rent				-.653 * (.027)	
Between housing and rent				-.771 * (.026)	
Standard deviation of error term in rent equation				42.341 * (.942)	

NOTE: Standard errors in parentheses.

[a]Weekly. Measured at zero hours of work. Coefficient is multiplied by 10.

[b]Weekly. Coefficient is multiplied by 10.

[c]Coefficient is multiplied by 10.

OLS = ordinary least squares.

*Statistically significant at the 90 percent level.

participants and nonparticipants is likely to show lower rent levels for participants, which might be mistakenly interpreted as a negative effect of participation on rental expenditures.

This suggestion is confirmed by the last column in Table III.2, which shows ordinary least squares regression estimates of the rent equation without any control for self-selection bias. The coefficients on all three benefit levels are in this case negative, and one of these coefficients is statistically significant. As a result, misleading conclusions would have been drawn from such estimates of the rent equation.

Table III.3 shows the run times for various models and provides evidence that it is computationally feasible to estimate these models using modern computers. The computer used for the estimation was a mainframe Amdahl, close in capability to a standard IBM mainframe. Microcomputers with 386 and 486 chips are somewhat slower than such mainframes but not so much as to make the times shown in the table unrepresentative. The first two rows of the table show the CPU minutes required for estimating the three participation equations only, but without any independent variables--that is, only with intercepts. We did not present the results of these estimates earlier because they are of no substantive interest; however, they do permit us to determine the effect of the independent variables themselves on run times. As the table shows, the run time for the intercept-only models was only 1.5 - 3.0 minutes, and the run time for the model consisting of the three fully specified participation equations (that is, the model shown in Table III.1) was much more--16.8 minutes. Therefore, the independent variables do indeed constitute most of the run time. When the rent equation is added, the run time is about 30 minutes of CPU time. This run time is well within the capability of most mainframes and most 386 and 486 micros as well.

The models estimated in Table III.3 were estimated sequentially, starting with the model in the first row and then proceeding to the model in the next row. The "starting values" for each row were obtained from the estimates obtained from the simpler model in the previous row. For this reason, perhaps a more accurate estimate of the total run time for each model would be the sum of the run

TABLE III.3

RUN TIMES FOR VARIOUS SPECIFICATIONS
OF THE MODEL

	CPU Minutes per Iteration	Approx. Total CPU Minutes	Cumulative Run Time
Three Participation Equations Only			
Intercepts only, no correlations	0.15	1.50	1.50
Intercepts only, correlations	0.24	3.00	4.50
Full specification, with correlations	0.85	16.80	21.30
Three Participation Equations plus Rent Equation	1.20	28.80	50.10

NOTE: CPU times are for an Amdahl mainframe roughly equivalent in power and speed to the IBM 3090 series.

times for that model and the previous ones. This is shown in the final column of the table as cumulative run times. For the final model, this cumulative run time is about 50 minutes. This is still within computational feasibility.[7]

Some experimentation was conducted on the number of "draws" required for estimation. The results presented in Tables III.1-III.3 are for 20 draws, a number determined by starting at a low number of draws and increasing that number until the estimates no longer "changed" with increasing numbers of draws. Different models estimated on different data sets may require more or less numbers of draws. We should note that the run time is roughly linear in the number of draws--that is, a model requiring 40 draws would require roughly double the CPU times shown in Table III.3, and a model requiring 10 draws would require roughly half the run times shown in the table.

Generalizability to Other Applications. The example we have illustrated here involves only three programs, and it involves a particular population group (female heads) and three particular programs (AFDC, Food Stamps, and public housing). Practical issues may arise when extending the technique to other applications.

One issue that might arise in other applications is the distribution of the sample across different program categories. In our SIPP data, a significant fraction of the sample participates in each of the three programs (30 percent in AFDC, 40 percent in FSP, and 17 percent in housing). Application of the technique to sets of programs where the sample is "thin" for some programs (e.g., less than 5 percent) may make estimation difficult. For example, studies of multiple program participation among husband-wife couples often suffer from small sample-size problems because there are some programs (e.g., AFDC-UP) for which their participation rates are quite small.

This problem is not unique to our estimating technique, for it arises in any participation study. However, it is more likely to arise here because multiple programs are considered and hence at least

[7] Each of the individual run time entries in the table is itself a sum of separate runs, each of which tried a set of "trial values" of all the coefficients, as described in the last section. Thus those run times represent how long it took to find the "best fitting" values of the coefficients for that model.

one of them may have a low participation rate. In addition, because our technique involves the estimation of the correlation of program participation, it implicitly requires sufficient numbers of households to participate in some combination of programs. This requirement may be difficult to meet in small samples.

A second issue of generalizability relates to the extension of the model to four programs. First of all, the run times given above are not linear with respect to the number of programs involved. We have illustrated only three programs, but the software we provide is capable of accommodating from one to four programs. Each additional program participation equation increases the run time more than proportionately because additional correlations and forms of self-selection bias must be estimated. In addition, the small sample size problem mentioned previously may make estimation with four programs difficult. If, for example, multiple program participation among AFDC, FSP, WIC, and either SBP or NSLP were considered, it is possible that samples might be quite small for some of the programs and some of the combinations.

Finally, we might note that the variable used as the dependent variable in the "outcome" equation does not affect the run times. Hence, using food expenditures instead of rent, for example, should have no effect on these computational results.

IV. SUMMARY

In this report we described a new method for handling the problem of self-selection bias in the context of estimating the effects of a single assistance program when there is multiple program participation. We also summarized the results of applying this program. The new method was applied to the SIPP, and a four-equation model consisting of three participation equations and one outcome equation was successfully estimated. The computational burden of the estimation is more than that associated with ordinary methods, but it is still well within the power of modern mainframes and high-powered microcomputers. The evidence we report is therefore favorable, and the technique appears to be suitable for application to problems involving self-selection bias for FSP recipients. We note that application of self-selection adjustment methods in general, as well as our method, requires the data set to contain variables that affect program participation but which do not directly affect the outcome variable of interest. We recommend that when data containing such variables but containing information on food expenditures or diet quality become available, program effects on those outcomes to be estimated with our proposed technique.

At the time of this writing, the data set most likely to be useful for these techniques is the 1989-91 CSFII, which has information on household food expenditures and individual food intake. The CSFII has approximately 1600 households in the low-income sample and 3500 in the population sample, which should be enough to generate sufficient numbers of observations in the major programs (FSP, AFDC, and perhaps WIC, SBP, and NSLP) with which FNS is concerned. The sample size may not be large enough to permit estimation of four separate participation equations (i.e., four programs), however, an issue we discussed previously. Another possible data set is the 1996 survey of food use currently under discussion, which will have information on household food use on a low-income sample of approximately 5000 households.

REFERENCES

Devaney, B., and R. Moffitt. "Dietary Effects of the Food Stamp Program." *American Journal of Agricultural Economics*, February 1991, pp. 202-211.

Fraker, T., S.K. Long, and C.E. Post. "Assessing Dietary Adequacy and Estimating Program Effects: An Application of Two New Methodologies Using FNS's Four-Day File for the 1985 CSFII." Washington, DC: Mathematica Policy Research, 1989.

Fraker, T., and R. Moffitt. "The Effect of Food Stamps on Labor Supply: A Bivariate Selection Model." *Journal of Public Economics*, vol. 35, 1988, pp. 25-56.

_____. "The Effect of Food Stamps on the Labor Supply of Unmarried Adults without Dependent Children." Washington, DC: Mathematica Policy Research, 1989.

