3983

MATHEMATICA
Policy Research, Inc.

3983

Contract No.: 53-3198-O-22
MPR Reference No.: 7925-013

EVALUATING THE FOOD STAMP PARTICIPATION ALGORITHMS USED IN MICROSIMULATION MODELS

June 10, 1991

Author:

Alberto Martini

Prepared for:

U.S. Department of Agriculture
Food and Nutrition Service
3101 Park Center Drive
2nd Floor
Alexandria, VA 22302

Project Officer:
Alana Landey

Prepared by:

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

Project Director:
Pat Doyle

This work was prepared as one task of a competitively awarded contract; the total amount of the contract is $1,812,081.

ACKNOWLEDGMENTS

The author would like to thank Thomas Fraker and Pat Doyle of **Mathematica** Policy Research for technical consultation and review, Nancy Heiser for writing the executive summary, and Thomas Good for editing the final document.

CONTENTS

CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Microsimulation models have been used widely since the 1960s to estimate the projected budgetary cost and distributional impacts of proposed changes to social programs. Microsimulation models are computer programs that consist of: (1) a microdata set that contains information on a sample of individuals and households; (2) a set of accounting rules that reflect the tax and transfer program regulations in effect at a given point in time; and (3) behavioral responses that reflect how individuals and households modify their behavior in response to changes in program parameters.

To evaluate the impact of proposed changes to the Food Stamp Program (FSP), the Food and Nutrition Service (FNS) currently uses three microsimulation models:

- The Micro Analysis of Transfers to Households (MATH) Model
- The Food Stamp Eligibility Routines (FOSTERS) Model
- The QC Minimodel

This report evaluates the computer routines in the MATH and FOSTERS models that determine which eligible households will be simulated as participating in the FSP--*the participation algorithms.* (The QC Minimodel does not contain a participation algorithm.)

Microsimulation Models and Participation Algorithms

The MATH model is used primarily to estimate the impact of proposed FSP changes that will affect eligibility. Its microdata set is the March Supplement of the Current Population Survey (CPS). These data are "aged" to represent the characteristics of the sample of individuals and households in a given future month; in the current MATH model, the given month is April 1991.

***The** FOSTERS model is used* primarily to simulate changes to the asset test and the food stamp unit definition. Its microdata set is the Survey of Income and Program Participation (SIPP). The simulation month in FOSTERS is not a future month, but rather the calendar month for which the data were collected

Both the MATH model and the FOSTERS model use a participation algorithm in two places. One participation algorithm is used to estimate participation under pre-reform program rules (the *base-law participation algorithm),* and the other participation algorithm is used to estimate participation after a program reform (the ***reform participation algorithm*****).** A comparison of the size and characteristics of the FSP caseload and program costs under the base law with those under the reform yields a measure of the relative impact of the reform.

The Base-Law Participation Algorithm

The base-law participation algorithms in the MATH and FOSTERS models select households as participating in the FSP among those households simulated to be eligible for the FSP. The participation algorithms assign each eligible household a probability of participation which, if greater than a randomly drawn number between zero and one, means that the household is simulated as participating in the FSP; otherwise, the household is simulated as not participating. The probability of participation for each household is computed whereby the total number and the characteristics of selected participants replicate as closely as possible those found in data drawn from program administration sources.

The most challenging aspect of designing and implementing a base-law participation algorithm is adjusting for the inconsistencies between measures of eligibles obtained from survey data and measures of participants obtained from administrative data. By definition, the number of households participating in a program is smaller than or equal to the number of households eligible for the program; yet results with the MATH model have indicated that, when participants and eligibles are cross-classified along a number of dimensions, the number of participating households (estimated with administrative data) can ***exceed*** the number of FSP-eligible households (estimated with CPS data) in some subgroups. Thus, because the participation rate in these subgroups exceeds 100 percent, the probability of participation assigned to the households in these subgroups exceeds one.

The adjustment procedure for this inconsistency must reduce the participation rate in these subgroups to less than 100 percent, while attempting to preserve both the total number of participants and their composition along four key dimensions (income, household size, receipt of public assistance, and elderly status). The procedures currently used to correct for this inconsistency in the MATH model include both a formal procedure (a computer algorithm) and an informal procedure (manual calibration). Although the manual calibration procedure is flexible and inexpensive, its outcome cannot be reproduced by other researchers, it is subject to human error, and it may not be applied consistently from year to year. ***To replace this informal manual calibration process with a more reliable one, we recommend exploring either "raking' algorithms or an ad hoc algorithm.*** Raking (or iterative proportional fitting) algorithms adjust the entries in a matrix of cross-classified data to conform with known marginal distributions. An ***ad hoc*** algorithm would replicate as closely as possible the procedures currently performed with manual calibration.

Another methodological concern (applicable only to the MATH model) pertains to the process used to further align participants along two additional dimensions with administrative data (the ratio of the household's benefit amount and its poverty threshold, and whether or not the household reported FSP participation for the CPS reference year). The formula used to improve the distribution of participants by benefit is based on regression coefficients estimated by Czajka (1981) in the context of a multivariate analysis of FSP participation. Several methodological problems are associated with using Czajka's participation equation: (1) the participation equation was not specified for a microsimulation context; (2) it was estimated with data from the 1979 Income Survey Development Program, which preceded the Elimination of the Purchase Requirement (EPR); and (3) it produces an estimate of the relationship between benefits and participation which contradicts common sense. ***Therefore, we recommend using a simpler method that allows the user to select participants in a subgroup on the basis of reported participation andlor a set of household characteristics that are correlated with participation.***

The Reform Participation Algorithm

The reform partiiipation algorithms in the MATH and FOSTERS models estimate the number of households that would change their current decision to participate in the FSP in response to reforms that affect FSP benefits, eligibility, or both. The models distinguish between (1) households that are already eligible under base law, and **(2)** those that would become eligible or ineligible under the reform. For the former, these models simulate the decision to stay, join, or leave the FSP in response to the reform; for the latter, these models simulate the decision to join the **FSP.** Currently, for the first case, the fraction of participants selected randomly to join or leave the program is equal to **{.0014** * (the change in benefits)). For the second case, the reform participation algorithm selects households according to base-law probabilities.

Several methodological concerns about the reform participation algorithm exist. First, the participation response is the same for all households, and does not vary by their benefit amount. For example, given the same absolute increase in benefits, the participation response is the same for households eligible for $10 as it is for households eligible for $200 worth of benefits. Second, no inflation adjustment is built into the algorithm. Third, **the** selection factor **(.0014)** is derived from aggregate time series data on program participation before the EPR

We recommend improving these three aspect3 of the reform participation algorithm by implementing a participation algorithm based on estimates of behavioral response derived from 1985 SIPP data. **This** reform participation algorithm would have the following features:

- The behavioral response parameter would be based on a logarithmic specification of the participation equation, so that the participation response depends on the percentage change in benefits, rather than on its absolute amount.
- The percentage change in benefits would be subject to an automatic inflation adjustment.
- The 1985 data would reflect the reality of the post-EPR program.

I. INTRODUCTION

Microsimulation models are large and complex computer programs used to analyze the effects of changes in government programs. These models simulate the size and characteristics of the population that would be eligible for a new program or for a program change, those among the eligibles who would be likely to participate, and those who would be the gainers or losers if the change were implemented. These models have been used widely as a policy analysis tool since the 1960s to forecast the budget cost and distributional impacts of various legislative proposals for social programs. Microsimulation techniques generate estimates to respond to such questions as the following:

- What would be the cost of a particular welfare reform plan?
- How many households would be made worse off and how many better off if the Food Stamp Program (FSP) were modified to replace the current shelter deduction by an increased standard deduction that varied by region?
- What impact would a tax reform proposal that replaced the current personal exemption with a refundable tax credit have on revenue and the disposable income distribution?

Microsimulation models contain two essential components: (1) a *microdata* set that contains economic and demographic information on a representative sample of individuals and families; and (2) a set of ***accounting rules***, that is, algebraic representations of the tax and transfer program regulations in effect at a given point in time. In addition, ***behavioral responses*** can be incorporated into a microsimulation model, as long as how individuals modify their behavior in response to a given program change is known, or-can be estimated or' conjectured.

Three microsimulation models are currently used by the Food and Nutrition Service (FNS) to evaluate the budgetary and social impact of proposed changes to the Food Stamp Program (FSP).

They are the Micro Analysis of Transfers to Households (MATH) model, the Food Stamp Eligibility Routines (FOSTERS) model, and the QC Minimodel. The specific goal of this report is to evaluate an important component of the MATH and FOSTERS models, the ***FSP participation*** algorithm--that is, a computer routine that determines which eligible households are participating in the FSP. Participation algorithms are used in the simulation model in two **places:** first when participation is estimated under pre-reform program rules (the ***base-law*** participation algorithm), and again when the program reform is simulated (the ***reform*** participation algorithm). The QC Minimodel is not considered here, although it represents perhaps the most frequently used simulation model for the FSP, because it does not contain a participation algorithm.[1]

The report is organized as follows. Section LA provides an overview of the general characteristics of the two simulation models, while Sections **I.B** and C discuss the objectives of the base-law and reform participation algorithms, respectively. A more detailed description of how the two types of algorithms are operationalized in each model, together with proposals for their improvement, are presented in Chapter II (for base-law) and Chapter III (for reform).

A. OVERVIEW OF THE MATH AND FOSTERS MODELS

The MATH model is the direct descendant of a long tradition of microsimulation models that started with the RIM model, used during the 1960s to evaluate alternative income maintenance programs. The MATH model is currently used primarily to evaluate the impact of proposed legislative changes that affect ***eligibility*** for the Food Stamp Program.

The MATH model uses the March Supplement of the Current Population Survey (CPS) as its microdata set. The data **from** the CPS are "aged" in order to represent the demographic, economic,

[1]The QC Minimodel is based entirely on administrative data **on food** stamp participants, rather than on survey data, as are MATH and FOSTERS. Therefore, it cannot be used to simulate the population eligible for the FSP, and consequently does not require an algorithm to select participants among the eligibles.

and labor-market situation at some future date, based on aggregate projections from several sources? For example, in the current MATH model, the March 1988 CPS, which collected data for calendar year 1987, has been aged to represent the economic and demographic situation projected for 1991. A number of imputations **and** simulations are also performed on the ***annual*** data in the CPS in order to obtain a representation of a cross-section of households in a selected future month, which is called the "simulation month" (April 1991 in the current MATH model). This process determines which households are ***eligible*** for the **FSP** during the simulation month, given their (projected) demographic and economic characteristics, and the program regulations that are expected to be in effect during the simulation month. In addition, the model determines which eligible households are ***participating in*** the program. This set of simulated outcomes is referred to as the "base law," or "base plan," and it represents the benchmark to which program reforms are compared. The comparison between FSP caseloads and expenditures under base law and those under reform yields a measure of the ***relative*** impact of the reform. This measure can be used to provide an estimate of the ***cost*** of the reform in a future budget year, under the assumption that the reform is fully implemented in that year.

FSP reforms are simulated in the MATH model first by modifying the appropriate program parameters of the model, and then by simulating eligibility and participation once again under the new program rules. At this stage, the model can also simulate ***changes in the behavior*** of households in response to the program change--for example, changes in the decision to work or to participate in welfare programs. The only behavioral response explicitly **modelled** in the version of the MATH model currently in use is the participation response to a change in food stamp benefits.

The FOSTERS model is a newcomer in the world of microsimulation of welfare programs. It uses data from the Survey of Income and Program Participation (SIPP), rather than from the March

---

'Doyle and Trippe (1989) provide a description and an evaluation of the aging process.

CPS. Unlike the MATH model, FOSTERS has been designed explicitly to simulate eligibility and participation in the Food Stamp Program, and it is used primarily to simulate changes to the asset test, changes to the food stamp unit definition, and other reforms involving data unique to SIPP.

