ERROR ANALYSIS OF FOOD STAMP MICROSIMULATION MODELS *

Alan Zaslavsky and Sally Thurston, Harvard University
Alan Zaslavsky, Department of Statistics, 1 Oxford Street, Cambridge, MA 02138

Microsimulation models are one of the main tools for analysing the effects of proposed changes in Federal programs, including changes in provisions of transfer payment programs. Food stamp programs, in particular, involve many complex rules for determining eligibility and setting benefit levels. These rules are frequently revised, and when they are, it is essential to estimate the effects of the changes on the aggregate costs of the program. Equally importantly, effects of the changes must be considered for particular population subgroups defined by variables such as geographical region, household size, household income, participation in other programs, and other household characteristics.

Three microsimulation models are used by the U.S. Food and Nutrition Service to estimate these effects: the QC Minimodel, the FOSTERS model, and the MATH model (in increasing order of complexity). Each of these models produces a variety of tables containing point estimates, but no estimates of error (standard errors or confidence intervals) associated with the tabulated estimates. Yet in fact there are many forms of error that may affect these estimates, including sampling error, stochastic simulation variability, and uncertainty about parameter values; each of these may affect the results of one or more of the models. The objective of this project is to explore methods for calculating and presenting error estimates for quantities estimated by these microsimulation models.

In our research, we have developed methods for estimation of various forms of uncertainty in the QC and FOSTERS models, and implemented these methods in software. In this paper, we first discuss the general principles that have motivated our efforts, and then briefly describe the details of the implementation in these two models. Our results are more fully documented in a technical report available from the authors.

*This research was supported by the Food and Nutrition Service, U.S. Department of Agriculture, Grant #59-3198-2-053. The authors thank John Czajka, Robert Cohen, Julie Sykes, and Carol Tripp of Mathematica Policy Research for their cooperation.

1 General principles

1.1 Elements of a microsimulation model

Typically, the development of a microsimulation model, such as the QC Minimodel, FOSTERS, or MATH models, involves the following steps:

1. *Selection of the base data set:* The base data set is a collection of records representative of the population of interest. It may be derived from a survey or from a sample of administrative records.

2. *Merging with supplementary files:* Sometimes, not all of the items required for the microsimulation are available in a single file derived from a single data source. In these cases, supplementary files may be merged into the original file by statistical matching. In statistical matching, units in the base file are matched with units in the the supplementary file with identical or similar values on a set of matching variables that are available in both files. (The alternative is "exact matching" of records that correspond to the same unit, but this is impossible when the two files are based on independent samples and therefore have little or no overlap.) Values for variables that are not available in the base file are drawn from the matched cases in the supplementary file.

3. *Processing and imputation:* New quantities may be required that are not present on the original file. There may be missing data for some variables. Or some required variables may be "missing" altogether so it is necessary to impute all of their values. For example, the file may contain only annual income and number of months worked, so that in order to calculate monthly benefits eligibility, it is necessary to impute monthly employment status and income.

4. *Simulation of individual behaviors:* In some microsimulation models, it is of interest to simulate behaviors of individual units. The procedure of "dynamic aging" involves simulation of changes in the status of individual units over time, projecting forward from the cross-sectional moment captured in the base data, in order to predict the makeup of the population at some future time. Some models simulate the "behavioral response" of individual units to hypothetical changes in policies or socioeconomic conditions. Dynamic aging and behavioral response modeling may coexist in the same microsimulation.

5. *Simulation of outcomes:* A policy-oriented mi-

crosimulation is concerned with estimating the effects of policy alternatives on some outcomes, which must first be calculated for individual units of the sample file. Food stamp models, and many others, involve application of a set of laws and regulations, fixed for the duration of the microsimulation run, to variables whose values are calculated in steps (1)–(4) above. Examples of outcomes, in various contexts, include food stamp eligibility and benefit levels, tax liability, and medical reimbursement levels.

6. *Aggregating and summarizing outcomes:* Once outcomes for individual units have been simulated, they must be summarized in meaningful ways, i.e. in a ways that corresponds to policy questions that are of interest. This involves calculation of some measure of costs and benefits, at various levels of aggregation corresponding to domains of interest, such as the entire nation, geographical regions, or non-geographical sectors of the population such as income groups.