Heckman, J. J. "Sample Selection Bias as a Specification Error." *Econometrica*, vol. 47, January 1979, pp. 153-161.

Keane, M. "A Computationally Practical Simulation Estimator for Panel Data." Mimeographed, University of Minnesota, 1990.

Keane, M., and R. Moffitt. "A Structural Model of Multiple Welfare Program Participation and Labor Supply." Mimeographed, University of Minnesota, 1991.

Long, S.K. "The Impact of the School Nutrition Programs on Household Food Expenditures." Washington, DC: Mathematica Policy Research, 1988.

McFadden, D. "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration." *Econometrica*, vol. 57, September 1989, pp. 995-1026.

Maddala, G.S. *Limited-Dependent and Qualitative Methods in Econometrics*. New York: Cambridge University Press, 1983.

Pakes, A., and D. Pollard. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica*, vol. 57, September 1989, pp. 1027-1057.

Contract No.: 53-3198-9-31
MPR Reference No.: 7890-009

SOFTWARE DOCUMENTATION FOR PROGRAMS TO ESTIMATE SELF-SELECTION MODELS WITH MULTIPLE EQUATIONS USING THE METHOD OF SIMULATION

December 3, 1992

Supplement to Report Entitled:

"THE ESTIMATION OF FOOD STAMP SELF-SELECTION MODELS USING THE METHOD OF SIMULATION"

Authors:

Michael Keane
University of Minnesota

Robert Moffitt
Brown University

Submitted to:

U.S. Department of Agriculture
Food and Nutrition Service
3101 Park Center Drive
Alexandria, Virginia 22302

Project Officer: Christine Kissmer

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Avenue, S.W.
Suite 550
Washington, D.C. 20024

Project Director: Thomas Fraker

CONTENTS

I. INTRODUCTION

This document provides instructions for the use of the programs SIMA.FOR and SIMB.FOR to estimate models of multiple welfare program participation, with or without an extra equation for an outcome variable. This document and the two programs are provided as supplements to the final report to FNS entitled "The Estimation of Food Stamp Self-Selection Models Using the Method of Simulation" by Michael Keane and Robert Moffitt (1992), submitted by Mathematica Policy Research. That report should be read prior to reading this document.

The statistical models estimable with the programs SIMA.FOR and SIMB.FOR are documented in detail in Appendix A of this report. They are also presented in the aforementioned final report to FNS. The model in SIMA.FOR consists of up to four welfare program participation equations.[1] Each equation can have different independent variables. The model in SIMB.FOR permits the addition of an outcome equation with a continuous dependent variable (e.g., food expenditures or some other variable that may be affected by program participation). That equation can contain regressors that do or do not overlap with those in the participation equations. In FNS applications, this equation will often contain the program benefit(s) and/or program participation dummy variables.

The error terms in the participation equations, and in the extra outcome equation if added, are assumed to be distributed according to a multivariate normal distribution with an unrestricted covariance matrix. If the extra outcome equation is added, this implies that the program fully accounts for the correlations that induce selection bias in that equation.

The programs employ the method of simulated moments (MSM) to estimate the models.[2]

---

[1] If additional programs are needed, the program can be adapted for that purpose. A number of the matrices in the Fortran program would have to be increased in dimension.

[2] McFadden, (1989) describes the method of simulated moments.

II. FILES USED IN ESTIMATION

The programs SIMA.FOR and SIMB.FOR are the major files used in the estimation of the two types of models of that were introduced in the previous chapter. Both are written in Fortran. An outline of the structure of the programs is given in Appendix B.

Both programs require that several input, output, and working files be opened on a disk or another medium. The OPEN statements in lines 77-82 of SIMA.FOR and lines 83-88 of SIMB.FOR must be set by the user to denote the locations of these files. Two input files must be made available:

1. ***INA.DAT and INB.DAT.*** These files contain user-set values of parameters that govern the iteration process, that determine which variables are included in the equations, and that supply starting values for the coefficients and other parameters to be estimated. These files are discussed in more detail in Section III.

2. ***DATA.FIL.*** A file containing the data set used in the estimation with this or another user-set name must be supplied.

The user must set the format for reading in DATA.FIL and must insert Fortran code to construct the variables. The section of SIMA.FOR where this insertion must be made begins on line 886, while the corresponding line in SIMB.FOR is 956. Each record of the data must supply the values of the dependent variables and independent variables to be used in the analysis. If SIMB.FOR is used, the record must also indicate whether the value of the outcome variable is or is not observed for that observation. This section may also be used to impose sample screens and exclusions.

The Fortran programs write up to three output files:

3. ***OUT6.DAT.*** A file containing all the printed output from the program, plus any machine statements or Fortran error messages. This file is assigned to device 6, which is ordinarily the default print device. If the user wishes the output to be printed according to the default on his or her machine rather than OUT6.DAT, the OPEN statement in the program for unit 6 can be deleted or commented out.

4. ***OUTPUT.DAT***. A file containing all the printed output from the program, but no machine statements or Fortran error messages.

5. ***PARAMS.OUT***. A file containing the final parameter values estimated by the run. The format for this file is identical to that of INA.DAT and INB.DAT, so that PARAMS.OUT can simply be renamed INA.DAT or INB.DAT and used to start another run of the Fortran program, using the parameter values in PARAMS.OUT as starting values for the next run. If the run of the Fortran program does not terminate normally, PARAMS.OUT may not be printed.

In addition to these files, the programs require that disk or other space be allocated for one working file, WORK.FIL, denoted as device 12. This working file is used to store values of certain parameters (see Appendix B). The files can be disposed of after estimation. Users who wish to hold these parameters in memory rather than on disk or other medium may modify SIMA.FOR and SIMB.FOR accordingly. In general, holding the files in memory is likely to decrease run time because I/O time is reduced.

The Fortran programs also require that a mathematical library be accessed containing subroutines to invert matrices and to draw unit normal random deviates. The programs as written use the LINPACK routines DPOFA, DGEFA, and DGEDI, and IMSL routine GGNML. The user may modify the relevant CALL statements if different routines are desired.

III. INA.DAT AND INB.DAT

A. INA.DAT

A sample INA.DAT file is included on the disk with the programs.

The first line of INA.DAT leaves room for a up-to-60-character user-supplied title for the run.

The second line of INA.DAT contains pre-set labels that do not have to be reset by the user.

The third line of INA.DAT contains parameters governing the iteration and estimation. The label for each parameter is given in line 2 of the file just above the location of the parameter to be set. The parameters to be set are the following:

- NPROG: The number of participation equations (=1, 2, 3, or 4).
- NITER: The total number of parameters to be estimated on the run. This includes the sum of all coefficients in all participation equations to be estimated (including intercepts), plus those of the correlation parameters across the equations to be estimated. The number of correlation parameters in a system of NPROG participation equations is NPROG*(NPROG-1)/2.
- IND: The number of independent variables used in the estimation, including all variables used in any of the equations. In the notation of Appendix A, IND is the number of variables that appear at least once in Z_1, Z_2, Z_3, or Z_4.
- NDRAW: The number of simulated draws per observation.
- MAXIT: The maximum number of iterations allowed.
- SSIZE: The beginning step size, usually set at 1.0 (see Section IV).
- TRANS: A character variable set equal to "YES" if the correlation parameters are to be transformed to a (-1,+1) interval and "NO" if not. Ordinarily this variable is set at "YES" for all iterations except the final run producing the final set of coefficient estimates (see Section IV).

The format for reading these parameters is (2x,I1,3x,I3,4X,I2,2X,I3,2X,I3,4X,F4.2,1X,A3).