SIPP provides ***monthly*** data on income, labor force participation, program participation, and household composition. SIPP has two main advantages over the CPS: (1) it eliminates the necessity of simulating monthly data with March CPS annual data, as is done in the MATH model; and (2) it provides information on household composition collected at the same time as the income data. By contrast, household composition in the CPS is observed in March of the interview year, while the income data pertain to the previous calendar year. The main disadvantage of SIPP is its small sample size, which makes it very difficult to use the FOSTERS model to simulate program changes that affect small segments of the low-income population.

Another important difference between MATH and FOSTERS is that the simulation month in FOSTERS is ***not*** a future month, but rather the calendar month for which the data were collected.[3] This represents an advantage--it eliminates the necessity of aging the survey data--but at the same time a disadvantage--it produces estimates of reform impacts that do not take into account changes in economic conditions between the time the data were collected and the time the reform would be implemented.

[3]The choice of a particular month of data from the many available in SIPP is dictated by three considerations: the timing of the release of SIPP data, the necessity of obtaining a large sample size, and the proximity of the chosen month to the interviews in which information for simulating FSP eligibility was collected.

B. OVERVIEW OF THE BASE-LAW PARTICIPATION ALGORITHMS

A base-law participation algorithm is a subroutine of the larger microsimulation model that determines which households are participating in a given program among those simulated to be eligible for that program on the basis of current legislation--that is, among ***base-law eligibles.***

In MATH and FOSTERS the selection of participants is performed ***stochastically.*** Each eligible household is assigned a probability of participation on the basis of a set of criteria (described later in this section) and this probability is compared with a randomly-drawn number between zero and one. The household is selected'to participate if its probability is greater than the random number.

In the existing FSP participation algorithms, the probabilities of participation are computed in such a way that the ***total number*** and the ***characteristics*** of the selected participants replicate us ***closely*** as ***possible*** those found in administrative data (actual or projected, depending on the model). The importance of a close replication of the size and composition of the program caseload should be emphasized. The ability of the simulation model to yield a credible estimate of the impact of a program reform depends in the first place on the realism of the model's representation of the pre-reform program caseload.

Several aspects of the process of replication of base-law participation are common to both the MATH and FOSTERS models, and give rise to the same methodological concerns: (1) the source of information on the program caseload; (2) the characteristics of the caseload that should be replicated; (3) the use of self-reported information on participation by CPS or SIPP respondents; and (4) the criteria used to assess the "closeness" between simulated and actual program caseloads. Because these issues are common to both models, they are discussed briefly here.

1. Source of Data on the FSP Caseload

The best available source of information on the size and characteristics of the FSP caseload is the Integrated Quality Control System (IQCS). The IQCS is a system of ongoing case record review designed to measure payment error rates in the Food Stamp, Aid to Families with Dependent Children (AFDC), and Medicaid programs. It is based on monthly probability samples drawn from all 50 States and the District of Columbia. A sample of active food stamp cases in two-month samples (usually July/August), weighted and edited for consistency, form the basis of estimates of the distributional characteristics of the food stamp population.

The use of the information contained in the IQCS files varies according to whether the caseload to be replicated is *actual* **or** projected. As discussed before, in the case of the **FOSTERS** model, the caseload to be replicated is the one observed in a particular month in the past. In the MATH model, the caseload to be simulated is not one observed in the past, but rather is projected for a future "simulation" month In this case, the IQCS provides information on the distributional characteristics of the caseload, which are assumed not to change between the observation and the simulation month, while the estimates of the total caseload and total benefit expenditures are obtained from the FSP Statistical Summary of Operations and projected forward to the simulation month with a variety of forecasting tools.

2. Which Dimensions of the Caseload Are Replicated?

The replication of the size and composition of the caseload can be pursued at three different levels:

1. replication of the overall size (actual or projected) of the caseload, in terms of households, individuals, or total benefits paid;

2. replication of the composition of the caseload, taking one characteristic at a time *(marginal* distribution)--for example, the composition in terms of household size; and

3. replication of the composition of the caseload, taking two or more characteristics simultaneously *(joint* distribution)--for example, the distribution of participants by income *and* household size.

The MATH and FOSTERS models pursue all three objectives at different stages of their base-law participation algorithms, giving different priorities to each objective. In general, the replication of the overall size of the caseload is given the highest priority. Replication of the joint distribution of caseload characteristics is attempted in both models--although for different characteristics--while the replication of marginal distributions is eventually used to judge how well a model has performed.

A related issue is the priority to be given to the replication of the various characteristics of the caseload. Since the model is used for policy simulations, it is conceivable that different characteristics play a different role according to how the simulation is used, and that their role changes according to the policy priority of the moment. The existing algorithms do not have enough flexibility to allow the user to set priorities explicitly among the various characteristics of the caseload.

3. The Use of Reported Participation

An issue that arises in the design of base-law participation algorithms is the use of program participation as reported by the respondents to the survey that forms the basis for simulating program eligibles. At first glance, one would think that participants could be selected among eligible households simply by using reported participation. However, two major complications preclude this simple solution:

- Participation in transfer programs is underreported in most, if not all, household surveys. If one were to use only reported participation, the overall size of the caseload would be underestimated Moreover, if underreporting is not random--in the sense that it is correlated with the characteristics of the caseload--not even the composition of the caseload can be replicated by using reported participation alone.

- Reported participation might not pertain to the same time period as that used in the simulation, either because the simulation period is in the future or because the length of the reference period used in-reporting participation differs from the length of the simulation period. For example, in the CPS participation is reported as "months on the program during a calendar year," while the MATH model simulates participation for a one-month period in the future.

The extent to which the two simulation models use reported participation in the base-law algorithm differs substantially. The **FOSTERS** model, due to its retrospective simulation and reliance on SIPP monthly data, uses this type of information extensively. By contrast, reported participation is used only marginally in the MATH model, due to the mismatch between reference periods for reporting and for simulation.

4. Evaluating the Closeness of the Simulated Caseload to Administrative Data

Because the overall objective of the base-law participation algorithm is to replicate as closely as possible the size and composition of a given FSP caseload, the last step in the algorithm entails evaluating how well the simulated caseload compares with the actual (or projected) caseload. Two types of comparisons are possible. First, the algorithm should be evaluated according to how well it performs in replicating the marginal (or joint) distribution of the characteristics used to align the two caseloads (*internal* validity). Although these characteristics are used as targets, this replication can be far from perfect, due to oddities in the underlying survey data, as will be explained in Chapter IL Second, the algorithm should be evaluated according to how well it replicates the characteristics of the caseload that were not used during the alignment process (*external* validity).

The closeness of the characteristics of simulated participants and administrative participants is currently evaluated in a rather informal way, by a simple comparison of the marginal distributions. This practice has clear advantages: it allows the analyst to freely incorporate information on current policy priorities, and it avoids the costs involved in exploring and implementing formal testing procedures. It also has two disadvantages: it is nonreproducible, and it can be rather arbitrary.

The use of formal goodness-of-fit tests should be explored. The methodology proposed by Birdsall and Andrews in Doyle and Trippe (1989) represents an interesting starting point for developing such goodness-of-fit measures. **However, we** believe that more conceptual work on these test procedures is needed before they can be incorporated into the simulation model.

C. OVERVIEW OF THE REFORM PARTICIPATION ALGORITHMS

The goal of what we define as reform participation algorithms is to incorporate into the microsimulation models a *participation response* to changes in program regulations. In other words, the main goal of the algorithm is to forecast how many households would change their decision to participate in a program in response to the changes introduced in that program.

Reforms to the Food Stamp Program may consist of different types of changes in the regulations, which might elicit different types of responses among the population currently or potentially eligible for the program. It is useful to distinguish among four broad types of FSP reforms: (1) reforms that affect both benefits and eligibility (such as changes in deductible expenses); (2) reforms that affect only the benefit amount among households already eligible (such as changes in the benefit reduction rate and in the maximum allotment); (3) reforms that affect eligibility without affecting the benefit amount (such as changes in the asset limit); and (4) other reforms that alter the characteristics of the program without affecting the benefit amount or the eligibility rules (such as changes in the type of issuance--for example, from coupons to checks or electronic benefit transfer--or changes in work registration requirements for FSP participants).

Three types of behavioral responses are of interest in simulating FSP reforms: (1) changes in the household's FSP participation decision; (2) changes in the decision to participate in other assistance programs; and (3) changes in labor market choices, such as the decision to work more hours or to leave a job.

In principle, one can incorporate into a microsimulation model each type of response for each type of program change, as long as one is able both to *model* how the program change alters the attractiveness of FSP participation and to *estimate* each type of response with the available data.

The microsimulation models that are evaluated in this report pertain only to the first three types of reforms, those that alter either benefits or eligibility or both. And they incorporate only the first type of behavioral response--the household's decision to participate in the **FSP** program. These models distinguish between households who are ***already eligible*** under base-law and those who ***become eligible*** under reform. For the former (base-law eligibles), these models simulate the decision to stay, join, or leave the program in response to the change in benefits. For the latter (reform eligibles), these models simulate the decision to join the program The distinction is important, because the algorithm necessary for simulating the participation decision differs in the two situations. Modelling the participation decision among already eligiile households requires an estimate of a behavioral parameter, one that indicates how households ***react*** to a change in benefits. By contrast, modelling the participation decision among newly **eligible** households does not differ conceptually from modelling the participation decision among households eligible under base-law.

It should also be pointed out that, when the reform consists of a change in benefits, these microsimulation models assume that the participation response does not depend on the source of the change, but only on its ***amount.*** For example, these models implicitly assume that the probability of participation by a given household will change to the same degree whether a $10 benefit increase is caused by an increase in the earnings deduction or by an increase in the cap on the shelter deduction. The discussion in this report will not challenge this assumption, in recognition of the fact that relaxing it would require a separate behavioral model for each source of benefit change. This does not seem feasible in light of the current state of knowledge about the determinants of FSP participation.

II. BASE-LAW PARTICIPATION ALGORITHMS

This chapter describes in more detail the base-law participation algorithms used in the MATH and **FOSTERS** models. It also discusses which aspects of these algorithms deserve closer scrutiny and possibly improvements in their design.

A. THE BASE-LAW PARTICIPATION ALGORITHM IN THE MATH MODEL

Before turning to the description of the FSP participation algorithm in the MATH model, it is useful to recall the salient features of this simulation model:

- It relies on the March CPS, which contains annual data.
- Monthly income is simulated from CPS annual income, using the household composition observed in March.
- FSP participation is collected in the March CPS as the number of months participating in the program during the previous calendar year.
- The MATH model is customarily used to simulate a future month.

1. A Stylized **Description of the Existing Algorithm**

We begin with a stylized description of the participation algorithm currently in use in the MATH model, and then discuss the specific aspects of the algorithm that make it more complex than the stylized description. The essential elements of the base-law participation algorithm in the MATH model can be described succinctly as follows:

- ***The eligible*** households in the MATH database are first cross-classified along four dimensions (income, household size, the receipt of public assistance, and the presence of elderly person), for a total of 96 cells (henceforth, ***primary*** cells).
- The count of ***participants*** in each of the % cells is obtained from administrative data: the ***total number*** of participants is obtained by projecting to the simulation month the number of participants obtained from the most recent Program

Operations Data, while the ***composition*** of the caseload by demographic and economic characteristics is obtained from the most recent IQCS data.

- The ratio of participants to eligibles in each cell is computed, yielding an estimate of **the** ***participation rate*** for that cell.

- Within each (primary) cell, a subcell participation rate is computed by using additional information--reported participation during the previous year and the benefit amount simulated for each household.

- Households are stochastically selected within each subcell to be participants or nonparticipants. A ***random number is*** generated and assigned to each household, and the household is selected if the participation rate for the household is greater than the random number.

The algorithm actually implemented in the MATH model is considerably more complex than the description just given. A number of reasons account for these additional complexities. We focus here on the three most important ones.

The first complication arises in the computation of the primary cell participation rates--that is, in the participation rates computed by dividing the number of participants by the number of eligibles in each cell. In some cells, the number of participants from the administrative data exceeds the number of eligibles simulated by the MATH model--which is equivalent to a participation rate in excess of 100 percent. A complex adjustment procedure--described in some detail in section A2 below--is necessary to correct this inconsistency.