There is an important practical distinction between two phases in the generation of microsimulation estimates. The microsimulation creation phase consists of all those activities which lead to the production of the "deliverable" microsimulation package. The microsimulation execution phase consists of those activities that are conducted by the end user, incorporating assumptions and simulated policy alternatives that are of particular interest to that user. The division between these phases is a matter of computational and technical resources and convenience, rather than of principle: those activities requiring particular expertise and judgement or manipulation of large data files must be completed in the creation phase, while those for which the end-user can easily set parameters and carry out the "run" using automatic procedures should be incorporated into the execution phase in order to allow flexibility to the end-user. Selection of the base data set, statistical matching, and imputation together typically lead to creation of the base file that is distributed as part of the microsimulation package. The remaining activities are conducted at each "run" of the microsimulation. The distinction between these phases is fuzzy when the producer and consumer are the same group of people, as when a microsimulation model is created for use by a single administrative office or research group.

1.2 Sources of uncertainty in microsimulation models

Each of the steps described above involves some uncertainties and therefore contributes to the total error of estimates from the model. Here we categorize these uncertainties and discuss how they appear in the steps of model creation and simulation.

1. *Sampling.* The base data set is typically derived from a sample, either from a survey (such as the CPS or SIPP) or from administrative records. Therefore, any estimates from the microsimulation are subject to sampling variability. If other data sets are used in creation of the final microsimulation database, these also contribute sampling variability to the results.

2. *Stochastic simulation uncertainty.* A stochastic (random) element may enter into the microsimulation model at several stages. In models involving simulation of behavioral responses, more than one response is possible for some subjects in the model, and therefore behaviors are simulated by randomly drawing the response from a predictive distribution. (For example, a subject may choose whether or not to enroll in the food stamp program when her/his benefit level increases.) When files are merged while creating a microsimulation database, there can be a stochastic aspect in the selection of the matching records from the secondary database for each record in the primary database. Similarly, when some items on the database are imputed, random draws may be taken from the predictive distribution of the missing items. Finally, correction of biases in the database may take the form of stochastic modifications of the original records. Note that stochastic elements may enter both in the model creation phase (imputation, merging and correction) and the model running phase (behavioral response and correction).

Each of the steps described above may be carried out without explicitly including a random element in the simulation. For example, we could impute a single value for each missing or biased item and unknown behavioral response using a deterministic rule, and merge databases by a deterministic method. However, inferences which do not take into account real uncertainties will be overly sharp. In other words, they do not properly represent our degree of uncertainty about the outcomes of interest.

3. *Model parameter estimation.* Models enter into the creation and use of the microsimulation database at several points. First, the survey that supplies the base data set may be subject to biases, such as coverage bias, nonresponse bias, and response error. Correction of these biases requires some modeling, at least implicitly. Second, merging of databases involves a model, explicitly or implicitly. The "conditional independence assumption," which states that non-overlapping variables on the base and supplementary databases are inde-

pendent, conditional on the overlapping (matching) variables, is only one of many possible assumptions upon which statistical matching may be based; it is the simplest but not necessarily the most plausible. Imputation of missing data also involves some model, more or less explicitly. Finally, simulation of behavioral response requires expressing, through a model, the probabilities of different possible responses.

For each of these models, parameter values must be estimated or posited. In some cases, the model parameters are derived from a study based on a supplementary data set, distinct from the database that is incorporated into the microsimulation. In such cases, there will be sampling variability in estimation of the parameters. In other cases, some of the required parameters are inestimable from any available data, and instead values must be posited based on prior beliefs about plausible relationships. Uncertainty about these prior beliefs is appropriately represented by a probability distribution on possible values of the parameters. Uncertainty of this type about nonsampling errors in estimates may be important even when there are data to be used in estimation.

4. *Model specification.* In each of the steps in which models enter into estimates, as described above, there may be alternative specifications for the model, and no compelling basis in either data or prior theory for choosing one over the others. If this is the case, then uncertainty about model specification may be a contributor to overall uncertainty about the outcomes.

5. *Macro effects.* Projections about outcomes may be affected by assumptions about influences that lie beyond the scope of the model. For example, projected food stamp expenditures in a future year depend in part on assumptions about the state of the economy in that year, which affects the probable employment status and income of food stamp units in the database. Because these assumptions affect *all* the units in the microsimulation in a systematic way, we refer to them henceforth as "macro effects"; these include effects of macroeconomic variables. Uncertainty about these variables contributes to uncertainty about model outcomes.