The arrays in SIMA.FOR are currently dimensioned for a maximum of NPROG=4, IND=30, NITER=80, and NDRAW=100. The user may wish to expand the dimensions if these maxima are restrictive. There is no limit on the number of input observations as the program is currently written.

The next section of INA.DAT contains lines with the starting values of the parameters to be estimated. Input formats are given after line 129 in SIMA.FOR. The starting values for the coefficients of the first participation equation are entered first, followed by the starting values of the coefficients of the second participation equation, and so on. An intercept is ordinarily one of the parameters. Following the starting values for the coefficients come the starting values for the correlation parameters across the equations in the order (1,2), (1,3), (2,3), (1,4), (2,4), and (3,4). Following these parameters must be set a value for RHO, a smoothing parameter (see Section IV and Appendix A). RHO should be set to be a "small" number such as .10.

The next section of IN.DAT contains lines indicating the names of the independent variables to be entered into any equation of the model. The names for the variables in the first participation equation are followed by those for the variables in the second equation, and so on, followed at the end by the names of the correlation parameters and the name of the smoothing parameter (RHO). The number of names in every participation equation must equal IND, the parameter set previously. The order in which the variable names are listed must correspond to the order of the IPARM parameters discussed momentarily and they must correspond to the order in which the variables are aligned in the variable vector DATA (to be supplied by the user in the data entry section of the Fortran program).

The next section of INA.DAT contains lines indicating which of the independent variables are to be included in each equation of the model and, simultaneously, which parameters of the model are to be iterated on (these variables are denoted as "IPARM" variables in the Fortran code). The first line of this section corresponds to the first participation equation, the second line corresponds to the second participation equation, and so on, followed by lines for the correlation parameters and the smoothing parameter. Each line of this section for a participation equation contains a "1" if the corresponding variable in the previous listing of variable names is to be included in that equation, and "0" if not. Each line therefore will contain IND characters, each being either "1" or "0". The lines

for the correlation parameters and the smoothing parameters are set equal to "1" if the user wishes to iterate on them and "0" if not. Ordinarily, the smoothing parameter indicator is set at "0".

B. INB.DAT

INB.DAT is for the most part quite similar to INA.DAT. All remarks on each section of the discussion of INA.DAT apply, but with the following additions:

- The parameter NITER must include the coefficients to be estimated in the outcome equation as well as those in the participation equations; the correlation coefficients, whose number is NPROG*(NPROG+1)/2; and the standard deviation of the error term in the outcome equation.
- The parameter IND must be set at the number of independent variables appearing in either the outcome or any of the participation equations.
- The starting values for the outcome-equation coefficients follow those of the participation equations immediately, as do their variable names and indicator (IPARM) values.
- An extra parameter (SIGMA) for the standard deviation of the error term in the outcome equation is introduced. Its starting value, name, and indicator (IPARM) value are located as indicated in the sample INB.DAT.

IV. REMARKS ON USAGE

Initial starting values of parameters are to be provided by the user. Initial starting values for the parameters in the outcome equation can be obtained from OLS estimates and initial starting values for the parameters in the participation equations can be obtained from single-equation probit estimates; the initial starting values for the correlation coefficients can be set at zero. Alternatively, the user may wish to start the parameters at values chosen by some other method. Our experience in the one application of this program is that initial starting values too far from plausible values can result in a failure of the program to iterate normally. The user is advised to begin by estimating a "small" model with as few parameters as possible, and to build up the model slowly by adding parameters. For example, beginning by including relatively few variables in the equations and by holding the correlation coefficients fixed at zero often provides adequate initial estimates.

The program is designed to be run repeatedly, sequentially copying the estimates written out to PARAMS.OUT from a particular run into INA.DAT or INB.DAT, and rerunning the program.

The smoothing parameter can be varied to detect sensitivity of the results. Because substantial bias may arise if the smoothing parameter is too large, it is to be preferred to set this parameter as close to zero as possible. However, setting it exactly equal to zero will prevent the program from running. We have found in our applications that a value of near 0.10 is a satisfactory compromise between these two considerations.

The number of draws must be determined by the user. Although MSM estimates are consistent in sample size for a fixed number of draws (even one draw), efficiency gains can be achieved by using more draws. Typically, efficiency gains will be negligible beyond 20 or 30 draws. In practice, the user may wish to set the number of draws at a low number (e.g., 10) for early iterations, and to increase this number to 20 or 30 once the estimation is near convergence.

The initial step size can be set at a lower value than 1.0, such as .2 or .1, if initial iteration from any particular set of starting values proves difficult. A lower step size will permit the program to move the parameters by only small amounts on each iteration, which sometimes provides a more stable iteration path. As iteration proceeds more normally, the step size can be reincreased to 1.0.

The TRANS parameter set to "YES" forces the correlation coefficients to stay in the proper range of (-1,+1), a range that may be violated otherwise if the program tries a value outside that range. After final estimates are obtained, the TRANS parameter should be set equal to "NO" for one last run in order to obtain correct standard errors on the parameters.

The convergence criteria in the programs require that the objective function in the problem be close to zero (viz., less than 1.0) in order for convergence to be declared.[1] As a practical matter, whether this low a value of the objective function can be achieved will depend on the application at hand as well as rounding error. The user may wish to use one of the many other convergence criteria available in the numerical optimization literature, or to simply define convergence as having been achieved when the iterations fail to move the parameters and their standard errors over a significant number of iterations.[2]

The program prints out two estimates of the asymptotic covariance matrix of the parameters and two chi-squared statistics. McFadden (1989) shows that the two covariance matrix estimates are asymptotically equivalent and that the second one that is printed out approaches the first one asymptotically from below. If the sample size and number of draws are large enough that the asymptotic formulas are reliable, the two matrices should differ only by a few percent. As for the two chi-squared statistics, both are Pearson statistics based on goodness-of-fit but the first uses the estimated probabilities in the denominator of the statistics while the second uses the actual

---

[1]See line 770 of SIMA.FOR and line 779 of SIMB.FOR. The objective function for any method-of-moments problem is the sum of the squared first-order conditions; the estimation procedure seeks to minimize this function.

[2]As in any optimization problem, different starting values should be tried to ensure that a global optimum has been achieved.

probabilities in the denominator. Because actual probabilities are sometimes small and give very high chi-squared statistics, we provide the first chi-squared statistic for more stable estimates.

REFERENCES

Keane, Michael and Robert Moffitt. "The Estimation of Food Stamp Self-Selection Models Using the Method of Simulation." Washington, DC: Mathematica Policy Research, 1992.

McFadden, D. "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration." *Econometrica*, vol. 57, September 1989, pp. 995-1026.

Marquardt, D.W. "An Algorithm for Least Squares Estimation of Nonlinear Parameters." *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, 1963, pp. 431-441.