Another complex aspect of the participation algorithm is the procedure adopted to augment the participation rate in the primary cells by incorporating additional information available at the household level--the household's ***reported*** FSP participation during the CPS reference year and the benefit amount to which the household is entitled. This modification increases the number of participation rates from % to 1,152. **This** step is discussed in more detail in section **A.3.**

The third major source of complexity pertains to the necessity of replicating the characteristics of the caseload--allowable deductions for shelter, dependent care, and medical expenses--for which no direct information is collected in the CPS. Such expenses must be ***imputed*** to each household, using the coefficients of expenditure equations estimated from other data sets. However, when applied to the set of simulated participants, these coefficients often yield distributions of deductions which differ significantly from those found in program data. The coefficients of the imputation equations must then be "calibrated" in order to produce distributions similar to those found in the administrative data. However, every calibration of the coefficients can produce a (slightly) different distribution of ***eligible*** households, which requires a new simulation of participation. The process could take several iterations, but is typically constrained only to two iterations (the initial simulation of participants, the calibration of coefficients, and the final simulation of participants.) The ***simultaneous*** nature of the selection of participants and the calibration of expense coefficients adds considerably to the complexity of the participation algorithm.

2. The Adjustment Procedure to Eliminate Excess Participants

The number of households participating in a program is, by definition, smaller or at most equal to the number of households eligible for the **program.**[4] However, when estimates of eligibles and participants based on two different data sets are compared, inconsistencies may occur (Doyle, 1990). The experience with the **MATH** model shows that the ***total*** number of participants estimated with administrative data is always substantially smaller than the total number of **FSP-eligible** households estimated with CPS data--that is, the overall participation rate is leas than 100 **percent.**[5] However,

[4]**The** only possible exception is the issuance of benefits in error, that is, to households that are no longer eligible. In this case, the **observed** participation rate **could** conceivably exceed 100 percent. In **1985, 3.67** percent of households with benefits were issued benefits in error (Doyle, **1990).** It is very unlikely that issuance in error contributes significantly to explaining the observed inconsistencies between survey and administrative data.

[5]**In** fact, it is typically between 50 and 60 percent, depending on the survey used to simulate eligibility and on the unit of observation.

experience also shows that when participants and eligibles are cross-classified along a number of dimensions the count of eligibles falls short of the count of participants in some cells.

Table 1 is reproduced from the Doyle and Trippe (1989) validation study of the MATH model. The entries in the table represent the difference between the MATH eligibles and the administrative participants. The negative numbers indicate which of the 96 cells have an excess of participants--or, perhaps more appropriately, a "shortage" of eligibles. Such a shortage is particularly severe among nonelderly public assistance households in the middle income ranges. Doyle and Trippe offer some conjectures on what might be causing this phenomenon. First, Census household surveys and administrative data sources differ in how they measure household composition and/or income, and in how they define the food stamp household The MATH model does not replicate the program rules governing the food stamp unit. Second, a possible source of the lack of low-income single adult households is the weighting process used by the Census Bureau to produce the CPS and SIPP public-use files. These procedures do not control for low-income groups, nor is the sample stratified so that it ensures that the low-income population is adequately represented.

Whatever the source of the discrepancy between eligibles and administrative participants, this problem represents perhaps the biggest challenge in designing and implementing a base-law participation algorithm. An adjustment procedure must be used to bring down all of the cell participation rates in excess of one, while at the same time preserving as much as possible both the total number of participants and their composition along the key dimensions.

There are two broad approaches to this adjustment procedure--a formal and an informal approach. The formal approach entails using a computer algorithm that iteratively redistributes participants to other cells until excess participants are eliminated. The informal approach entails redistributing excess participants to other cells "by hand," according to some rule of thumb.

TABLE 1

NUMBER OF MATH-ELIGIBLES LESS ADMINISTRATIVE PARTICIPANTS
BY THE DIMENSIONS OF THE PARTICIPATION ALGORITHM

	Household Size				
			3-5	6+	Total
Gross Income < $1					
Non-public assistance < 60	215,023	169,020	214,136	23,976	622,155
Non-public assistance 60+	18,425	12,608	4,654	338	36,024
Public assistance < 60	2,096	0	1,206	0	3,302
Public assistance 60+	0	1,540	0	0	1,540
Total					663,021
Gross Income $1-$199					
Non-public assistance < 60	120,807	67,8370	129,348	13,546	331,539
Non-public assistance 60+	58,993	34,234	1,198	038	94,423
Public assistance < 60	-145,129	61,025	46,796	1,146	-36,159
Public assistance 60+	-8%	-2,882	-2,347	-1,280	-7,406
Total					382,397
Gross Income $200-$499					
Non-public assistance < 60	252,994	188,104	131,748	8,109	50,954
Non-public assistance 60+	1,134,656	220,525	37,200	3,917	1,396,298
Public assistance c 60	-52,492	-149,401	-114,008	10,031	-305,878
Public assistance 60+	252,704	89,655	6,284	4,586	353,231
Total					2,024,614
Gross Income $500-$599					
Non-public assistance < 60	81,818	138,782	168,232	12,725	401,557
Non-public assistance 60+	188,345	149,581	9,716	-103	347,534
Public assistance < 60	-2,699	46,291	-63,012	-23,550	-48,170
Public assistance 60+	-1,605	41,873	23,822	935	65,025
Total					765,950
Gross Income $600-$749					
Non-public assistance < 60	175	262,754	281,706	30,229	574,865
Non-public assistance 60+	30,288	245,158	37,754	0	313,200
Public assistance < 60	-6,732	10,494	95,139	49,454	148955
Public assistance 60+	-9,530	19,480	12,963	4,741	27,654
Total					1,064,073
Gross Income $750+					
Non-public assistance < 60	0	1,887	1,021,327	266,803	1,290,017
Non-public assistance 60+	3,030	91,472	47,941	2,839	145,282
Public assistance < 60	-1,339	-8,571	91,750	-3,459	78,383
Public assistance 60+	0	-174	17,523	27,314	44,663
Total					1,558,345
Gross Income Total					
Non-public assistance < 60	670,816	828,386	1,946,497	355,388	3,801,087
Non-public assistance 60+	1,433,737	753,578	- 138,460	6,991	2,332,766
Public assistance < 60	-206,295	-40,358	57,871	28,622	-160,160
Public assistance 60+	240,674	149,494	58,244	36,295	484,706
Total	2,138,932	1,691,099	2,201,071	427,2%	6,458,399

SOURCE: Doyle and Trippe (1989)

The current practice with the MATH model uses both approaches. An ***iterative algorithm is first*** applied to the 96-cell matrices of eligibles and administrative controls; this algorithm redistributes the excess participants to all cells that do not already contain excess participants, proportionally to the ***number of eligibles in*** each cell. This simple strategy guarantees that the total number of simulated participants remains very close to the administrative **total.** However, as currently designed, this algorithm does ***not*** attempt to preserve the distribution of participants along the key dimensions. Nothing constrains the algorithm to minimize the difference between the marginal distributions found in the administrative data and those obtained after the adjustment. Not surprisingly, this algorithm performs rather poorly in terms of marginal distributions.

A subsequent manual intervention is required in order to obtain a better fit between the original administrative data on participants and the adjusted data. The disadvantages of this manual procedure are that its outcome is not reproducible by other researchers, it is subject to human error, and it might not be applied consistently from year to year. The advantages are its flexibility and its low cost--low when compared with the investment that might be required to design an algorithm that completely eliminates the need for manual intervention. As a matter of fact, the feasibility of such an automatic algorithm cannot be assessed completely without actually trying to design it.

We believe that the investigation of this issue should continue in two directions. First, the applicability of existing "raking" algorithms to this problem should be investigated. Raking (or ***iterative proportional fitting***) algorithms are designed to adjust the entries in a matrix of cross-classified data to conform with known marginal **distributions.**[6] The problem faced here (to redistribute the values in the cells so that the new marginal distributions deviate as little as possible from the original ones, under the constraint that each cell does not exceed a **prespecified** value) is not one that is typically solved by raking algorithms. However, it is **possible** that minor modifications of existing procedures

[6]Bishop et al. (1975) describes raking procedures.

would make them applicable to this problem. It should be emphasized that, despite their complexity, these procedures are inexpensive to execute and can be run on a microcomputer, since they do not operate on a raw data set, but rather on a matrix of cross-classified data.

As an alternative to using raking procedures, one could examine the feasibility of an ad ***hoc*** algorithm that redistributes participants according to a set of "rules of thumb," similar to those used by an informed and intelligent "calibrator. We believe that a relatively simple and inexpensive algorithm can be designed, although we cannot guarantee its effectiveness a priori. An intuitive description of such an algorithm is the following (as with the raking procedures, this algorithm would be applied to matrices, rather than to raw data): Excess participants are redistributed one "unit" at a time, where one unit can be defined as some multiple of 1,000 participants (weighted). Let us start at some arbitrary cell j, containing x units of excess participants: (These x units need not be the original excess units, some of which might already have been reallocated in previous iterations.) The algorithm "looks" at all the other cells with no excess participants (henceforth, ***feasible*** cells). For each feasible cell i, the algorithm computes a "score," based on the "distance" between cell i and cell j, and on the percentage difference between adjusted participants and administrative targets in cell i. Intuitively, "closer" cells should receive a higher score than "distant" cells. For example, if cell j is "PA-nonelderly-two person-income $1-299," "PA-nonelderly-two person-income $300-499" would be considered a close cell, while "non-PA-elderly-four person-income $1-299" would be a more distant cell. Also, cells in which a large percentage difference already exists between adjusted participants and administrative targets are assigned a lower score, since they have already "paid a price" during previous iterations. After a score has been assigned to all feasible cells, the excess unit from cell j is assigned to the cell with the highest score. Then the algorithm moves to another cell with excess participants, a new score is assigned to all other feasible cells, and one excess unit is reassigned. The process is repeated for all cells and for all excess units in those cells. The data presented in Table

1, taking a unit of 1,000 participants, would require repeating the process about 600 times, and computing about 55,000 scores. Although this might seem to be a formidable task, a fast microcomputer could perform it in a relatively short period of time, and at relatively low cost.

It should be emphasized that, due to the "atheoretical" nature of this algorithm, only an actual attempt at its design could reveal whether this algorithm is indeed able to completely eliminate the need for manual calibration.

3. *The* Procedure To Compute Subcell Participation Rates

The objective of the procedure described in section A.2 is to align participants with administrative data along four key dimensions--gross income, household size, the receipt of PA, and elderly status. All the remaining observable characteristics that might help predict participation among eligibles--for example, the level of benefits, the presence of earnings, the race and education of the household head, and the presence of children--are ignored in the first-stage alignment.

The MATH model uses a separate procedure that expands the 96-cell participation rates by adding two dimensions: the ratio between the household's benefit amount and its poverty threshold, and whether or not the household reported FSP participation for the CPS reference year. Each of the 96 primary cells is broken down into 12 subcells (six benefit-to-poverty categories by two reporting status categories). The participation rate in each subcell is obtained by computing its ***deviation*** from the overall cell participation rate, so that the distribution of participants across the 96 primary cells is not altered by this additional procedure. The result is a much expanded matrix of 1,152 participation rates. These rates are used for the final stochastic selection of participants.

It should be emphasized that the objective of this additional procedure is not to improve the fit of the distribution of participants along the four key dimensions, but rather to improve the distribution along additional dimensions, in particular the benefit amount received by participating households. Although income and household size are the major determinants of FSP benefits, it is possible that aligning participants by income and household size does not produce a satisfactory fit with the benefit distribution observed in administrative data.

The formula that computes the 12 subcell participation rates for each of the 96 primary cells is based on regression coefficients estimated by Czajka (1981) in the context of a multivariate analysis of FSP participation. In other words, rather than exploiting the information on the distribution of the benefit amount found in administrative data, this procedure uses estimates of the net effect of benefits on participation obtained with econometric methods.