6. *Welfare measures.* In a complex multiobjective microsimulation, there may be competing aggregate measures of welfare (costs and benefits), which weight the importance of the various aspects of welfare in different ways. Different "consumers" of the microsimulation may have different preferences among these alternative measures. From the point of view of the microsimulation "producer," it is important to know how sensitive comparative outcomes of interest are to the choice among these alternative measures. In this sense, the choice of welfare measures may be regarded as a quantifiable component of uncertainty for the producer, even though no consumer regards the choice among measures as random.

1.3 Treatment of uncertainty in microsimulation models

As we have shown, many types of uncertainty affect the outputs of public policy microsimulation models. Yet almost all such models produce only point estimates, without estimates of uncertainty. This lack of attention to uncertainty may be based on a belief that decision-makers are uninterested in measures of uncertainty and would be confused to see more than one number for any given estimand. While we agree that results must often be presented in a simplified form, we still believe that analysis of uncertainties is an essential task for the analysts who create and use microsimulation models.

First, when results are reported without measures of uncertainty, the consumer of the analysis is misled into a spurious sense of the accuracy of the every detail in the results. There is then no way of distinguishing between features of the results that are well established and those that are controversial or are no more than random noise in the data. At the very least, the analyst has a responsibility to guide decision-makers away from reliance on estimates that are essentially a throw of the dice.

Second, a full analysis of uncertainties in an analysis makes it possible to report the extent to which its conclusions are affected by competing assumptions, whether these assumptions pertain to alternative analytic methods or to beliefs about "macro" effects exogenous to the model). At best, this analysis may show that the results are robust against a range of assumptions, so that users of the analysis may avoid unnecessary debate about irrelevant issues. Even if alternative estimates differ substantially, it advances the debate to have the effect of each controversial choice presented within a common framework, rather than in a set of separate and competing reports.

Finally, an analysis of uncertainties in existing microsimulations is the basis for design of future data collection, modeling, and simulation efforts. It tells the analyst to what extent the accuracy of results is affected by various limitations, such as model uncertainty, limitations on available data, and restrictions on computational effort.

It is important to recognize that not every analysis must include the most complicated possible analysis of uncertainty. Once some experience is gained showing that particular sources contribute only a trivial fraction of the total uncertainty for a variety of different estimands, then it is unnecessary to examine that source repeatedly. Also, some of the details of the error analysis may be of more importance for the design of future simulations than for reporting of the results that are immediately required.

1.4 Strategies for estimating and representing uncertainty

Conceptually, we may divide the sources of uncertainty in a simulation into those which correspond to a choice among identifiable alternatives, and those which correspond to unstructured random "noise" in the process. The first set includes uncertainties about model specification, macro effects, and utility measures. For these uncertainties, it is appropriate to conduct a sensitivity analysis, in which the effects of changing particular inputs to or features of the model are tabulated individually. This permits consumers of the model to observe the effects of varying particular assumptions of interest to them.

The second set unambiguously includes stochastic simulation variability and sampling variability. No analyst would be interested in studying the results obtained from a particular stream of random numbers or a particular survey sample as compared to another produced from the same design; what is of interest, rather, is the amount of variability among the results as a whole. This can naturally be summarized by a variance component.

The line between these categories is not a sharp one, however. For example, uncertainty about model parameters values is caused, in part, by sampling variability in the studies which provided the estimates. However, other uncertainties within these studies, and even controversies over the validity of some or all of the estimates, may be bigger contributors to the range of uncertainty about these parameters. Conversely, disagreements about the appropriate choice of model specifications may be resolved, from a subjective probability standpoint, by positing a probability distribution over the competing specifications, representing a combination of prior plausibility and the strength of the evidence for the alternatives. Furthermore, even when is is not appropriate to regard some set of options as possessing a probability distribution (for example, in choosing among measures of welfare), it is still useful to summarize the sensitivity analysis with a rough measure of the importance of each factor by calculating a variance component for the main effect of that factor. We therefore consider it appropriate to report both the sensitivity analysis and the variance components for systematic effects such as these.