APPENDIX A

STATISTICAL MODELS AND ESTIMATION METHOD

A. SIMA.FOR

The statistical model estimated in SIMA.FOR is the following, for the case of the maximum of four participation equations:[1]

$$(1)\quad P_{1i}^{*} = Z_{1i}\,\gamma_1 + \nu_{1i}$$
$$(2)\quad P_{2i}^{*} = Z_{2i}\,\gamma_2 + \nu_{2i}$$
$$(3)\quad P_{3i}^{*} = Z_{3i}\,\gamma_3 + \nu_{3i}$$
$$(4)\quad P_{4i}^{*} = Z_{4i}\,\gamma_4 + \nu_{4i}$$

where:

P_{ji}^{*} = latent indicator for individual i, whose sign determines a binary choice for program j (j=1,...,4)

Z_{ji} = regressor variables in equation j

γ_j = coefficients in equation j

ν_{ji} = error term in equation j

The parameters are estimated by the method of simulated moments. The four error terms are assumed to be distributed multivariate normal with unit variances and an unrestricted correlation matrix. There are 2**4=16 alternative combinations j created by the four choices. The first-order conditions (FOC) for the problem are the derivatives of the log-likelihood function with respect to the unknown parameters in that function, and are given by:

$$(5)\quad \text{FOC} = \sum_{i=1}^{N} \sum_{j=1}^{16} \left[d_{ij} - P(j \mid \theta, U_i)\right] W_j(\theta, U_i)$$

[1] These equations are written in cross-sectional form but could be applied to panel data as well if the panel has four or fewer time periods. In such an application, each dependent variable in (1)-(4) would be a latent index for an individual i in wave j of the panel.

where d_{ij} is a dummy equal to one if individual i chooses combination j; $P(j|\theta,U_i)$ is the probability of choosing combination j conditional on the union of all four observed regressor sets (U_i) and the parameter vector θ, which consists of all parameters in the problem; and W_j is a weighting matrix. Following McFadden (1989), we use an unbiased simulator $f(j|\theta,U_i)$ for $P(j|\theta,U_i)$ and we also simulate the optimal weighting matrix, which consists of gradients of the log probabilities. When these optimal weights are simulated, the MSM estimator is asymptotically as efficient as maximum likelihood (as the number of draws is increased, that is). We minimize the objective function corresponding to (5) by the Gauss-Newton method as modified by Marquardt (1963). The random normal deviates drawn are held fixed throughout each run, and are drawn separately for the construction of the FOC and the weighting matrix. We also implement a logit smoothing technique suggested by McFadden which adds a extreme-value error term with a coefficient r onto each of the equations in the model, so that the probability of the individual choosing a one or zero for each alternative is a logit. As $r \to 0$ this model approaches the probit model. McFadden suggests that the smoothing parameter r be set as close to zero as possible in light of the multivariate normality assumption.

B. SIMB.FOR

The model in SIMB.FOR consists of equations (1)-(4) plus the equation:

(6) $Y_i = X_i \beta + \epsilon_i$

The error term ϵ_i is assumed to be jointly normally distributed with the error terms in (1)-(4), thereby permitting a full representation of selection bias. The vector X_i may include participation dummies or program benefits. The value of Y_i may not be observed for all observations in the sample. The program permits such observations to be noted.

The estimation of the model proceeds with MSM as before. The only difference is that two moments are added to (5), namely, $(y_i - X_i\beta)$ and $-(y_i - X_i\beta)/\sigma^2$, where σ is the standard deviation of ϵ_i. These are the first two moments of the estimated residual for ϵ_i. When combined with

appropriate weights, the derivatives of the log-likelihood function for the model including (1)-(4) plus (6) has the form of (5) expanded to include these additional weighted residuals.

As mentioned previously, the program SIMB.FOR permits Y_i to be unobserved for part of the sample. We wish to note that the two new residuals have zero mean values only in the total population, i.e., only if Y_i is observed for all $_i$. If the Y_i are observed only for some subsample, and if selection bias is present so that $E(\epsilon_i | Y_i \text{ observed}) \neq 0$, then the residuals in the selected sample no longer have zero mean values. Without the mean-zero property, the estimator in the program is no longer a method-of-moments estimator. Instead, since the program simulates the the derivatives of the log-likelihood function in any case, the estimator in the program is a simulated maximum likelihood estimator. That estimator is consistent only as the number of draws grows large. Hence users with partial observability for Y_i should expect to use many more draws than if Y_i is completely observed.

APPENDIX B

STRUCTURE OF THE FORTRAN PROGRAMS

The Fortran programs are set up with a MAIN routine and several subroutines. MAIN opens all files, reads the parameters in INA.DAT or INB.DAT and writes them out, and calls the subroutine TLOOP. The subroutine TLOOP runs the iteration process. All relevant arrays are initialized, a seed for the random number generator is set, and iteration is performed by repeatedly calling the subroutine PLOOP at different parameter values and checking for convergence. The subroutine PLOOP loops through the data and estimates the probabilities as well as gradients w.r.t. the parameters by simulation (random number draws) at the parameter values set in TLOOP. The first pass through PLOOP uses the estimated probabilities to calculate the weighting matrix (see Appendix A), and the subroutine WEIGHT is called from PLOOP on that first pass; these weights are held fixed throughout the run and consequently the subroutine WEIGHT is not recalled. TLOOP subsequently repeatedly calls PLOOP to try different step sizes within each iteration and to update the parameters over iterations.

The elements of the weighting matrix calculated in the first pass, as well as probability gradients used in that calculation, are written out to an external file and read back in for each subsequent pass through the data. Users who wish to hold these numbers in memory rather than repeatedly use I/O operations may wish to modify the data accordingly.

The input data are also read in repeatedly on each call to PLOOP. The user may wish to modify the program to read in the data only once and to hold it in memory to reduce I/O time.

The random numbers drawn in the first pass through PLOOP are used to calculate the weights. The random numbers drawn on the second and subsequent passes are used for iteration. The same seed is used for the second and all subsequent passes, and hence the same random numbers are repeatedly drawn. The user may wish to modify the program to draw these numbers only once and to either hold them in memory to write them out to and later read them in from a temporary disk. Our experience is that redrawing the random numbers is as quick as the time-consuming I/O alternative, but this may vary by system.

THE SELECTION BIAS PROBLEM

IN THE EVALUATION OF FOOD PROGRAMS

ROBERT MOFFITT

BROWN UNIVERSITY

FNS

JUNE 9, 1993

I. THE GENERAL PROBLEM*

(A.) DEFINING WHAT WE ARE INTERESTED IN

- LET Y = OUTCOME VARIABLE OF INTEREST
  (FOOD EXPENDITURES, NUTRIENT AVAIL., ETC.)
- WANT TO KNOW "THE EFFECT OF PROGRAM X ON Y"
- COMPLICATION: SUPPOSE THE EFFECT IS DIFFERENT FOR DIFFERENT FAMILIES & INDIVIDUALS?
- LET

  Y_i^* = VALUE OF Y FOR FAMILY i IF NOT A PROGRAM PARTICIPANT, OR RECIPIENT

  Y_i^{**} = VALUE OF Y FOR FAMILY i IF IS A PROGRAM PARTICIPANT, OR RECIPIENT

- THEN THE EFFECT OF THE PROGRAM FOR FAMILY i IS:

$$\alpha_i = Y_i^{**} - Y_i^*$$

- BUT IF α_i IS DIFFERENT FOR DIFFERENT FAMILIES, WHAT POPULATION ARE WE INTERESTED IN?
  WHOLE U.S. POPULATION? NO, USUALLY.