Czajka specified the benefit-to-poverty ratio variable in six intervals (less than 5 percent, 5-9, 10-14, 15-19, 20-24, and more, than 25 percent). The j-th estimated coefficient represents the difference between the participation rate of eligibles in the j-th benefit-to-poverty ratio and that of eligibles in the lowest benefit-to-poverty category.

We see several methodological and practical problems with using Czajka's coefficients in this phase of the participation algorithm:

- The participation equation used by Czajka was not specified to be used in a microsimulation context. The breakdown of the benefit-to-poverty ratio in six levels is too detailed (compared, for example, with the four levels used to align participants by household size), causing the final matrix of participation rates to be unduly large. Moreover, the benefit amount is divided by poverty, while at the same time gross income used to align participants in the first phase of the algorithm is left unscaled.

- The participation equation was estimated with data from the 1979 Income Survey Development Program (ISDP)--that is, from a period of time preceding the Elimination of the Purchase Requirement. For this reason alone, these coefficients might no longer reflect the reality of the FSP.

- The major criticism with using Czajka's coefficients lies in the fact that they imply a very odd relationship between benefits and participation. According to these coefficients, everything else held constant, participation is lowest among households whose benefits are between 5 and 9 percent of poverty, and highest among those entitled to benefits between 10 and 14 percent of poverty. We believe that such a pattern contradicts both economic theory and common sense, and we are very uncomfortable with its use in simulation.'

If the present structure of this portion of the participation algorithm is to be maintained, at a minimum Czajka's coefficients should be replaced with equivalent ones estimated with more recent data *and* with the specific purpose **of** being incorporated into the algorithm.* However, we believe that this portion of the algorithm should undergo a more critical revision. We believe that relying directly on the coefficients of an equation estimated with survey data to replicate caseload characteristics observed in administrative data is generally a risky undertaking, no matter how "good" and "reasonable" the estimated coefficients might seem. If replicating the distribution of benefits observed in administrative data is deemed important, we believe that the benefit amount should be included in the list of key dimensions along which participants are aligned in the first phase of the algorithm (section A.2). The level of benefits could replace the income variable, given that aligning along household size, benefits, and income could be at least partially redundant.

Moreover, the procedure described above is extremely complex, and it is not clear whether this complexity adds to the ability of the model to replicate the administrative data. At the same time, this procedure is very difficult to understand and expensive to modify. Most of its complexity derives from the attempt to calculate subcell participation rates as deviations from the overall cell participation rate. As a possible alternative to this procedure, we propose a simpler method that would allow the user to select participants within each cell on the basis of reported participation

[7]A more extensive discussion appears in Chapter III of this report, and in Chapter V of Allin and Martini (1990).

*Marginal modifications of the participation equation estimated in Allin and Martini (1990) would serve the purpose.

and/or a set of household characteristics that are correlated with **participation.**[9] This method is based on the idea that, given that the ***number*** of participants to be selected in each cell has been determined by the first phase of the algorithm, the only remaining objective of the second phase is to exploit the available information on the households in each cell so that the eligibles who are more likely to be participants are also more likely to be selected as such.

It should be emphasized that the method proposed here does not in any way "undo" the distribution of participants along the four key dimensions produced by the first phase of the algorithm. In other words, the alignment reached in the first phase of the algorithm is not altered. The same principle inspires the procedure currently used in the MATH model.

The method proposed here allows the user to choose among using reported participation, predicted participation, or both.

a. Using Reported Participation

The rationale for using reported participation is obvious: despite all the underreporting and measurement error taking place in a survey, those who do report participation might be more "similar" to the participants observed in administrative data than those who do not report participation. While participation reported in the CPS cannot be used to align participants along the key dimensions in the first phase of the algorithm, it is likely that using reported participation to select participants ***within*** each of the primary cells would yield better results than a pure random selection.

The reporting information would be used in the following **fashion.**[10] Let pj be the participation rate for the j-th primary cell, determined after the adjustment described in section **A.2.**

[9]**This** set of characteristics would exclude the four characteristics used to align participants in the first phase of the algorithm.

[10]**A** similar algorithm is presently used in the MATH model to select public assistance participants **(PAPRAT** routine).

Let rj be the proportion of ***reporters* within** the j-th cell--that is, the ratio between the number of reporters and the number of eligibles. We have two cases:

(i) $r_j < pj$ **:** the proportion of reporters is less than the participation rate (this includes the cases where pj = 1). In this case ***all*** reporters within the cell would be selected as participants, while a fraction of nonreporters would have to be selected stochastically to "fill the gap," with probability equal to:

(1) $S_{nonreporters} = (p_j - r_j)/(1 - r_j)$.

It is easy to see that when pj = 1, the fraction selected would be equal to l--that is, all nonreporters would be selected (in addition to all reporters).

(ii) $r_j > p_j$**:** the proportion of reporters is greater than the participation rate (this **includés** the **cases** where pj = 0). In this case, only a fraction of reporters is stochastically selected, with probability equal to:

(2) $S_{reporters} = p_j/r_j$

while none of the nonreporters would be selected.

b. Using Predicted Participation

If it is believed that FSP participation reported in the CPS contains little useful information, an alternative that could be explored is to use predicted participation--that is, a one-zero variable predicted for each household using its observable characteristics and the coefficients from a participation equation estimated with data from a different **survey**[11] (presumably one, such as SIPP, in which the quality of reporting is better than in the CPS).

The process would entail the following steps:

(i) For each household, a probability of participation is computed analytically from the coefficients of the relevant participation equation estimated with the other data set. Let X be the vector of the characteristics of the household, and **B** the

[11]It should be noted that a separate participation equation could be estimated for different subpopulations, if there is evidence that the estimated coefficients vary substantially across such subpopulations. Sample-size constraints are likely to limit the number of subgroups for which a separate equation can be estimated.

vector of estimated coefficients from a probit participation equation. The probability of participation is expressed as:

(3) probability of participation = $\Phi(XB)$,

where $\Phi(\ )$ is the cumulative density function of the standard normal distribution.

(ii) A random number from a uniform distribution, u, is drawn for each household. The predicted participation of the household is computed as:

$$
(4)\quad \begin{aligned} &\text{predicted participation} = 1 \text{ if } u < \Phi(XB), \\ &\text{predicted participation} = 0 \text{ if } u > \Phi(XB). \end{aligned}
$$

(iii) Once each household is *classified as* a predicted participant or nonparticipant, the actual *selection* of participants within each cell would follow the same step as the selection of reporters in subsection 3.a (which involves drawing another random number for each household).

c. Combining Reported and Predicted Participation

As a third possibility, reported and predicted participation could be used concurrently. When the proportion of reporters in a cell is lower than the participation rate (case (i) in subsection **3.a),** reporters would be selected first, and predicted participants would be selected next among non-reporters. Conversely, when the proportion of reporters in a cell is larger than the participation rate (case (ii) in subsection **3.a),** predicted participants would receive higher priority among reporters.

This discussion on using an estimated probability of participation leads us to briefly address a more general point--the microsimulation application of participation rates determined analytically from the coefficients of a participation equation (logit or probit). **We** believe that predicted participation should be used only in the second stage of a participation algorithm, only after the participants are aligned with administrative data along important dimensions. Using analytic probabilities of participation as an alternative to the first stage alignment is bound to fail, because the estimated coefficients will reflect all the underreporting of participation and other types of measurement error present in the survey data used in the estimation. Thus, some form of *calibration* of the coefficients would be necessary in order to reach the administrative targets for program participants. However, this calibration would almost entirely eliminate the advantage implicit in using a predicted probability of participation.

B. THE BASE-LAW PARTICIPATION ALGORITHM IN THE FOSTERS MODEL

Before describing the base-law participation algorithm in the FOSTERS model, it is useful to recall the salient characteristics of this model:

- It relies on the SIPP, rather than on the CPS.
- SIPP contains monthly data on income and program participation, and data on most deductible expenses.
- Monthly income is thus observed, not simulated.
- Moreover, SIPP contains information on household composition collected at the same time as the income information, thus eliminating one of the major drawbacks of the CPS.
- FOSTERS simulates a past month, rather than a future month.

1. A Stylized Description of the Existing Participation Algorithm

The overall characteristics of the FOSTERS model have a significant influence on the design of the base-law participation algorithm. The monthly data collection of SIPP and the fact that FOSTERS simulates a past month imply that ***reported participation*** can have a much larger role in FOSTERS than in the MATH model. We begin by providing a stylized version of the algorithm, and then proceed to discuss some complications that arise in its implementation.

- ***The eligible*** households are cross-classified according to two dimensions (income and household size).
- The corresponding **count** of ***participants*** for each cell is obtained from administrative data.

- The eligible households that ***report*** FSP participation in the observation month are assigned participation status with a probability of one. Due primarily to the underreporting of FSP participation in SIPP, the count of reporters in a cell typically falls short of the corresponding count of participants in the administrative data. In order to reach the administrative count, the algorithm must determine which nonreporters should be selected as participants (in addition to all the reporters). The procedure for determining the probabilities of the selection of nonreporters is described in more detail in section B.2.

- Once the selection probabilities are determined, nonreporters are selected stochastically within each cell to become participants: a random number is generated and attached to each household, and the household is selected if the selection probability for the household is greater than the random number.

Let us define as **E the** matrix that contains, in each cell, the number of eligibles in a given income/household-size category; QC is the matrix that contains the number of participants in each income/household-size category; and **R is** the matrix that contains the corresponding counts of eligible reporters. Tables Al, **A.2,** and **A.3** in the Appendix present the matrices QC, E, and R used in the context of the simulation of the August 1985 FOSTERS model.

2. **Determining the Selection Probabilities for Nonreporters**

In principle, the selection probability can be determined for each income/household-size cell by dividing the difference between the number of IQCS participants and the number of SIPP reporters (the "gap to be filled") by the number of nonreporters in that cell. Using the above definitions:

$$(5)\quad S = (QC - R)/NR$$

where S is a matrix of selection probabilities, and **NR is the** matrix of eligible nonreporters. It should be emphasized that this algebraic operation is performed cell by cell. In general, this operation should yield a number between zero and one for each cell. However, in some cases, the result may be either less than zero or greater than one, a result that cannot be used as a selection probability without some adjustment. These anomalies may arise for two different reasons. (We use a subscript i to indicate a single cell in a matrix.)

(i) $S_i > 1$. The selection probability S_i is greater than one when the numerator in expression (5) above is greater than the denominator:

(6) $QC_i - R_i > NR_i$

By a simple transformation of this expression, we can see that this condition will hold if the number of FSP participants in the IQCS exceeds the number of FSP eligibles in **SIPP:**

(7) $QC_i > NR_i + R_i$ or $QC_i > E_i$

We discussed a similar problem in the context of the MATH model, with the eligibles falling short of the IQCS counts in some cells. Table A4 shows the matrix E **- QC:** a negative entry indicates a cell in which the IQCS count ***exceeds*** the estimate of eligibles from SIPP.

(ii) S_i c 0. A selection probability that is less than zero is indicative of a different problem with the data, a problem that does not arise in the context of the MATH algorithm because its design is different. The fact that an entry in the matrix S is negative implies that in the corresponding cell the IQCS count of participants is ***smaller*** than the number of SIPP reporters:

(8) $QC_i - R_i$ c 0 or $QC_i < R_i$

It is unlikely that the explanation for this phenomenon is the ***overreporting*** of FSP participation. Most likely, this problem is due to the small sample size in some cells, to the presence of measurement error in the variables that are used to cross-classify households (in this case, income and household size), or to inconsistencies between administrative and **SIPP** data. Table A.5 contains the QC - R matrix for the August 1985 FOSTERS model: a negative entry represents a cell in which the number of SIPP reporters exceeds the number of participants in the IQCS data set.

Table A.6 is a tabulation of the S matrix of selection probabilities, obtained by applying formula (5) with no corrections for cases outside the unit interval. An inspection of Table A.6 shows that the values of about 40 percent of the cells are either-negative or greater than one.