The choice of levels of the various random and systematic factors may be regarded as a classical experimental design question. A straightforward approach is to repeatedly calculate estimates while varying the inputs over the points of a factorial design. The resulting estimates can then be analyzed directly by classical ANOVA methods to obtain estimates of variance components and main effects. We may write down models for the ways in which the effects of the various uncertainties are reflected in the estimated variance components.

Simulation experiments involve repeating the entire simulation process with alternative values of the parameters controlling the simulation, e.g. alternative model parameter values, alternative matching methods, alternative random number seeds for the stochastic component of the simulation, and alternative subsamples corresponding to simulation of sampling error. Parameters corresponding to simulation activities that take place in the *execution phase* may be modified on repeated runs of the model. Parameters corresponding to simulation activities that take place in the *creation* phase must be modified during that phase; therefore the microsimulation database itself must contain alternative realizations corresponding to different parameter values. (This closely follows the increasingly standard practice of multiple imputation of public-use databases.)

2 Experience with the QC and FOSTERS models

2.1 Description of models and sources of variability

The QC Minimodel is based on a sample of approximately 63,000 records from a quality control audit file of food stamp participants. The information in this model is only that which would be available on the food stamp records.

The FOSTERS model is based on the SIPP, and contains about 22,000 records. Many of these correspond to units that would not be eligible for food stamp benefits under any plausible plan, so the sample size is effectively much smaller.

Both the QC and the FOSTERS models are static models, i.e. they work with a fixed cross-sectional sample rather than simulating changes over time. In the QC model, there is no flexibility in modeling individual behaviors, as all units included in the data file are assumed to participate in the food stamp program. Therefore, the only obvious source of uncertainty is sampling variability due to the fact that the file is a sample of all food stamp records. There are probably other uncertainties that affect QC model estimates but are not explicitly represented in the model, such as issues about data quality and about effects of benefit levels on participation. We have chosen not to model the effect of these other sources of error.

FOSTERS models the participation decision of potential beneficiaries whose eligibility or potential benefits change under a modified plan. Part of the uncertainty in estimates from the FOSTERS model therefore stems from uncertainty regarding the correct specification and parameter values for the mathematical model that predicts participation. There is also a stochastic element to the prediction of individual response. While modeling of individual behaviors in FOSTERS is limited relative to the more sophisticated MATH model, we anticipate that the paradigm developed in this work can be extended to the more complex model.

2.2 Factors and experimental design

In order to calculate variability estimates, we first must define a set of "conditions" that correspond to different possible realizations of the data and behavioral outcomes, and then create multiple versions of each of the summary arrays that are aggregated by the model code as it passes through the data, one for each condition. Therefore, the data structures for the arrays are modified by adding a dimension to every array, corresponding to the alternative conditions. The program is correspondingly modified by adding another level of looping within the main loop that processes the records and aggregates the arrays. For each record of the data file, this new inner level is executed once for each condition. It repeats the calculation of the contribution to the arrays for each condition (to the extent that the calculation varies between conditions) and aggregates that contribution into the level corresponding to that condition in every array.

The term "conditions" is used by analogy to the usual concepts of experimental design; we are performing a simulation experiment to calculate the variance of any estimate due to uncertainties in the model. Each condition is specified by giving the levels of one or more factors.

The conditions created to incorporate sampling variability involve using a grouped jackknife estimator. At each level of the jackknife "factor," we omit some fraction of the records. The omitted records are a systematic sample of the records, e.g. every fourth record. We chose this method both because it was easy to implement, and because if there is any stratification implicit in the order of the records, this stratification is preserved. Note that the additional calculation required for these jackknife "conditions" is minimal, because all that is required in the loop over conditions is to determine whether or not to include a particular record in the aggregate for the corresponding level of the arrays.

In the FOSTERS model, another factor represents alternative values of participation parameters. The probability of participation for a newly eligible food stamp unit is based upon observed rates (calculated from the SIPP) for various categories of units. These rates are modified to correct for underreporting of participation in the SIPP. We represented uncertainty about this correction by introducing a parameter for a shift in this probabilities by an arbitrarily selected amount on the logistic scale.

Because the FOSTERS model is stochastic, another component of variability in its results is that due to the variability of the random numbers used in simulation of the participation decision. This component can be estimated by calculating estimates for several "conditions" that are in fact identical except for the random number seed used; estimates for these conditions differ only due to stochastic simulation variability. For each random number seed, we used the resulting stream of random numbers and the "antithetic" stream obtained by subtracting those numbers from one; this induces a negative correlation between conditions that reduces the stochastic variability of means across runs.