---

* REFERENCE:
R. MOFFITT, "PROGRAM EVALUATION WITH NONEXPERIMENTAL DATA" EVALUATION REVIEW, JUNE 1991

- USUALLY WANT TO KNOW AVERAGE α OF THOSE PARTICIPATING IN THE PROGRAM:

$\bar{\alpha}^P$ = AVERAGE EFFECT OF PARTICIPANTS ("P")

= AVERAGE OF THE $(Y_i^{1*} - Y_i^{*})$'S OF ALL PARTICIPANTS

- PROBLEM: $\bar{\alpha}^P$ WILL CHANGE IF THE CASELOAD CHANGES, OR ITS COMPOSITION CHANGES, BECAUSE THOSE WHO ENTER OR LEAVE MAY HAVE DIFFERENT α_i'S

(B) THE SELECTION BIAS PROBLEM

- THE CONCEPTUAL PROBLEM IS THAT NEITHER α_i NOR $\bar{\alpha}^P$ CAN EVER BE MEASURED, EVEN IN PRINCIPLE, BECAUSE Y_i^{*} CAN NEVER BE OBSERVED: IT NEVER "HAPPENS" BECAUSE A PROGRAM PARTICIPANT IS, BY DEFINITION, ON THE PROGRAM. SO WE CAN NEVER KNOW WHAT THE Y_i FOR THAT FAMILY WOULD HAVE BEEN AT THAT EXACT SAME TIME AND WITH ALL OTHER FACETS OF THE FAMILY'S SITUATION THE SAME, HAD THE PROGRAM NOT BEEN AVAILABLE.

- ALL WE CAN DO IS ESTIMATE (I.E., GUESS) WHAT Y_i^* IS.

-THERE ARE FOUR GENERAL METHODS:

(1) ASSUME THAT THE Y_i OF NONPARTICIPANTS = Y_i^*'S WE WANT

(2) RANDOMIZATION, CONTROLLED EXPERIMENT: USE THE Y'S OF A CONTROL GROUP TO ESTIMATE Y_i^*

(3) USE THE Y_i'S OF PROGRAM PARTICIPANTS AT A PRIOR POINT IN TIME WHEN THEY WERE OFF THE PROGRAM TO ESTIMATE Y_i^*

(4) FIND A GROUP OF SIMILAR FAMILIES WHO WERE NOT EVEN OFFERED THE PROGRAM (OR WERE AT LEAST OFFERED A DIFFERENT SET OF PROGRAM CHARACTERISTICS) AND USE THEIR Y_iS TO ESTIMATE Y_i^*S (IDEALLY: FIND A GROUP THAT WOULD HAVE PARTICIPATED IF THEY HAD BEEN OFFERED IT).

-TAKE EACH ONE IN TURN. BUT REGARDLESS OF WHICH WE USE, WE WILL ESTIMATE

$\hat{\bar{\alpha}}^P$ = AVERAGE OF Y_i^*'S OF PARTICIPANTS
-AVERAGE OF ESTIMATED Y_i^* S

- $\hat{\bar{\alpha}}^P = \bar{\alpha}^P$ ONLY IF OUR ESTIMATE OF Y_i^* IS CORRECT, I.E., ONLY IF IT EQUALS WHAT THE Y OF PARTICIPANTS WOULD HAVE BEEN OFF THE PROGRAM

1.) NONPARTICIPANTS

WHICH ONES? VS. POP? NO. ELIGIBLE NON-PARTICIPANTS? BETTER. BUT: WHY DID SOME FAMILIES WHO ARE ELIGIBLE PARTICIPATE AND OTHERS DID NOT? IN ALL LIKELIHOOD, THEY DIFFER IN THEIR LEVELS OF Y_i^{0*}.

EX 1: THOSE WITH LOWER Y_i^* PARTICIPATE (THE "WORST OFF")

$\Rightarrow$ GET $\bar{\alpha}^P$ TOO LOW

EX 2: THE BETTER OFF PARTICIPATE

$\Rightarrow$ GET $\bar{\alpha}^P$ TOO HIGH

$\therefore$ THIS APPROACH IGNORES SELECTION BIAS

1-2) CONTROLLED EXPERIMENTS

- RANDOMIZATION ENSURES THAT Y_i^*'S OF CONTROL GROUP ARE APPROX SAME AS Y_i^*'S OF EXPERIMENTAL GROUP
- EXPERIMENTS DIFFICULT TO IMPLEMENT IN MANY CASES: COST, ETHICAL & PRACTICAL PROBLEMS IN DENYING SERVICES, ATTRITION, CROSS-OVERS, ETC.
- ALSO: POINT OF RANDOMIZATION IS IMPORTANT BUT MAY ALTER ESTIMATE AWAY FROM $\bar{\alpha}^P$.

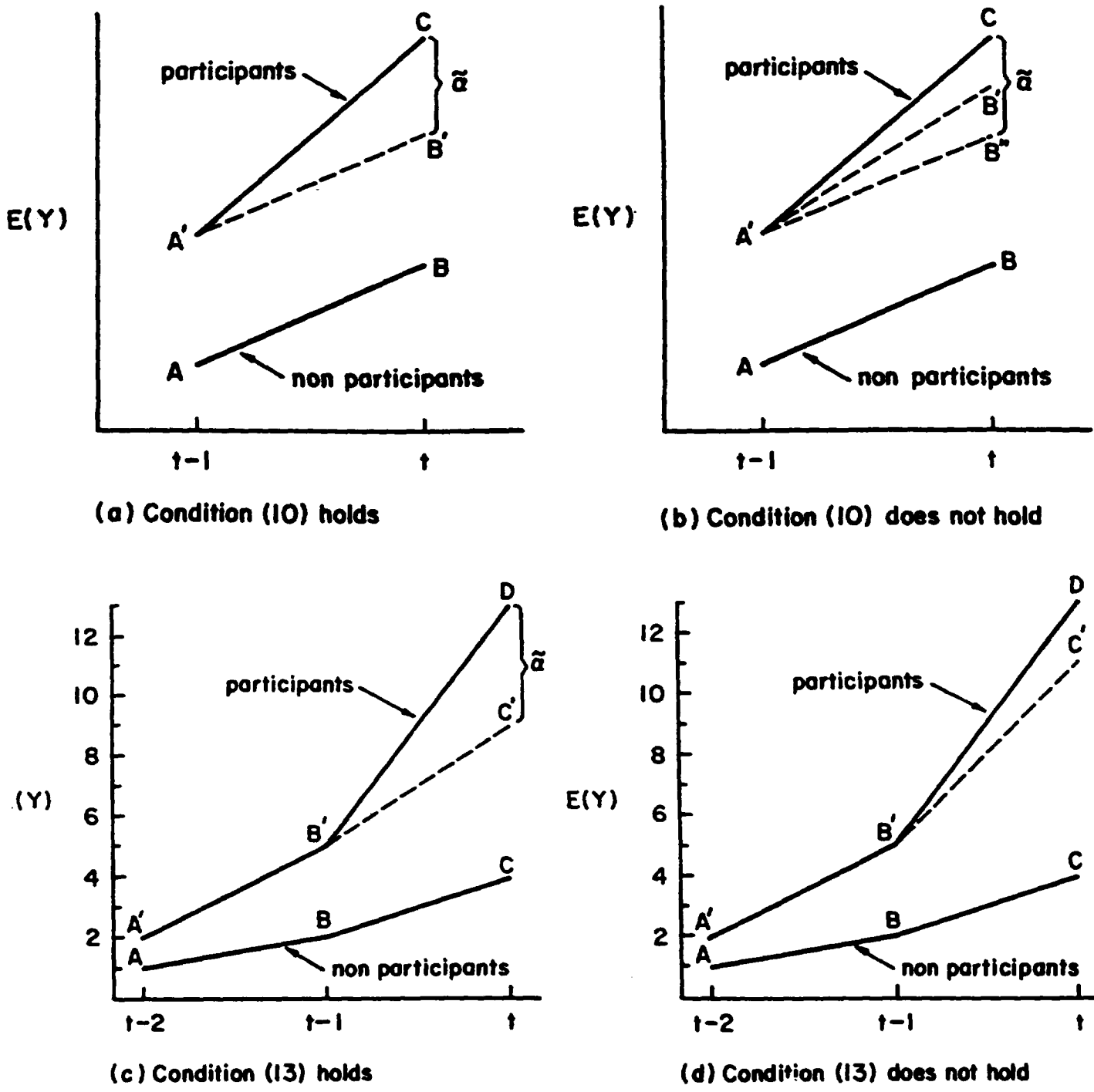
EXAMPLE:

FLIP COIN AT POINT OF APPLICATION OR CERTIFICATION. IF YOU INCREASE NO. OF CERTIFICATIONS IN ORDER TO KEEP THE # SERVED APPROX. UNCHANGED, $\bar{\alpha}^P$ MAY CHANGE BECAUSE THE "EXTRA," "NEW" PARTICIPANTS MAY HAVE DIFFERENT (LOWER?) α_i'S

(3) PRE-PROGRAM Y'S

$\Delta Y = Y^{**}_{it} - Y^{*}_{i,t-1}$: t = on program
$t-1$ = off program

- "Before + after"
- AT LEAST IT IS THE SAME PEOPLE
- BUT: OTHER THINGS MAY CHANGE FROM $t-1$ TO t (ECONOMY, FAMILY-SPECIFIC EVENTS, ETC)
- BETTER: COMPARE ΔY OF PARTICIPANTS (AT t) TO ΔY IF THOSE NOT ON PROGRAM AT $t-1$ OR t
- CALLED "DIFFERENCES-IN-DIFFERENCES"
- HAS ADVANTAGES: FOR EX., EVEN IF PARTICIPANTS AND NON-PARTICIPANTS DIFFER IN THEIR LEVELS OF Y, THEY MAY NOT DIFFER IN GROWTH RATES OF Y
- DISADVANTAGE: WE CANNOT BE SURE THAT THE GROWTH RATE OF Y OF PARTICIPANTS WOULD HAVE BEEN THE SAME AS THAT OF NON-PARTICIPANTS, HAD THE PARTICIPANTS NOT PARTICIPATED. AFTER ALL, WHY DID THE PARTICIPANTS CHOOSE TO PARTICIPATE AND THE NON-PARTICIPANTS DID NOT? PROBABLY SOMETHING WAS DIFFERENT IN THEIR SITUATIONS, SOMETHING WE HAVE NOT MEASURED



igure 1: Alternative Trends for Participants and Nonparticipants

(4) Y_i: OF THOSE OFFERED NO PROGRAM OR A DIFFERENT PROGRAM

- RARELY HAVE A CASE WHERE A GROUP CAN BE FOUND THAT IS NOT OFFERED AN ON-GOING PROGRAM AT ALL, AND YET THEY ARE THE SAME (DEMOGRAPHICALLY) AS INDIVIDUALS & FAMILIES WHO ARE OFFERED IT
- EX: FSP. AVAILABLE IN ALL COUNTIES & BENEFIT LEVELS THE SAME
- EX: AFDC. AVAILABLE IN ALL COUNTIES BUT BENEFIT LEVELS AND ELG. CONDITIONS DIFFER
- EX: WIC. FEDS ESTABLISH PRIORITY RANKINGS BUT STATES HAVE DISCRETION IN WHICH TO SERVE, IN SETTING INCOME ELIGIBILITY CHARACTERISTICS, ETC. THOUGH PROGRAM AVAILABLE EVERYWHERE.
- IF GEOGRAPHIC VARIATION IS AVAILABLE, COMPARISONS (Y_i^*'s) WILL BE VALID ONLY IF THE GEOGRAPHIC VARIATION IS NOT A RESULT OF, OR OTHERWISE CORRELATED WITH, THE VALUE OF Y_i's IN THE AREA.

  EX ARE THE Y_i's OF (INELIGIBLE) FAMILIES IN AN UNGENEROUS STATE (E.G., ALABAMA) WHO WOULD HAVE BEEN ELIGIBLE IN A GENEROUS STATE (E.G., NEW YORK) THE SAME AS THE Y_i^* OF SIMILAR PARTICIPANT FAMILIES IN NEW YORK? WHAT IF NEW YORK WERE MORE GENEROUS BECAUSE ITS FAMILIES HAVE LOW Y_i^*'s?

GENERALIZATION OF THIS APPROACH

- MORE GENERALLY, WE NEED A VARIABLE — CALL IT "Z" — WHICH AFFECTS THE PROBABILITY OF PARTICIPATION IN THE PROGRAM BUT WHICH DOES NOT DIRECTLY AFFECT, OR IS NOT DIRECTLY CORRELATED WITH, Y.
- SUCH A Z IS CALLED AN "INSTRUMENTAL VARIABLE" OR "INSTRUMENT"
- EX: $Z_i = \begin{cases} 1 & \text{IF FAMILY } i \text{ IS ELIGIBLE IN ITS STATE OF RESIDENCE} \\ 0 & \text{" " " " " INELIGIBLE " " " OF RESIDENCE} \end{cases}$

IF (A) WE HAVE A GROUP OF FAMILIES IDENTICAL IN OBSERVED WAYS (INCOME, ETC)

(B) Z_i VARIES ACROSS THEM, AND

(C) ELIGIBILITY RULES ARE NOT DIRECTLY CORRELATED WITH Y_i.

THEN Z_i IS A "VALID" INSTRUMENTAL VARIABLE.

- TECHNICAL DEFINITION OF EFFECT OF PROGRAM:

$$\hat{\alpha}^P = \frac{\text{AVERAGE } Y_i \text{ OF THOSE WITH } Z_i = 1 \;-\; \text{AVERAGE } Y_i \text{ OF THOSE WITH } Z_i = 0}{\text{PARTICIPATION RATE OF THOSE WITH } Z_i = 1}$$

(NOTE THAT NOT ALL OF THOSE WITH $Z_i = 1$ WILL PARTICIPATE)

OTHER Z'S

- DISTANCE TO NEAREST PROGRAM OFFICE OR FACILITY ("ACCESS")

  PROBLEM: CERTAINLY WILL AFFECT PROBABILITY OF PARTICIPATION, BUT PROGRAM LOCATIONS ARE PROBABLY PICKED TO SERVICE THE NEEDIEST

- LEVEL OF BENEFITS

MORE GENERAL ESTIMATE:

$$\hat{\alpha}^P = \frac{\text{AVERAGE } Y_i \text{ OF THOSE WITH ONE LEVEL OF } Z_i \;-\; \text{AVERAGE } Y_i \text{ OF THOSE WITH DIFFERENT LEVEL OF } Z_i}{\text{PARTICIPATION RATE OF THOSE WITH FIRST LEVEL OF } Z_i \;-\; \text{PARTICIPATION RATE OF THOSE WITH SECOND LEVEL OF } Z_i}$$

THE "IDENTIFICATION PROBLEM" IS THE PROBLEM OF FINDING A LEGITIMATE Z_i.

THE "SEARCH FOR Z'S"

(5) COMBINING APPROACHES (3) + (4) ("NATURAL EXPERIMENTS")

- COMBINING CHANGE FROM $t-1$ TO t WITH A CHANGE IN Z: SUPPOSE ONE AREA CHANGES ITS Z (ELIGIBILITY CONDITIONS, BENEFIT LEVELS, ETC) AND ANOTHER AREA DOES NOT.
- THEN ONE CAN EXAMINE ΔY OF SIMILAR FAMILIES FROM $t-1$ TO t IN THE TWO AREAS TO ESTIMATE $\hat{\alpha}^P$

II. MATHEMATICAL STATEMENT

TWO EQUATIONS.