In implementing the 1985 FOSTERS model, three expedients were used to deal with these **out-**of-range values: collapsing rows or columns in the S matrix, bounding from above the value in a cell when it exceeded one, and bounding a cell value from below when it was negative. We now comment on these solutions in more detail.

1. ***Collapsing some rows and columns. This is*** a natural first approach to deal with the problem of out-of-range values. The collapsing tends to alleviate the small-sample-size problem. Table A.7 illustrates the result obtained by reducing the size of the matrix of selection probabilities to 10 (income) by 4 (household size), from the initial 12 by 8.[12] Although the "problem" cells are reduced in number, they do not disappear entirely. Of 40 cells, 11 still have out-of-range values. Cells whose ratio is greater than one remain for larger households whose incomes are above $500, and for all households up to size 5 whose incomes are between $100 and $299. Cells with negative values (the number of reporters is larger than the IQCS number) are still found throughout the matrix

2. ***Values greater than one replaced by ones.*** Eliminating values that exceed one implies that in the corresponding cells the IQCS target is scaled down to equal the number of eligibles. This loss implies that the overall number of simulated participants will tend to be below the total number of participants in the IQCS data.

3. ***Values less than zero replaced by zeroes. This*** solution does not reduce the ability of the model to reach the IQCS total. It implies that no additional households will be selected to participate from the cell, beyond those households that are explicitly reporting participation.

From the standpoint of the replication of the ***overall*** IQCS count of participants, the corrections 2 and 3 tend to compensate each other. However, they do not compensate each other in terms of the ability of the model to replicate the ***joint*** distribution of IQCS participants by income and household size. Table A.8 contains the final matrix of selection probabilities, obtained after the application of the three solutions.

Table 2 compares the percentage distribution of FSP participants by income and household size found in the IQCS data with that produced with SIPP data by the simulation described above. In addition, the table compares mean income and household size, as well as the total (weighted) number

---

[12]In addition, the income categories $500-599 and $600-699 were collapsed only for household sixes one and two.

of participants observed in the IQCS (7.114 million) with that obtained through the simulation (7.054 million). The difference between these two figures represents the *net loss of participants* from applying the adjustments (2) and (3) to the selection probabilities for eligible nonreporters. This loss represents less than 1 percent of the total IQCS count of participants.

TABLE 2

COMPARISON OF ACTUAL AND SIMULATED DISTRIBUTIONS
OF FSP PARTICIPANTS BY INCOME AND HOUSEHOLD SIZE

Income Class			Household Size		
	IQCS	Simulated		IQCS	Simulated
Zero	6.86%	6.92%	1	33.55%	33.67%
l-99	1.94	1.96	2	21.32	21.54
loo-299	24.49	24.17	3-5	37.91	38.13
300-499	40.62	40.97	6+	7.21	6.68
500-599	9.87	9.08			
600-699	6.33	6.92			
700-799	3.26	3.23			
800-899	2.43	2.20			
900-999	1.48	1.49			
1000+	2.73	3.05			
	100.00	100.00		100.00	100.00
Mean Gross Income	$398	$402		-	-
Mean Household Size	-	-		27	2.6
Total Count	7,114	7,054		7,114	7,054

Source: SIPP 1985 FOSTERS model; August 1985 IQCS data

While creating a relatively small loss on the total count of participants, the adjustment procedures induce only minor distortions in the distribution of participants by income and household size. Only in some of the categories that contain less than lo-percent of the sample is the discrepancy between the simulated and the IQCS distribution greater than 5 percent--for example, the $500-599 income class and the 6+ household-size category. These figures suggest that the algorithm used in the FOSTERS model is able to align simulated participants to IQCS counts rather well. Based on the results for the 1985 model described above, enhancements of the model to improve its ability to fit the income and household-size distributions do not seem warranted.

However, aligning the model along these two dimensions does not guarantee that other dimensions are also well aligned. (By "other" dimensions we mean those variables that are not used to compute selection probabilities for nonreporters.) Table 3 compares the FOSTERS-simulated participants and the IQCS participants along the following dimensions: the receipt of public assistance, the presence of elderly persons, the receipt of earnings, the presence of school-age children, and the distribution of and the average FSP benefits. These figures suggest that the algorithm performs rather poorly along these dimensions. The only exception is the level of FSP benefits, for which both the IQCS mean and the overall distribution are replicated very closely.

TABLE 3

COMPARISON OF ACTUAL, AND SIMULATED DISTRIBUTIONS
OF FSP PARTICIPANTS BY VARIOUS CHARACTERISTICS

	IQCS	Simulated
Receiving Public Assistance	49.0%	54.4%
Elderly Person Present	21.4	26.7
Receiving Earnings	19.6	23.9
School-Age Children	46.3	42.3
FSP Benefits		
$10 or less	8.7	12.3
11 - 50	14.2	14.9
51 - 75	10.9	9.5
76 - 100	19.2	18.5
101 - 150	18.9	18.1
151 - 200	11.5	10.4
201 or more	16.7	16.3
Mean FSP Benefits	$116	$113

Source: SIPP 1985 FOSTERS model; August 1985 IQCS data

III. REFORM PARTICIPATION ALGORITHMS

This chapter describes the reform participation algorithm in the MATH and FOSTERS models and discusses possible improvements. Section A descrii the algorithm currently in use. Section B discusses the estimation of **the** ***participation*** response to a change in benefits. This is essentially an econometric problem, and represents a special case of the more general issue of estimating behavioral responses to policy changes. Section C explains how the estimated behavioral response can be incorporated into the simulation models.

A. THE EXISTING REFORM PARTICIPATION ALGORITHMS

In this section we describe and evaluate the reform participation algorithms embedded in the the MATH and FOSTERS models. The reform participation algorithms in these two models are virtually identical, and they use the same behavioral response parameter. In the remainder of this section, we will use the term "MATH" to signify both models.

1. The MATH Algorithm

The reform participation algorithm in the MATH model is rather **simple.**[13] When a reform is simulated, the algorithm distinguishes among four cases:

(i) ***Households eligible but n&participating under base law.*** By design, these cases are affected only by an ***increase in*** benefits. The fraction of previously nonparticipating households that will participate after the reform is equal to the absolute dollar amount of the increase times a **fixed** factor that represents the participation response. In more formal terms, let us define the household's potential benefits under base law as **$BASELAW,** and the household's potential benefits after the program reform as **$REFORM.** The fraction of nonparticipants who are selected randomly to become participants is equal to:

[13]**This** description is based on pages 321-43 of the ***MATH Technical Description*** (Doyle et al., 1989).

(8) fraction selected = **.0014 * ($REFORM - $BASELAW).**

For example, households that are eligible nonparticipants under base law and that are entitled to a $10 increase will be selected randomly at a rate of 1.4 percent to become participants under reform law.

(ii) ***Households eligible and participating under base law.*** By design, these cases are affected only by a ***decrease in*** benefits. Their treatment in the model is perfectly symmetrical to that of nonparticipants subject to a benefit increase. The probability that a previously participating household will not participate after reform is equal to the absolute dollar amount of the decrease times the fixed factor. In more formal terms, the fraction of participants selected randomly to become nonparticipants is equal to:

(9) fraction selected = **.0014 * ($BASELAW - $REFORM).**

(iii) ***Households not eligible under base law, becoming eligible under reform. These cases*** are selected to participate according to the base-law probabilities of participation, which guarantees that the characteristics of selected participants are similar to those found in administrative data.

(iv) ***Households eligible under base law, becoming ineligible under reform. These*** households can no longer participate in the program, and no simulation is necessary.

Before discussing the possible improvements for this algorithm, we must stress the difference between Case (iii) and Cases (i) and (ii). In Case (iii) (households ***becoming*** eligible under reform) no participation response is involved. We see no reason to deviate from the current practice of selecting participants with base-law probabilities, since in this case becoming a participant is not the behavioral response to a change in benefits, but rather a decision similar to that already made by the households eligible under base law. The only change with respect to current practice pertains to the FOSTERS model. Currently, the FOSTERS model uses the matrix of base-law participation probabilities obtained from the MATH model. We believe that this matrix should be obtained from the FOSTERS model directly.[14]

[14] The 1988 version of the FOSTERS model already implements this recommendation.

In the remainder of this chapter, we discuss only the algorithm that applies to Cases (i) and (ii) (households eligible under both base-law and reform programs). However, it is important to observe that, under many FSP reforms, the ***total*** impact of the reform is a mixture of all four cases. In other words, the distinction among the four cases is one that pertains to the economic situation of a given household under base law, not to the particular type of reform.

2. Areas for Improvement

We see three areas for improving the algorithm that simulates the participation response to a change in benefits among Case (i) and Case (ii) households:

- The participation response could be ***made to depend on the base-law level of the benefits*** of the household, rather than being the same for all households. For example, in the current system, given the same absolute increase in benefits, the participation response is the same for households eligible for $10 or eligible for $200 worth of FSP benefits.
- In the current system, ***no inflation adjustment*** is built into the algorithm, despite the fact that the participation response depends on the absolute dollar amount of the change in benefits. This means that the participation response has become larger and larger since the algorithm was first implemented, since inflation has caused changes in benefits to become larger in purely ***nominal*** terms.
- The selection factor currently in use (.0014) is derived from aggregate time series data on program participation in the pre-EPR program of the 1970s.[15] Therefore, this estimate, regardless of its validity at the time that it was calculated, might ***no longer reflect the reality of the current FSP.***

[15] An exact source for this estimate is no longer available.

B. ESTIMATING THE PARTICIPATION RESPONSE TO A CHANGE IN FSP BENEFITS

In the previous section, we discussed how the behavioral response embedded in the MATH and FOSTERS models no longer reflects the reality of the **FSP.** In this section, we discuss the methodological issues involved in estimating a behavioral response. We then present estimates of the participation response to a change in FSP benefits obtained from SIPP.

1. Methodological Issues in Estimating a Behavioral Response

Most of the methodological problems that must be confronted to estimate the participation response to a change in FSP benefits are common to all attempts to estimate behavioral responses to policy changes (Burtless, 1989). The research on policy evaluation conducted in the last three decades has produced two major methodological approaches for estimating behavioral responses to policy changes--one based on controlled experiments, the other on behavioral models estimated with nonexperimental data.[16]

The essence of ***experimental methods is*** to subject a sample of individuals to a "treatment," designed to mimic as closely as possible the policy change that is being studied. Depending on the specific context, a variety of methods can be used to compare the observed response to a counterfactual behavior, such as that of a similar group not subject to the treatment (a control group) or that of the treatment group itself before the treatment (a pre-post comparison). If the counter-factual is chosen appropriately, the comparison yields a measure of the response that precludes having to model behavior explicitly.

The fact that the Food Stamp Program is an entitlement program precludes implementing a controlled experiment to study the participation response to FSP benefit changes. Such a hypothetical controlled experiment would in fact require changing the level of benefits by a nontrivial amount either for a randomly selected group of eligiile households or for all eligible households in

[16]Although this distinction is not always clear-cut, it is used here for simplicity.

randomly selected geographical locations, in order to observe the changes in the participation rate that follow the change in benefits.[17]

A result conceptually similar to that obtained with controlled experiments can be provided by "natural experiments," which occur when a policy change is actually implemented, and its impact on the target population can be observed. To obtain a correct behavioral response from an analysis of natural experiments, one must adequately control for all the possible confounding factors. This requirement is rarely met in practice, particularly when the analysis is conducted with aggregate time-series data. In the specific case of the FSP, too few and too small changes in the program have been implemented after the elimination of the purchase requirement in 1979 to allow us to disentangle their effect on participation from that of the other determinants of participation. Moreover, as recent experience has indicated, rather dramatic changes in FSP participation can occur without any significant change in the level of benefits (Mathematica Policy Research, 1990).