The design for error analysis in the QC Minimodel is a one-way layout, because only one source of error (sampling variability, represented by jackknife replication) is considered. The design for the FOSTERS model, as now implemented, is a four-factor full factorial design; the four factors are jackknife replication, participation parameter value, random number seed, and stream. Clearly, the number of conditions could grow very rapidly as more parameters are added; this motivates our interest in applying fractional factorial designs in future research.

2.3 Calculation and display of results

The existing QC and FOSTERS models display results in a series of tables. We have modified the program procedures that calculate table entries to give standard errors for each table entry as well as point estimates. The standard errors are calculated by comparing estimates based on aggregates calculated under the various conditions (properly reweighted in the case of the jackknife estimates). For the QC model, which is affected by a single type of variability (sampling variance), we add a line to each table showing the standard error of every entry. For the FOSTERS model, in which there is more than one source of variability, we offer a breakdown of the variance of each entry into its sources in an ANOVA-like format.

The user of microsimulation models may be interested in various calculated quantities that are not presented directly in the tables, such as differences between plans or differences between population groups in the effect of a change of plan. Point estimates for these quantities may be obtained by simple arithmetic on values available from the standard tables. Standard errors, however, cannot be calculated from tabulated standard errors, because they depend upon covariances between table entries. Given the large number of covariances and possible calculated estimands, it is obviously not possible to tabulate either all the covariances or all the possible estimands with their standard errors.

Instead, we have adopted the approach of attaching an interactive, query-driven interface to the models to permit the model user to obtain estimates with standard error for a wide range of possible estimands. Our assumption is that most estimates of interest can be obtained by summing across some or all levels of a table in some direction, calculating differences between two levels of the table in some direction, or some combination of the two operations. We call such linear combinations of table entries "aggregates" (for simple sums) or "contrasts" (when a difference is calculated between levels on one or more dimensions), and we have designed a simple interface which prompts the user to describe an aggregate or contrast and then returns a point estimate with standard error. These user queries are logged on file for later examination. As with table entries, an ANOVA-like breakdown of sources of uncertainty can be prepared when there is more than one such source.

2.4 The interactive query module

Within a plan, the user may ask for a point estimate and its standard error for the aggregate of table entries, collapsed over one or more dimensions. For dimensions that are ordered, the user is asked for a lower and upper category, and all levels within this range are then collapsed. For unordered dimensions, the user is asked for up to 2 categories which are then combined. Estimates for two different plans may be contrasted, but not added. In requesting contrasts involving two plans, the user may ask for aggregates of table entries collapsed over cells in the same way, and a contrast of this quantity between the two plans of interest is then calculated.

Aside from the dimension of plan, the other unordered categories which occur in various tables are: welfare status (10 overlapping categories), deduction type (medical, child care, shelter, earnings, and standard), household categories (9 overlapping categories), and region (northeast, midwest, south and west). Ordered categories that occur in various tables are: household size (1, 2, ..., 7+), gross income as a percent of poverty level (5 or 7 categories, depending on the table), and participation categories (the gainer/loser status under a proposed plan as compared to the base plan).

The new program module calculates point estimates and standard errors of these aggregates and contrasts; a detailed description of the possible estimands for various tables is omitted. The module is interactive and is executed following the calculation of the tables.

An example of input and output from using this interactive part of the program follows. (This output is based on a subset of 6000 records.)

```
For what table do you want a contrast?
0=standard summary table, 1=table1,
2=table2, 3=gainer/loser table,
4=regional gainer/loser table
1
For what variable do you want
information? 1=hhold counts,
2=persons, 3=benefits
1
Please enter 2 plan numbers (same
number if want comparison within single
plan) Note: Since hhold counts and
persons do not vary by plan, entering 2
different plans for these will give
mean=SE=0
5 5
Please enter lower then upper category
numbers (same number if want contrast
within 1 category) These numbers are
poverty levels.  1= <=0%, 2=1-50%,
3=51-100%, 4=101-130%, 5= >130%
2 4
Please enter lower then upper hhold
size (1-7, where 7 is hholds of size
>=7)
1 2

Mean=547167.1408      SE=8091.2274
```