OUTCOME EQN:

$$Y_i = X_i\beta + \alpha P_i + \varepsilon_i \qquad (1)$$

PARTICIPATION EQN:

$$P_i^* = X_i\delta + Z_i\gamma + \nu_i \qquad (2)$$

WHERE

X_i = INDIVIDUAL, FAMILY, & AREA CHARACTERISTICS (AGE, RACE, EDUCATION, INCOME, UNEMP. RATE, ETC)

$P_i = \begin{cases} 1 & \text{IF PARTICIPATE} \\ 0 & \text{IF NOT} \end{cases}$

P_i^* = "PROPENSITY" (TENDENCY) TO PARTICIPATE

ε_i, ν_i = UNOBSERVED & UNMEASURED INFLUENCES

- THE PROBLEM IS THAT ε_i WILL ALWAYS EXIST, AND IT MAY BE CORRELATED WITH P_i
- FOR Z_i TO BE "VALID", IT MUST AFFECT P_i^* AND IT MUST NOT BE CORRELATED WITH ε_i (WHICH IT WILL BE IF Z_i IS A RESULT OF Y_i)
- WITH BENEFIT LEVEL (B_i):

$$Y_i = X_i\beta + \alpha(B_iP_i) + \varepsilon_i \qquad B_iP_i = 0 \text{ FOR NON-PARTICIPANT}$$

$$P_i^* = X_i\delta + \phi B_i + Z_i\gamma + \nu_i \qquad B_i > 0 \text{ FOR NON-PART.}$$

(B_i = POTENTIAL BENEFIT)

III. ESTIMATION METHODS

- ASSUMING APPROACH (3), (4), OR (5) IS TAKEN, THERE IS THE QUESTION OF HOW TO ESTIMATE $\bar{\alpha}^P$
- THERE ARE 3 COMMONLY-USED METHODS:

(1) "INSTRUMENTAL VARIABLES" ESTIMATION (A 2-STEP PROCEDURE)

STEP 1: ESTIMATE EQN (2) WITH PROBIT OR LOGIT

STEP 2: CONSTRUCT A PREDICTED PART. PROB $\hat{P}_i$ FROM THE RESULTS AND ESTIMATE EQN (1) WITH OLS, REPLACING P_i WITH $\hat{P}_i$

NOTE: THE FORMULAS FOR $\bar{\alpha}^P$ GIVEN ON PAGES 8 + 9 ARE OF THIS TYPE, BUT WITH NO X'S AND ONLY 1 Z; WITH TWO VALUES

(2) "HECKMAN LAMBDA" TECHNIQUE (A 2-STEP PROCEDURE)

STEP 1: ESTIMATE EQN (2) WITH PROBIT OR LOGIT

STEP 2: CONSTRUCT A VARIABLE CALLED "LAMBDA" FROM THE RESULTS AND ESTIMATE EQN (1) WITH IT INCLUDED

(3) FULL-INFORMATION MAX. LIKELIHOOD (FIML) (A 1-STEP PROCEDURE)

ESTIMATE EQNS (1) + (2) SIMULTANEOUSLY BY SEARCHING FOR VALUES OF $\beta, \alpha, \delta, \gamma$ THAT BEST "FIT" THE DATA.

WRINKLES : WHAT IF Y_i IS ITSELF A 0-1 VARIABLE?
OR A VARIABLE WITH A LARGE # OF 0'S?

- THEN MODIFICATIONS IN THE TECHNIQUES MAY BE REQUIRED

BOTTOM LINE

- ALL THREE TECHNIQUES ARE "SUPPOSED" TO GIVE APPROX. SAME RESULTS, SO IT SHOULD NOT MATTER WHICH IS USED
- HOWEVER, ONE MAY BE "BETTER" THAN THE OTHERS (I.E., GIVE AN ESTIMATE CLOSER TO THE TRUTH) IN CIRCUMSTANCES DIFFICULT TO DEFINE BEFOREHAND. OFTEN IT IS BEST TO USE ALL 3.
- THEY DIFFER IN TERMS OF COMPUTATION. (1), (2), & (3) ARE INCREASINGLY MORE BURDENSOME TO COMPUTE.

NOTE

- IN MANY WRITE-UPS IT IS DIFFICULT TO SEPARATE THE MOST IMPORTANT ISSUES FROM THE LESS IMPORTANT ONES.
- THE IDENTICATION PROBLEM, THE SEARCH FOR Z'_i, THE METHOD OF [redacted] IDENTIFYING Y_i^* ARE ALL KEY ISSUES
- THE CHOICE OF ESTIMATION METHOD ((1)-(3), FOR EX) IS LESS IMPORTANT

IV. EXAMPLES OF STUDIES

(A) DEVANEY, HAINES, & MOFFITT*

Y_i = NUTRIENT AVAILABILITY

P_i = FSP

DATA: 1979-80 SFC-LI

X'S: MANY, INCLUDING INCOME VARIABLES

Z'S: (1) EDUCATION OF HEAD

(2) EMPLOYMENT STATUS OF HEAD

(3) WHETHER HOUSEHOLD OWNS HOME

EST. METHOD: FIML

RESULTS FOR MOST NUTRIENTS, ADJUSTING FOR SELECTION BIAS INCREASED THE (+) EFFECT OF FSP ON NUTRIENT AVAILABILITY (I.E., THERE WAS NEGATIVE SELECTION). HOWEVER, THE MAGNITUDES WERE SMALL.

ISSUE CHOICE OF Z'S QUESTIONABLE. COULD EASILY BE DIRECT DETERMINANTS OF FOOD USE & CONSUMPTION MIX.

---

* "ASSESSING THE DIETARY EFFECTS OF THE FOOD STAMP PROGRAM" MPR FINAL REPORT TO FNS, 1989

(B) SCHWARTZ, GUILKEY, AKIN, POPKIN *

Y_i : BREASTFEEDING BEHAVIOR

P_i : WIC

DATA : 1988 NMIHS

X's : MANY

Z : WIC PROGRAM EXPENDITURES PER CAPITA IN THE STATE OF RESIDENCE (AND WIC ADMIN. EXPENDS)

EST. METHOD : FIML

RESULTS : WIC HAD LITTLE EFFECT ON INCIDENCE OF BREASTFEEDING (IF NOT A NEGATIVE EFFECT)

- EFFECT OF SELECTION BIAS ADJUSTMENT NOT SHOWN, AT LEAST IN DRAFT FINAL REPORT

ISSUES THE Z_i USED IS OPEN TO QUESTION.

(1) PER CAPITA: TOO BROAD? WANT ELIGIBLE POP.

(2) TOTAL EXPENDITURES DO NOT DIRECTLY PROXY THE ELIGIBILITY CRITERIA AND OTHER PROGRAM CHARACTERISTICS THAT AFFECT PARTICIPATION. TOTAL EXPENDITURES ARE INSTEAD LIKELY TO BE AFFECTED BY THE INCOME, FAMILY STRUCTURE, ETC. OF THE ELIGIBLE POPULATION (AND ITS SIZE).