When experimental data cannot be generated and natural experiments are not available, econometric techniques that rely on nonexperimental data can be used. We focus particularly on the use of cross-sectional household survey data.[18] A useful distinction among types of econometric analyses of nonexperimental data can be made between structural models and reduced-form models.

---

[17]An increase in potential benefits in selected geographical locations would actually elicit two different responses--a participation response among currently eligiile households plus a response from households that are not currently eligible but might change their behavior in order to take advantage of a more generous program. Disentangling the two effects might be very difficult.

[18]We ignore here the possibility of using panel data. In principle, panel data offer two advantages: (1) the occurrence of actual policy changes coupled with observations on the same households for a long enough time period allows the analyst to observe the behavior of the same household under different policy regimes; and (2) the repeated observations allow the analyst to examine changes in behavior, rather than ***differences in*** behavior, as is the case with cross-sectional data. In practice, however, several factors prevent us from using panel data to estimate a participation response: very few reforms that affect the benefit amount have been implemented in recent years; a substantial fraction of the "change" observed in longitudinal data is due to measurement error; and even in the most complete longitudinal data sets (e.g., SIPP), most of the information necessary for simulating eligibility and the benefit amount is not available for all the subperiods covered by the survey.

Structural models attempt to capture the underlying determinants of observed behavior. A structural model contains parameters that represent the important elements involved in the decision-making process under investigation (such as the individual's tastes and preferences, the prevailing prices for goods and the factors of production, the information available to the individual, the parameters that describe existing policies, and the relevant characteristics of the economy). If individual behavior is modelled correctly (and the necessary data are available), the estimated parameters of the model can be used to predict how individuals would react when the constraints they face change--particularly, how they would react to changes in the policy parameters. The weakness of structural models is that a great deal must be known about how people make decisions to make these models a credible representation of reality; moreover, the data requirements of these models are rarely satisfied by existing data sets.

Designing a structural model of FSP participation would be a very challenging undertaking. The existing evidence on the reason for nonparticipation among FSP eligibles indicate that stigma, the costs of participation, and a lack of knowledge about the FSP are three important determinants of participation (General Accounting Office, 1988). We believe that a credible structural model of FSP participation should model the following *explicitly*: (1) the trade-off between the costs of participation (such as stigma) and the utility to be derived from additional disposable income; (2) the sources of the costs of participation; and (3) how knowledge about program regulations is disseminated among the eligible population. Such a model would require an almost completely different framework than the income-leisure choice model commonly used to analyze the labor supply responses to changes in the welfare system. Moreover, estimating such an ambitious model would require data that are not currently available. We believe that developing a structural model of FSP participation is a task worth pursuing, but one that is not within reach given the current state of knowledge on this topic.

This state of affairs limits our choice to a ***reduced-form*** approach. Reduced-form models simply attempt to capture the correlation between observed outcomes and observed characteristics. These models produce estimates of the "net effect" of a unit change in one of the explanatory variables on the outcome of interest, when all the other variables are held constant. However, the interpretation of such net effects as behavioral responses is not always possible. Such interpretation implies that, if households in group A and B differ along characteristic X (e.g., a program variable such as benefits) and this difference is associated with a difference in the outcome (participation rate), then type-A households start behaving like type-B households (participate at the same rate) when given the same value of X that type-B households currently have.

Some necessary, although not sufficient, conditions can be identified for this assumption to be valid. The most important of these conditions is that the variation in the policy variable across is exogenous; that is, the variation does not depend on choices made by the households. This condition requires that the model be specified correctly, so that ***all*** the determinants of the outcome that are also correlated with the policy variable are held constant. In the example above, if type-A and type-B households quality for different benefits amounts, but another factor (e.g., income) is determining the difference in ***both*** benefits and the participation outcome, changing the benefits amount for type-A households will not induce them to participate at exactly the same level as type-B households, unless their level of income is also changed.

A second condition that must be satisfied to interpret cross-sectional differences as behavioral responses pertains to the speed of the adjustment to the policy change. In general, cross-sectional differences in participation can be used more appropriately to approximate the ***long-run response*** to a program reform. In order to use them as an approximation of a short-run response, one must assume that **eligible** households adjust to a program reform very rapidly by modifying their behavior accordingly. One essential element for justifying this assumption is that households be well informed about program characteristics--that is, that they know pre-reform regulations and are aware about when and how they are changed. The available evidence suggests that a large proportion of eligible nonparticipants report that the reason they did not participate was that they did realize they were eligible (General Accounting Office, 1988). Eligible households that are not informed about their eligibility are not likely to know the amount of the benefit to which they are entitled, and are even less likely to know whether this amount is being altered by a program reform.

For these reasons, the adjustment of the participation rate to a program reform that increases the size of the benefit is bound to be slow.[19] Eventually, the program caseload reaches a new equilibrium. It is this new equilibrium level that one can hope to forecast by using cross-sectional differences in participation by the amount of the benefit. However, most policy simulations are applied to the short run--what happens to the program caseload and expenditures in the year immediately following the reform. Such a short-run response could very well differ from the long-run response.

2. Estimates of Participation Response from SIPP Data

In the previous section, we argued that reduced-form models and cross-sectional data, despite their shortcomings, are the only viable solution for estimating a participation response to changes in FSP benefits. SIPP provides the best available cross-sectional data on household income, assets, and program participation on a sub-annual basis. This data set has been used by Allin and Martini (1990) to conduct a multivariate analysis of FSP participation. Allin and Martini estimate a reduced-form participation equation in which (1) the universe is represented by households that are simulated to be eligible for the FSP on the basis of their characteristics in August 1985; (2) the dependent variable

[19] We are considering here relatively "small" changes in the program. The adjustment to large, dramatic changes might be faster, but large changes are seldom implemented, and their effect is much more difficult to predict on the basis of the past history of the program.

is reported FSP participation status in the same month; and (3) the explanatory variables include household demographic and economic characteristics, as well as the benefit amount for which the household is (simulated to be) eligible. The estimated coefficient on the benefit variable in the participation equation can be interpreted as the participation response to a benefit change (with the *caveat* stated in the previous section.)

Allin and Martini discuss several methodological issues relevant to their estimation. We recall the most important:

- ***Specification of the benefit variable. While*** intuition suggests that the relationship between participation and the benefit amount should be non-negative (i.e., either zero or positive), intuition is of less help in suggesting the ***shape*** of such a relationship. Allin and Martini experiment with three different assumptions about how the benefit variable enters the participation equation: linear (as is the case in most of the literature), piecewise linear, and logarithmic. They conclude that the logarithmic specification is preferable. A logarithmic specification implies that the change in the probability of participation is proportional to the ***percentage*** change in benefits, rather than to the ***absolute*** change (which would be implied by a linear specification[20]).

- ***Inclusion of household income and size.*** Allin and Martini include both household gross income and household size in the participation equation. Thus, the estimate of the benefit effect they obtain is ***net*** of income and size effects.

- ***Low variation in the benefit amount. The*** FSP benefit amount varies only to a limited extent among households of the same size and with the same gross income, because FSP benefits are computed with a formula that includes the maximum allotment (a function of household size) and net income (which is equal to gross income minus allowable deductions).[21] Moreover, due to the institutional

[20]For example, a linear specification implies that a $10 increase has the same effect for households currently receiving $50 or $200 in benefits. A logarithmic specification implies that a $10 increase has an effect on participation that is four times as large for a household receiving $50 than for a household receiving $200. The behavioral response currently in use in the MATH and FOSTERS models assumes that the change in the probability of participation is proportional to the ***absolute*** change in benefits.

[21]For households of the same size and with the same total income-, benefits can vary for a number of reasons: categorical differences (e.g., elderly and disabled households are allowed a deduction for medical expenses, and have no cap on shelter expenses); differences in sources of income (e.g., the (continued...)

characteristics of the FSP, the benefit amount does not vary by geographic area, as does the Aid to Families with Dependent Children program.[22] This lack of variation implies that, when a cross-section of households is used for estimation, it might be difficult to identity the effect of the benefit amount on participation separately from the effects of income and household size.

- ***Benefit amount not observed for nonparticipants. Since*** the benefit variable is not observed for nonparticipants, it must be either imputed or simulated on the basis of the household's demographic and economic characteristics as reported in the survey. Thus, the simulated or imputed benefit variable is sensitive to a wide range of reporting errors and missing information. For example, households that underreport income during the interview are simulated to be eligible for a benefit amount larger than the amount for which they are actually eligible.

The estimates of the behavioral response obtained by Allin and Martini using the three different specifications (linear, piecewise linear, and logarithmic) are presented in Table 4. The table contains the percentage change in the probability of participation for an average household, and for a hypothetical $10 change in benefits.

According to the linear assumption, a $10 increase in benefits is associated with approximately half of a percentage point increase in the probability of participation, regardless of whether the increase involves a household that currently receives, say, $10 or $220 worth of benefits.

The participation response implied by the two other specifications differ considerably from those implied by the linear specification. The piecewise linear specification allows a more flexible response, but also an "irregular" one. We obtain a negative (albeit small) response in the $80 to $150 range, and a positive response in all other ranges. The negatively sloped segment can easily be seen in Figure 1.

[21](...continued)
benefit reduction rate is lower for earnings than for unearned income); and differences in expenses (e.g., some households have child care expenses while others do not).

[22]This is true for the continental United States, while a different maximum allotment is used in Alaska, Hawaii, and other U.S. territories.

TABLE 4

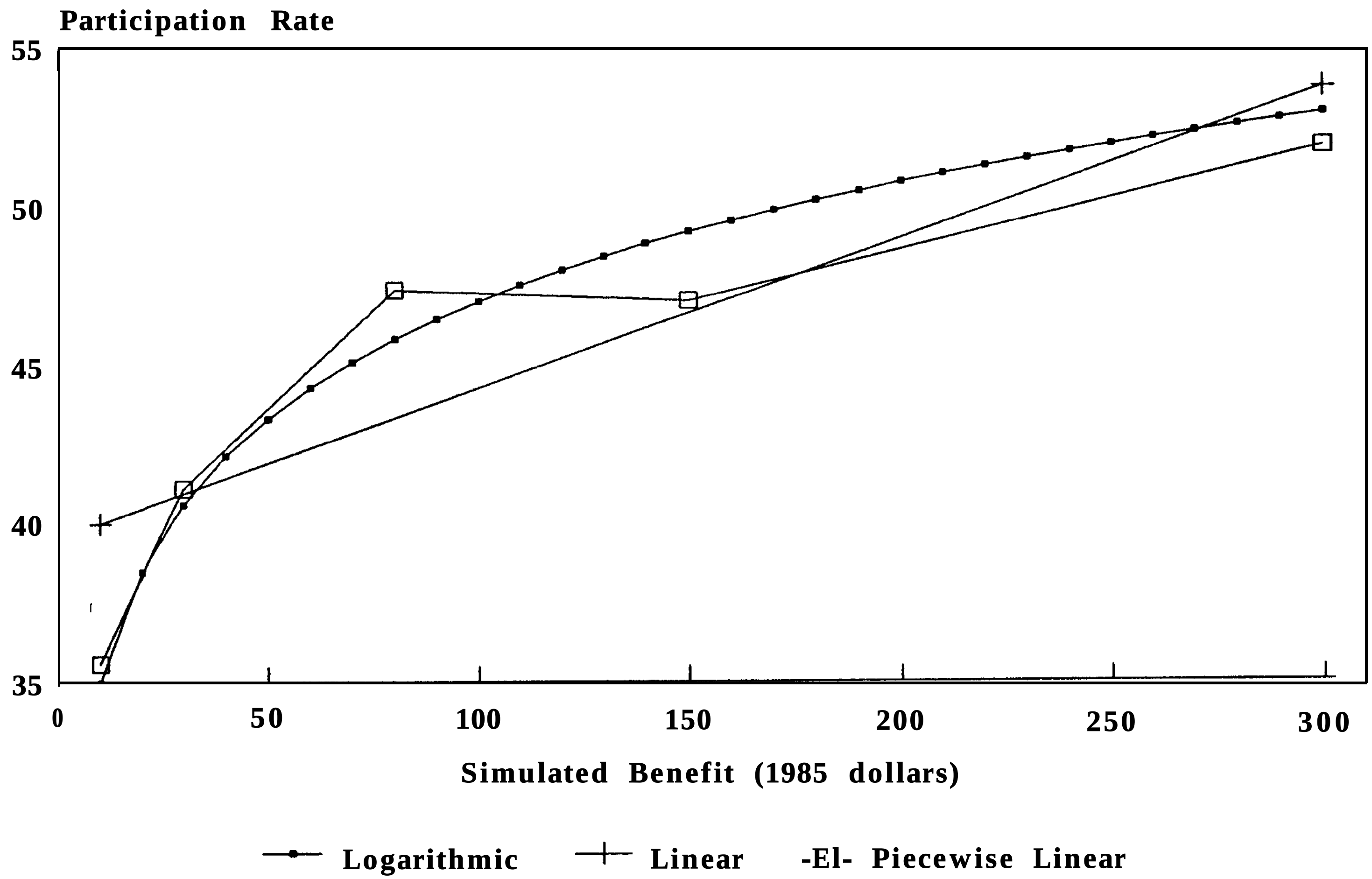
ESTIMATES OF THE CHANGE IN PARTICIPATION
ASSOCIATED WITH A $10 INCREASE IN BENEFITS,
COMPUTED AT ALTERNATIVE INITIAL LEVELS OF BENEFITS

	Specification of the Benefit Variable		
Level of Benefit	Linear	Piecewise Linear	Logarithmic
	Percentage Point Change in Participation:		
$10	.462	2.69	3.52
$30	.465	1.24	1.52
$80	.470	-.51	0.63
$150	.475	0.32	0.35
$220	.477	0.32	0.25

SOURCE Allin and Martini (1990).

The logarithmic specification follows a pattern very similar to that of the piecewise linear specification, as shown in Figure 1. In terms of the participation response to a $10 change in benefits, the logarithmic specification implies a 3.5 percentage point increase in the probability of participation among households currently entitled to $10 worth of benefits (for whom, in other words, benefits would double), but a much smaller response, a quarter of a percentage point, among those entitled to a $220 benefit. This concave pattern--that is, always increasing but at a decreasing rate--is a mathematical property of the logarithmic function. However, this is roughly the pattern followed by the piecewise linear specification. In this context, the logarithmic specification seems to be a defensible way to "smooth out" the irregular, pattern created by the piecewise linear specification.

FIG. 1 ALTERNATIVE SPECIFICATIONS FOR
THE BENEFIT-PARTICIPATION RELATIONSHIP
ALL ELIGIBLE HOUSEHOLDS



C. INCORPORATING THE PARTICIPATION RESPONSE INTO THE SIMULATION MODELS

In this section we discuss how the SIPP-based participation response (presented in the previous section) can be incorporated into these models as an alternative to the estimate currently in use.

The Allin-Martini estimates of the FSP participation response presented in section A.2 can form the basis for a new algorithm that improves upon the existing one. Two features of the Allin-Martini estimates are particularly relevant. One is their timeliness: using SIPP data from 1985 provides estimates that reflect the reality of the post-EPR FSP program. The other important feature is that the participation response was estimated using a logarithmic specification, implying that the change in the participation rate depends on the *percentage* change in the benefit amount, rather than on the ***absolute*** change. Using the percentage change seems more in line with intuition. Moreover, using the percentage change automatically corrects for the lack of an inflation adjustment.

While stressing the positive features of the All&Martini estimates, we should also remember that a great deal of uncertainty surrounds the validity of these estimated behavioral responses. Due to limitations in the data and to possible model misspecifications, these estimates might still be very far from capturing the true participation response to benefit changes. Nevertheless, an algorithm based on these estimates is likely to represent ***a significant advance*** over that currently in use.

With this caveat in mind, we now discuss the more technical aspects of incorporating the SIPP estimates into the existing microsirnulation models. The first issue pertains to the fact that the SIPP estimates cannot be transposed mechanically into the model and substituted for the existing selection factor (.0014), due to the fundamental inconsistency between the marginal effects obtained from a participation equation and the selection factor used to simulate reforms. The next section explains how such inconsistency arises and how it can be resolved.

1. Resolving the Inconsistency between the Response Estimate and the Selection Factor

To simulate the participation response to a benefit increase among eligible households (Case (i), Section Al), the existing algorithm selects a fraction of the households not participating under base law and makes them participants. Analogously, to simulate the effect of a benefit decrease--Case (ii)--a fraction of participants are selected to become nonparticipants. In other words, the selection is never applied to the entire universe of eligibles, but rather to the two subsets of such a universe.

We believe that this feature of the existing algorithms should be maintained. However, the behavioral response produced by estimating the participation equation is equivalent to a ***percentage change in the participation rate among eligibles.*** To be used in simulation, this measure must be transformed into a selection factor that can be applied to either the subset of base-law nonparticipants or the subset of base-law participants.

An example will clarify the point. Let us assume that the **FSP** reform under simulation consists of a $10 ***increase in*** benefits for all base-law eligible **households.**[23] In the current reform participation algorithm, 1.4 percent (equal to **.0014** x $10 x **100)** of base-law nonparticipants are selected randomly to become participants. However, the change in the participation rate among **base-**law eligibles is different **(*smaller*)** than 1.4 percent. More precisely, the change in the participation rate induced by the reform simulation depends on the level of the base-law participation rate itself. ***The higher*** the base-law participation rate, the ***lower*** the absolute number of nonparticipants who can be affected by the increase in benefits, and the ***lower*** the effect of the reform on the participation rate. Table 5 illustrates how the change in the participation rate varies according to the pre-change level of the participation rate. In both cases, we use a base of 10,000 eligiiles and apply the same selection factor **(.0014).**

---

[23]This is an unrealistic example because most reforms that modify the FSP benefit formula also affect the number of eligible households. However, the lack of realism does not affect the usefulness of the example.

TABLE 5

HOW THE CHANGE IN THE PARTICIPATION RATE VARIES
ACCORDING TO THE BASE-LAW PARTICIPATION RATE

Base-Law Participation Rate	Base-Law Nonparticipants	New Participants	Participation Rata After Reform	Percentage Change in the Participation Rate
30%	7,000	98	31.0%	1.0%
80%	2,000	28	80.3%	0.3%

NOTE: The number of new participants is computed as (.0014) * ($10) * (number of base-law nonparticipants).

It follows that the participation response estimated in a participation model must be appropriately transformed to be applied in the same way as the current selection factor used in the MATH model. To understand how this transformation can be implemented, we must introduce a few simple relations. Let us define as μ the MATH-type selection factor, used to select a fraction of base-law nonparticipants (or participants). In the case of a $1 change in benefits, we have:

(10) new participants = μ * (base-law eligible nonparticipants)

or

(11) ex-participants = μ * (base-law participants).

Let us define δ as the change in the probability of participation estimated from the participation equation. In other words, depending on the direction of the change in benefits, the following is the behavioral interpretation of 6:

(12) new participants = δ * (base-law eligibles) for a benefit increase

or

(13) ex-participants = δ * (base-law eligibles) for a benefit decrease.

The relationship between μ and 6 can easily be established by equating the two right-hand sides of equations (10) and (12) (for the case of benefit increase):

(14) $\mu^{increase}$ * (base-law eligible nonparticipants) = δ * (base-law eligibles)

from which,

(15) $\mu^{increase}$ = 6 * (base-law eligibles)/(base-law eligible nonparticipants)

and

(16) $\mu^{increase}$ = δ /[(base-law eligible nonparticipants)/(base-law eligibles)],

which can be rewritten as:

(17) $\mu^{increase}$ = δ /[1 - (base-law participation rate)],

where the participation rate is expressed as a probability, that is, a number between zero and one.

Equation (17) tells us that the participation response obtained from the participation equation can be *transformed* directly into a MATH-type selection factor simply by dividing it by one minus the base-law participation rate. For example, a participation response equal to .00056, combined with a 60 percent base-law participation rate, would translate into a .00056/(1 - .60) = .0014 value for the corresponding MATH-type selection factor.

Analogously, for benefit decrease, we can equate the two right-hand sides of equations (11) and (13):

(18) $\mu^{decrease}$ * (base-law participants) = 6 * (base-law eligibles).

Using a procedure similar to the one used for a benefit increase, we obtain:

(19) $\mu^{decrease}$ = δ / (base-law participation rate).

Equation (19) tells us that for a benefit decrease the participation response obtained from the participation equation can be transformed into a MATH-type selection factor simply by dividing it by the base-law participation rate.

2. **Operationalizing the New Algorithm**

The new algorithm can be operationalized in two steps. First, the MATH-type selection factor, μ, is computed for each household. Second, a random number is drawn for each household, and the comparison of the random number with the selection factor determines **which** households change their base-law participation status under reform. The following is a more detailed description of these two steps.

Step 1. The MATH-type selection factor, μ, is computed for each household, distinguishing between the benefit increase and the benefit decrease. The necessity of computing μ for each household stems from the fact that, unlike in the current simulation models, the participation response varies from household to household according to observable characteristics. For a benefit increase, the formula for the selection factor is derived by adding **to equation** (17) the dollar amount of the increase:

$$(20)\quad \mu_i^{\text{increase}} = [\delta_i \,/\, (1 - \text{base-law participation rate})_i] * [\text{dollar amount of the increase}]_i$$

All the terms in (20) vary from household to household, as indicated by the subscript i. The term δ_i is computed by using the estimated coefficients of the **probit** participation equation and the characteristics of the ***i*th** household, in the following way:

$$(21)\qquad \delta_i = [\, \beta \,/\, (\text{base-law benefits amount})_i \,] * \phi(X_i B).$$

The term β is the coefficient of the logarithm of benefits in the probit participation equation. This coefficient is divided by the benefit amount for which the household is eligible under base law.[24] The term $\phi(X_i B)$ is required to transform the probit coefficient into a marginal effect (see Allin and Martini, 1990): $\phi(\ )$ is the density of the standard normal, X is the vector of household characteristics used as explanatory variables in the participation equation, and B is the corresponding vector of probit coefficients.

Another term in equation (20) that must be discussed is the term that contains one minus the base-law participation rate. Its role is to convert the marginal effect obtained from the participation equation into the MATH-type selection factor. The numerical value for this conversion factor varies from household to household and is equal to the base-law probability of participation, that is, the probability by which each eligible household was selected to participate under base law.[25]

Equation (20) is modified as follows for a decrease in benefits:

$$(22)\quad \mu_i^{\text{decrease}} =$$

$$[\delta_i \,/\, (\text{base-law participation rate})_i] * [\text{absolute dollar amount of the decrease}]_i$$

The only differences between (20) and (22) are that the ***absolute*** value of the decrease in benefits is used, and that the conversion factor is the base-law participation rate rather than one minus the base-law participation rate.

---

[24] This is due to a mathematical property of the logarithmic function: the first derivative of $f(x) = \beta \log(x)$ with respect to x is equal to β/x.

[25] It should be mentioned that for certain types of eligible households, the base-law probability of participation is equal to one, that is, they are selected with certainty to participate in the program see Chapter II). ***Prima facie,*** this might appear to be a problem in applying equation (13), because it would involve a division by zero. However, equation (13) is ***never*** applied to households with a base-law probability of participation of one: none of these households can become a ***new*** participant, because they are all already participating!

Step 2. Once the selection factor is computed for the each household, the second step of the algorithm consists of drawing a random number from a uniform (0,1) distribution for each household. If the random number is less than the selection factor μ, the nonparticipating household becomes a participant (benefit increase), or the participating household becomes nonparticipant (benefit decrease).

IV. SUMMARY OF RECOMMENDATIONS

Following the structure of the report, we summarize our recommendations separately for the base-law and reform algorithms.