---

* "THE EFFECT OF THE WIC PROGRAM ON THE INITIATION AND DURATION OF BREASTFEEDING" RTI REPORT TO FNS, 1992

(C) DEVANEY, BILLIMEIER, SCHORR*

Y_i : MEDICAID COST SAVINGS IN FIRST 60 DAYS AFTER BIRTH (AMONG OTHER Y'S)

P_i : PRENATAL WIC

DATA: MEDICAID ADMINISTRATIVE DATA, VITAL STATISTICS, WIC PROGRAM DATA, 1987-88

X'S : MANY

Z'S : COUNTY-LEVEL VARIABLES: PCT OF ELIGIBLES SERVED, DENSITY OF WIC CLINICS, PHYSICIAN DENSITY, PER CAPITA INCOME, POP DENSITY

RESULTS UNADJUSTED RESULTS SHOWED GENERAL MEDICAID SAVINGS FROM WIC PARTICIPATION. HOWEVER, ADJUSTED RESULTS HIGHLY SENSITIVE AND IMPLAUSIBLE, RANGING FROM NEGATIVE TO POSITIVE IN LARGE SWINGS.

ISSUE Z'S DID NOT "EXPLAIN" MUCH OF THE VARIATION IN PARTICIPATION RATES. MANY OF THEM WERE PROBABLY DIRECTLY CORRELATED WITH POP. CHARACTERISTICS (AND THEREFORE WITH HEALTH OF CHILDREN, THE Y). VARIABLES MORE DIRECTLY AFFECTING ELIGIBILITY MIGHT HAVE WORKED BETTER.

---

* "THE SAVINGS IN MEDICAID COSTS FOR NEWBORNS AND THEIR MOTHERS FROM PRENATAL PARTICIPATION IN THE WIC PROGRAM" REPORT FROM MPR TO FNS, OCTOBER 1990

V. MULTIPLE PROGRAMS

- MANY, IF NOT MOST, FOOD-PROGRAM RECIPIENTS PARTICIPATE IN OTHER PROGRAMS AS WELL
- NON-FOOD: AFDC, PUBLIC HOUSING
- FOOD: FSP, SBP, NSLP, WIC
- WE ARE GENERALLY INTERESTED IN THE EFFECT OF PROGRAM "A" BOTH IN THE PRESENCE & ABSENCE OF PROGRAM "B", AND IN THE COMBINED EFFECTS OF PROGRAMS "A" AND "B"
- SELF-SELECTION THEN OCCURS INTO PROGRAM COMBINATIONS: THE TYPES OF FAMILIES IN FSP AND AFDC ARE DIFFERENT FROM THOSE IN FSP ALONE. AND THE Y_i^*'s OF THE GROUPS MAY DIFFER
- NEED A MULTIPLE PROGRAM MODEL
- IDEALLY, NEED Z's FOR ALL PROGRAMS
- A SEVERE PROBLEM ARISES WITH THE ISSUE OF ESTIMATION: ALL 3 TECHNIQUES ARE VERY DIFFICULT
- ESTIMATION PROBLEM CAN BE SOLVED WITH NEW "SIMULATION" METHODS *
- SEE ILLUSTRATIONS ON ATTACHED TABLES

---

* KEANE & MOFFITT, "THE ESTIMATION OF FOOD STAMP SELF-SELECTION MODELS USING THE METHOD OF SIMULATION" REPORT FROM MPR TO FNS, 1992

II. APPLYING THE MSM TO A MODEL OF MULTIPLE PROGRAM SELF-SELECTION

We have applied the new MSM method to a prototype model drawn from past work on self-selection into the FSP and other programs. Our example has three possible programs, although the software we are providing permits up to four. The mathematical representation of the model is as follows:

$$(1) \quad Y_i = X_i\beta + \alpha_1 B_{1i} + \alpha_2 B_{2i} + \alpha_3 B_{3i} + \epsilon_i$$

$$(2) \quad P^*_{1i} = Z_{1i}\gamma_1 + \nu_{1i}$$

$$(3) \quad P^*_{2i} = Z_{2i}\gamma_2 + \nu_{2i}$$

$$(4) \quad P^*_{3i} = Z_{3i}\gamma_3 + \nu_{3i}$$

The variables in these equations have the following meanings:

Y_i = outcome variable of interest (food expenditures, dietary intake, etc.) for individual i

X_i = variables determining Y, excluding program benefits themselves

B_{1i} = benefit received from program 1 (=0 for nonrecipients)

B_{2i} = benefit received from program 2 (=0 for nonrecipients)

B_{3i} = benefit received from program 3 (=0 for nonrecipients)

P^*_{1i} = variable representing the "propensity" to be a recipient of program 1

P^*_{2i} = variable representing the "propensity" to be a recipient of program 2

P^*_{3i} = variable representing the "propensity" to be a recipient of program 3

Z_{1i} = variables affecting the propensity to be a recipient of program 1 (including the program benefit)

Z_{2i} = variables affecting the propensity to be a recipient of program 2 (including the program benefit)

Z_{3i} = variables affecting the propensity to be a recipient of program 3 (including the program benefit)

TABLE III.2

RESULTS OF THE ESTIMATION: THREE PARTICIPATION EQUATIONS AND RENT EQUATION

	AFDC Part. Eqn.	FSP Part. Eqn.	Housing Part. Eqn.	Rent Eqn.	OLS Rent Eqn.
Program Benefit[a]	.081 *	.054 *	-.038 *	–	–
	(.011)	(.020)	(.017)		
Hourly Wage Rate	-.308 *	-.270 *	-.208 *	15.274 *	5.899 *
	(.067)	(.065)	(.074)	(2.323)	(.1950)
Nonlabor income[b]	-.042 *	-.060 *	-.056 *	1.321 *	.261
	(.013)	(.010)	(.011)	(.231)	(.195)
Education	.046	.027	-.035	-4.310 *	-.690
	(.033)	(.032)	(.037)	(1.162)	(.976)
Age	-.012 *	-.009	-.001	-.777 *	-.401 *
	(.007)	(.007)	(.008)	(.248)	(.209)
South Dummy	-.108	-.364 *	-.023	-5.802 *	-5.837 *
	(.098)	(.082)	(.096)	(3.092)	(2.678)
No. Children Younger Than 18	.207 *	.195 *	-.142 *	–	–
	(.044)	(.053)	(.052)		
White Dummy	-.312 *	-.353 *	-.537 *	10.572 *	7.461 *
	(.081)	(.079)	(.092)	(3.016)	(2.572)
SMSA Dummy	–	–	–	6.301 *	10.015 *
				(2.007)	(2.491)
Fair Market Rent in Area[c]	–	–	–	.249	2.656 *
				(.488)	(.531)
AFDC Benefit	–	–	–	2.195 *	-.043
				(.293)	(.377)
FSP Benefit	–	–	–	1.774 *	-2.417 *
				(.442)	(.564)
Housing Benefit	–	–	–	2.725 *	-1.172
				(.215)	(.271)
Constant	.272	.953 *	.458	45.510 *	16.346
	(.375)	(.351)	(.476)	(14.086)	(12.599)
Correlation Coefficients:					
Between AFDC and FSP				.962 *	
				(.010)	
Between AFDC and housing				.450 *	
				(.045)	
Between FSP and housing				.500 *	
				(.044)	

TABLE III.2 (continued)

	AFDC Part. Eqn.	FSP Part. Eqn.	Housing Part. Eqn.	Rent Eqn.	OLS Rent Eqn.
Between AFDC and rent				-.706 * (.026)	
Between FSP and rent				-.653 * (.027)	
Between housing and rent				-.771 * (.026)	
Standard deviation of error term in rent equation				42.341 * (.942)	

NOTE: Standard errors in parentheses.

[a]Weekly. Measured at zero hours of work. Coefficient is multiplied by 10.

[b]Weekly. Coefficient is multiplied by 10.

[c]Coefficient is multiplied by 10.

OLS = ordinary least squares.

*Statistically significant at the 90 percent level.