A. RECOMMENDATIONS FOR THE **BASE-LAW** PARTICIPATION **ALGORITHMS**

The first set of recommendations concerns the adjustment procedure used in the first phase of the MATH algorithm. This procedure attempts to correct for the "excess" of participants with respect to eligibles in some of the cells of the **96-cell** matrix defined by income, household size, receipt of public assistance, and elderly status (primary cells). The current practice involves the use of a computer algorithm that **redistributes** excess participants to other cells, followed by a manual calibration. We recommend exploring the feasibility of an algorithm that would avoid any manual intervention, and propose two alternatives: (1) exploring the applicability of existing raking procedures; (2) developing an ad ***hoc*** algorithm that redistributes participants using a set of "rules of thumb" similar to those used by an informed and intelligent "calibrator." In either case, the ability of the algorithm to avoid any subsequent manual calibration ***cannot*** be determined a priori.

The second set of recommendations concerns the subsequent phase of the MATH participation algorithm, in which the **96-cell** matrix of participation rates is expanded to 1,152 rates by incorporating information on the benefit amount and reported participation. The participation rates by the level of benefits are computed using the coefficients of a participation equation estimated by Czajka (1981): we have doubts on the appropriateness of the use of these coefficients in microsimulation. We believe that, if replicating the benefit distribution observed in program data is an important objective, the benefit variable should be added as one of the key dimensions in the first phase of the algorithm, possibly replacing either income or household size. In addition, we suggest that the selection of participants ***within*** each of the primary cells give priority to: (1) households that report participation; (2) households predicted to participate based on an estimated participation equation; or (3) a combination of predicted and reported participation.

Regarding the FOSTERS model, we conclude that the procedure currently in use is quite sensible and no enhancements are warranted to improve the ability of the model to reproduce the distribution of simulated participants by income, household size, and benefit amount. However, we find that the simulated participants do not "fit" the QC participants along other dimensions, such as receipt of public assistance and earnings.

B. RECOMMENDATIONS FOR THE REFORM PARTICIPATION ALGORITHMS

The bulk of our recommendations concern households already eligible under base law. (For households ***becoming*** eligible under reform law, we recommend only that they be selected using the currently observed participation rates.) Households eligible under base law face two choices that must be simulated by the model. If they are not participating before the reform, an ***increase in*** benefits might induce them to participate. On the other hand, if they are already participating, a decrease in benefits might induce them to leave the program. The algorithm currently in use in both MATH and FOSTERS first computes a probability equal to 0.0014 times the ***absolute dollar amount*** of the benefit change. Then, in the case of benefit reduction, a fraction of base-law eligible participants is randomly selected--using the above probability--to become nonparticipants. Symmetrically, in the case of a benefit increase, a fraction of base-law eligible nonparticipants is selected to become participants.

We find three major weaknesses in this algorithm: (1) the participation response only depends on the absolute amount of the increase, not on the pre-reform level of benefits; (2) there is no inflation adjustment built into the algorithm, (3) the behavioral response parameter currently in use (.0014) was derived from aggregate time series data on program participation in the pre-EPR program of the 1970s.

After an in-depth discussion of the methodological difficulties implicit in the estimation of the behavioral response to a benefit change, we present estimates of this behavioral response derived from 1985 SIPP data. Building on these estimates, we propose a new reform participation algorithm, with the following features: (1) the behavioral response parameter is based on a logarithmic specification of the participation equation, so that the participation response depends on the percentage change in benefits, rather than on its absolute amount; (2) the use of the percentage change in benefits includes an automatic inflation adjustment; and (3) the use of 1985 data reflects the reality of the post-EPR program.

REFERENCES

Allin, Susan, and Alberto Martini. "A Multivariate Analysis of Participation in the Food Stamp Program." ***Current Perspectives on Food Stamp Program Participation.*** Alexandria, VA Food and Nutrition Service, U.S. Department of Agriculture, 1991 (forthcoming).

Bishop, Yvonne, Stephen Fienberg, and Paul Holland, ***Discrete M&variate Analysis: Theory and Practice*, *The* MIT** Press, Cambridge, MA, 1975.

Burtless, Gary. "Labor Supply Response in Microsimulation Models." Paper commissioned by the CNSTAT Panel to Evaluate Microsimulation Models of Social Welfare Programs, 1989.

Doyle, Pat. "Food Stamp Program Participation Rates: August 1985." ***Current Perspectives on Food Stamp Program Participation. Alexandria,*** VA Food and Nutrition Service, U.S. Department of Agriculture, 1990.

Doyle, Pat, and Carole Trippe. "Validation of the Food Stamp Microsimulation Model." Final Report prepared for the Food and Nutrition Service, U.S. Department of Agriculture. Washington, DC: Mathematica Policy Research, Inc., 1989.

Doyle, Pat, et al. ***MATH Technical Description.*** Washington, DC: Mathematica Policy Research, Inc., 1989

General Accounting Office. "Food Stamps: Reasons for Nonparticipation." Washington, DC: U.S. GAO, 1988.

Mathematica Policy Research, Inc. ***Recent Trends in Food Stamp Program Participation. A Preliminary Report to Congress.*** Alexandria, VA Food and Nutrition Service, U.S. Department of Agriculture, 1990.

APPENDIX A

DETAILED TABLES FROM THE 1985 FOSTERS MODEL

TABLE A.1

QC MATRIX:
NUMBER OF FSP PARTICIPANTS FROM THE IQCS FILE
(IN THOUSANDS)

Income	Household Size 1	2	55	4	5	6	7	8+
Zero	278	75	17	44 7	17 5	10	3 3	6
$1-99	74	32	90	36	19	0	1	
$100-199	399	127	226	97	42	5	7	8
$200-299	363	311	230		51	14		3
$300-399	988	359	208	182	85	23	9	2
$400-499	249	312		148				
$500-599	26	188	219	132	75	27		10
$600-699	4	86	113	134	58	38	1:	13 10
$700-799	2	15	53	51	58	32	13	9
$800-899	4	4	24	40	37	43 27	11	10
$900-999	0	1	13	30	25	19	13	4
$1000+	0	7	0	28	48	38	33	40

TABLE A.2

E MATRIX:
NUMBER OF ELIGIBLE HOUSEHOLDS FROM THE FOSTERS SIMULATION ON SIPP DATA
(IN THOUSANDS)

Income	Household Size 1	2	3	4	5	6	7	8+
Zero	381	163	56	58	29	11	5	8
$1-99	176	52	32	28	8	4	0	4
$100-199	284					3		7
$200-299	464	257 179	285	45 79	22 51	22	10 6	5
$300-399	1,579	312		158	47	17	17	14
$400-499	996	409	240	165	74	19	14	7
$500-599	684	410	255	160	77	21	37	3
$600-699	74	388	170	116	67	14	4	13
$700-799						38	4	3
$800-899	20	138	149	113	55	32	14	0
$900-999	10	16	183	106	44	11	9	19
$1000+		8	B	191	277	108	72	97

TABLE A.3

R MATRIX:
NUMBER OF ELIGIBLE HOUSEHOLDS REPORTING FSP PARTICIPATION IN SIPP
(IN THOUSANDS)

Income	Household Size 1 117	2	3	4	5	6	7	8t
Zero		61	27	13	22	3	0	3
$1-99	56	27	18	11	0	3	3	0
$100-199	180	121	99	28	13			7
$200-299	173	183	130	73	42		6	5
$300-399								
$400-499	210	213		144	32	1618	15	
$500-599	32 724	119	21 66	14 72	44 50	4	11 33	12 3
$600-699	8	19	45	51	42 35	-34	4	13
$700-799	0	4	28	33		24	2	
$800-899	0	0	12	25	31	11	11	7
$900-999					8		9	13
$1000+	0	0	0	22	75	19	18	35

TABLE A.4

E - QC MATRIX:
DIFFERENCE BETWEEN SIPP ELIGIBLES AND IQCS COUNTS
(IN THOUSANDS)

Income	Household Size 1	2	3	4	5	6	7	8+
Zero	103	88	15	14	12	4	2	2
$1-99	102	20	23	21	3	-2	-3	4
$100-199		52		9			9	5
$200-299	-115	-54	55	-18	-4	8	-1	2
$300-399	501	97	-32	-17	-11	-6	7	12
$400-499	658	222	36	28	2	-8	26	-3
$500-599	70	302	57	-18	9	-17 -18		-10
$600-699	21	158	96			11	-9	3
$700-799	7	34	131	62	11	-11	-13	-6
$800-899	0	15	70	116 76	18		-3	-10
$900-999					19	-8		15
$1000+	0	8	8	191	277	108	72	97

TABLE A.5

QC - R MATRIX:
DIFFERENCE BETWEEN IQCS COUNTS AND SIPP FSP REPORTERS
(IN THOUSANDS)

Income	Household Size 1 161	14	3	4	5	7	7	8t
Zero		5	28	31	-5	-3	0	3
$1-99	18	6	-1	-4	5	2	-3	0
$100-199	219	128	-9	24	6	-2	1	-5
$200-299	190		96	38	0		-6	-2
$300-399					7			
$400-499	264 39	146 96	12	5	53	13 9	-4	-10 3
$500-599	-6	40	60		25		-22	10
$600-699		-33			16	24		
$700-799	4	-4	8	6:7	23	28 -7	150	-3 6
$800-899	0	-0	-4	5	6	19 8	4	10
$900-999		1	1		17			-9
$1000+	0	0	0	-22	-75	-19	-18	-35

TABLE A.6

S MATRIX:
UNCORRECTED SELECTION PROBABILITIES

Income	Household Size 1	2	3	4	5	6	7	8t
Zero	0.61	0.14	0.97	0.70	-0.73	0.89	-0.06	0.60
$1-99	0.15	0.19	-0.68	-0.22	0.63	-2.20	0.00	0.00
$100-199	2.11	0.10	2.31	0.49	0.67	0.00	-0.26	0.00
$200-299	0.65	1.48	[illegible]	3.60	0.02	-0.37	0.00	0.00
$300-399	0.31	0.50	-0.01	2.69	2.74	1.93	-2.12	-3.64
$400-499	0.05	[illegible]	0.27	0.23	1.26	10.43	-1.55	0.00
$500-599	-0.01	0.15	0.62	0.19	0.92	3.36	-5.62	0.00
$600-699	0.05	-0.12	0.45	1.42	0.64	2.77	0.00	0.00
$700-799	0.38	-0.02	0.07	0.00	0.67	-1.90	6.60	0.00
$800-899	0.00	-0.01	-0.03 0.02	0.05 0.06	0.24	2.51 0.00	0.12 0.00	-1.50 0.00
$900-999	[illegible]	0.06						
$1000+	0.00	0.00.	0.00	-0.13	-0.37	-0.21	-0.34	-0.57

TABLE A.7

SELECTION PROBABILITIES
AFTER COLLAPSING

Income	Household Size			
		2	-	6+
Zero	0.61	0.14	0.67	0.65
$1-99	1.04	0.19	0.00	0.03
$100-299*	0.18	1.01	0.63 1.30	-0.66
$300-499*		0.83		0.35
$500-599	0.00*	0.01*	0.58	1.03
$600-699	0.00*	0.01*	0.73	3.45
$700-799		-0.02	0.15	2.26
$800-899	0.68	-0.01	0.03	2.57
$900-999	0.00	0.06	0.12	0.53
$1000+	0.00	0.86	-0.06	0.19

* Indicates collapsed rows, columns, or cells

TABLE A.8

SELECTION PROBABILITIES
AFTER BOUNDING OUT OF RANGE VALUES

Income	1	Household Size		
	0.61	2	3 5	6+
Zero		0.14	0.67	0.65
1-99	0.15	0.19	0.00	0.03
100-299"	1.00	1.00	1.00	0.00
300-499*	0.18	0.83	0.63	0.35
600-699	0.00*	0.01*	0.58	1.00
700-799	0.00* 0.09	0.01* 0.00	0.73 0.15	1.00
700-799				1.00
800-899	0.38	0.00	0.03	1.00
900-999	0.00	0.06	0.12	0.53
1000+	0.00	0.86	0.00	0.19

* Indicates collapsed rows, columns, or cells